

INTÉGRATION SUR LES VARIÉTÉS p -ADIQUES
[d'après Coleman, Colmez]

par Christophe BREUIL

TABLE DES MATIÈRES

Introduction	1
Conventions et notations	3
1. Construction de l'intégration p -adique	3
1.1. La construction de Colmez	3
1.1.1. Énoncé du théorème	3
1.1.2. Description de la preuve	5
1.1.3. Le point de vue de Zarhin	9
1.2. La construction de Coleman	9
1.2.1. Quelques mots d'introduction sur la géométrie rigide	9
1.2.2. Intégration des formes de seconde espèce	10
2. Applications	15
2.1. L'accouplement des périodes p -adiques	15
2.1.1. Quelques rappels sur B_{dR}^+	15
2.1.2. Accouplement des périodes p -adiques : énoncé du théorème	16
2.1.3. Esquisse de preuve et relations de Riemann p -adiques	18
2.2. Les polylogarithmes p -adiques	21
2.2.1. Itération de l'intégration sur les courbes	21
2.2.2. Construction des polylogarithmes p -adiques	24
2.3. Autres applications	27
RÉFÉRENCES	29

Introduction

Soit X une variété algébrique lisse sur le corps \mathbf{R} des nombres réels. Les points de X à valeurs dans \mathbf{R} ou dans \mathbf{C} , que l'on note $X(\mathbf{R})$ et $X(\mathbf{C})$, forment une variété analytique

réelle ou complexe, sur laquelle on dispose de tout l'arsenal de la théorie classique de l'intégration, développé principalement avant le début du XX^e siècle. Depuis l'introduction par Hensel des nombres p -adiques et de la topologie p -adique dans les années 1900, les mathématiciens n'ont eu de cesse d'étendre à \mathbf{Q}_p ce qu'ils savaient faire sur \mathbf{R} . Ainsi peut-on se demander s'il existe une théorie de l'intégration sur $X(\mathbf{Q}_p)$ lorsque X est une variété algébrique lisse sur \mathbf{Q}_p . Or, la topologie p -adique est très différente de la topologie classique car elle est totalement discontinue, c'est-à-dire que tout point possède une base de voisinages à la fois ouverts et fermés ou encore qu'il n'y a pas d'obstruction à passer du local au global ! Comme tout le charme réside essentiellement dans cette obstruction, cette tâche a longtemps semblé vouée à des sorites sans grand intérêt. Pourtant, l'existence du logarithme p -adique, connue dès les tout débuts, fournissait un exemple tout à fait non trivial d'une fonction p -adique qui avait bien l'air d'une "primitive p -adique"...

La situation change au début des années soixante avec l'introduction par Tate des espaces analytiques rigides ([Ta2]). L'idée en soi est naturelle : puisque la topologie p -adique a trop d'ouverts, on va en sélectionner certains plus "jolis" que les autres. On connaît depuis l'essor de cette théorie, dont l'un des aboutissements fut les théorèmes d'uniformisation p -adique de Mumford et Raynaud ([Mu2], [Ra3]). Coleman le premier, dans les années 80, réussit à utiliser la "topologie rigide" pour définir dans [Cole3] une théorie de l'intégration des 1-formes différentielles algébriques de seconde espèce sur les variétés algébriques projectives (ou propres) et lisses sur \mathbf{Q}_p avec bonne réduction. Motivé par les applications aux fonctions L p -adiques, il étend dans le cas des courbes son intégration à toutes les 1-formes rationnelles, et même à des formes non nécessairement algébriques ou rigides, ce qui lui permet par exemple de définir des polylogarithmes p -adiques. Enfin, au début des années 90, Colmez dans [Colm1], et indépendamment Zarhin dans [Za1], parviennent à rendre possible l'impossible en intégrant sur du totalement discontinu : non seulement ils étendent l'intégrale de Coleman au cas de variétés algébriques supposées seulement lisses sur \mathbf{Q}_p et à toutes les 1-formes rationnelles fermées, mais ils le font sans passer par la géométrie rigide. Cependant, leurs méthodes ne s'appliquent pour l'instant qu'aux 1-formes algébriques et ne permettent pas de retrouver les polylogarithmes p -adiques par exemple. Une fois débarrassé des hypothèses de propreté et de bonne réduction, Colmez généralise alors de façon systématique dans [Colm2] beaucoup des théorèmes de Coleman, et obtient en outre une preuve originale et particulièrement élégante, car plus proche de son analogue complexe, de l'accouplement des périodes p -adiques des variétés abéliennes. Dernièrement, Besser a donné dans [Bes] une généralisation de l'intégrale de Coleman (sur les variétés propres lisses avec bonne réduction) au cas de n -formes fermées avec $n > 1$.

L'objectif de cet exposé est de présenter dans une première partie l'intégration de Colmez et l'intégration de Coleman, puis dans une seconde partie certaines de leurs applications. Il a été conçu plutôt comme une introduction au sujet et non comme un "survey". Ainsi, j'ai pris le parti d'exposer de façon assez complète les deux constructions de l'intégration, puis de développer avec quelques détails deux applications particulières (essentiellement une par auteur), enfin de mentionner très brièvement quelques autres. L'exposé n'est donc pas exhaustif (certains aspects ne sont pas du tout abordés) et je prie le lecteur - ou l'auteur - averti de me pardonner s'il ne trouve pas développée ici son application favorite.

Durant la préparation de cet exposé, j'ai bénéficié de conversations orales ou "électroniques" avec plusieurs personnes dont : Amnon Besser, Jean-Benoît Bost, Robert Coleman, Pierre Colmez, Jean-Louis Colliot-Thélène, Philippe Gille, Bernard Le Stum, William Messing et Michel Raynaud. Qu'ils soient tous chaleureusement remerciés pour leur aide. Je suis reconnaissant à Amnon Besser, Robert Coleman et Glenn Stevens de m'avoir envoyé respectivement les références [Bes], [CI] et [St] (parmi d'autres ; la lecture de [St] en particulier m'a facilité la rédaction de la partie 2.2). Enfin, je remercie Jean-Benoît Bost, Laurent Clozel, Jean-Marc Fontaine, Michel Raynaud et surtout Pierre Colmez pour leurs remarques sur des versions préliminaires de ce texte.

Conventions et notations

Dans tout cet article, on fixe un nombre premier p et une clôture algébrique $\overline{\mathbf{Q}}_p$ de \mathbf{Q}_p . On note \mathbf{C}_p la complétion p -adique de $\overline{\mathbf{Q}}_p$, $\mathcal{O}_{\mathbf{C}_p}$ l'anneau des entiers de \mathbf{C}_p et $|\cdot|$ la valeur absolue p -adique sur \mathbf{C}_p donnée par $|x| = \frac{1}{p^{\text{val}(x)}}$ où val est la valuation p -adique normalisée par $\text{val}(p) = 1$. Le logarithme d'Iwasawa est l'unique fonction $\log : \mathbf{C}_p - \{0\} \rightarrow \mathbf{C}_p$ telle que :

$$\begin{aligned} \log(1-x) &= -\sum_{n=1}^{+\infty} \frac{x^n}{n} \quad \text{si } |x| < 1 \\ \log(xy) &= \log(x) + \log(y) \\ \log(p) &= 0. \end{aligned}$$

Le choix $\log(p) = 0$ est arbitraire (on peut même décider que $\log(p)$ est une indéterminée, cf. [Colm2]) mais naturel du point de vue de la théorie des nombres. Une variété est un schéma X séparé de type fini et géométriquement intègre sur un corps K . Nous ne considérerons en fait que des variétés lisses, c'est-à-dire telles que $X \rightarrow \text{Spec}(K)$ est un morphisme lisse. Une variété abélienne sur K est une variété X sur K munie d'une structure de K -schéma en groupes et telle que $X \rightarrow \text{Spec}(K)$ est propre. Cela entraîne $X \rightarrow \text{Spec}(K)$ lisse. Une fonction rationnelle (resp. une 1-forme rationnelle) sur une

variété X est un élément du corps des fonctions $K(X)$ (resp. des différentielles de Kähler $\Omega_{K(X)/K}^1$) qu'on peut voir comme une fonction (resp. une 1-forme) régulière sur un ouvert de Zariski U de X . Si X est une variété sur K , on supposera toujours $X(K) \neq \emptyset$.

1. CONSTRUCTION DE L'INTÉGRATION p -ADIQUE

1.1. La construction de Colmez

Dans cette partie, K désigne soit \mathbf{C}_p , soit une extension finie de \mathbf{Q}_p dans \mathbf{C}_p .

1.1.1. *Énoncé du théorème.* Soit X une variété lisse sur K de dimension d . On a supposé $X(K) \neq \emptyset$. Pour tout ouvert de Zariski affine U de X , il existe une immersion fermée $U \hookrightarrow \mathbf{A}^m$ pour $m \gg 0$ et on définit la topologie p -adique sur $U(K)$ comme la topologie induite par la topologie p -adique sur $\mathbf{A}^m(K) = K^m$. C'est indépendant de l'immersion choisie. Par recollement, on obtient la topologie p -adique sur $X(K)$. On appelle ouverts p -adiques les ouverts pour cette topologie. Tout point de $X(K)$ admet pour la topologie p -adique une base de voisinages ouverts et fermés puisque tel est le cas sur K^m : cette topologie est donc totalement discontinue. Comme X est lisse, pour tout $x \in X(K)$, il existe un ouvert de Zariski $U_x = \text{Spec}(K[x_1, \dots, x_n, z_1, \dots, z_d]/(f_1, \dots, f_n))$ contenant x où $(\frac{\partial f_i}{\partial x_j})_{1 \leq i, j \leq n}$ est de déterminant inversible et tel que $x \in U_x(K)$ correspond à $(x_i = 0, z_j = 0)$. Pour définir la structure *analytique* p -adique sur $X(K)$, rappelons le :

Lemme 1.1.1.1. — *La projection sur les d dernières composantes z_1, \dots, z_d induit un homéomorphisme entre un ouvert p -adique V de $U_x(K) \subset X(K)$ contenant x et la boule ouverte $B(0, \delta_-)^d = \{(z_1, \dots, z_d) \in K^d \mid |z_i| < \delta, 1 \leq i \leq d\}$ pour un δ suffisamment petit dans $\mathbf{R}^+ - \{0\}$.*

Cela résulte d'un algorithme élémentaire (théorème des fonctions implicites!), voir par exemple ([Colm2], A.3.4). Remarquons que tout ouvert p -adique peut s'écrire comme une réunion *disjointe* de boules ouvertes comme en (1.1.1.1) (car si B_1^d et B_2^d sont deux telles boules, on a soit $B_1^d \cap B_2^d = \emptyset$, soit $B_1^d \subset B_2^d$, soit $B_2^d \subset B_1^d$).

Définition 1.1.1.2. — *Une fonction localement analytique sur un ouvert p -adique W de $X(K)$ est une fonction f de W dans K telle que, pour tout $x \in W$ et pour un choix (donc tout choix) de paramètres locaux z_i autour de x comme en (1.1.1.1), il existe $S \in K[[z_1, \dots, z_d]]$ et $r \in \mathbf{R}^+ - \{0\}$ ($r \leq \delta$) tels que $S(z_1, \dots, z_d)$ converge si $|z_i| < r$ et coïncide avec f sur ce voisinage.*

Exemple 1.1.1.3. — Toute fonction rationnelle f régulière sur un ouvert de Zariski U est une fonction localement analytique sur $U(K)$. Soit $\log(f)$ la fonction définie comme la composée $U(K) - \{x \mid f(x) = 0\} \xrightarrow{f} K^\times \xrightarrow{\log} K$: $\log(f)$ est une fonction localement analytique sur $W = U(K) - \{x \mid f(x) = 0\}$.

On définit l'espace analytique p -adique X^{an} associé à X comme l'espace annelé $X(K)$ muni de la topologie p -adique et du faisceau $\mathcal{O}_{X^{an}}$ des fonctions localement analytiques. Si \mathcal{O}_X est le faisceau structural sur X , $\mathcal{O}_{X^{an}}$ restreint à la topologie de Zariski est un faisceau de \mathcal{O}_X -modules et, en utilisant (1.1.1.1), on peut définir une différentielle $d : \mathcal{O}_{X^{an}} \rightarrow \mathcal{O}_{X^{an}} \otimes_{\mathcal{O}_X} \Omega_{X/K}^1$ où $\Omega_{X/K}^1$ est le faisceau des différentielles de Kähler. Si f est une fonction rationnelle, on a $d \log(f) = \frac{df}{f}$. On note $K(X)_{log} = K(X)[\log(f), f \in K(X)^\times]$ où les éléments de $K(X)$ sont vus comme des fonctions localement analytiques. Les fonctions f de $K(X)_{log}$ sont toutes localement analytiques sur un ouvert de Zariski $U_f(K)$ de $X(K)$. Le lemme suivant sera utilisé plusieurs fois :

Lemme 1.1.1.4. — *Soient $f \in K(X)_{log}$ et $W \subset X(K)$ un ouvert p -adique. Si $f|_W = 0$ alors $f = 0$.*

Voir ([Colm2], II.1.7) pour la preuve. Inutile de dire que cet énoncé est faux pour des fonctions localement analytiques quelconques.

Notre objectif est de définir une théorie de l'intégration pour les 1-formes rationnelles fermées sur X . À cause du lemme de Poincaré, toute 1-forme fermée s'intègre localement pour la topologie p -adique sur une boule suffisamment petite, donc s'intègre globalement. Le problème est qu'il y a une multitude de constantes locales d'intégration : une sur chaque boule ! Et pourtant, parmi cette infinité de primitives, il en existe miraculeusement une plus jolie que les autres :

Théorème 1.1.1.5. — ([Colm2], [Za2]) *Il existe une unique façon d'associer à une variété X lisse sur K et une 1-forme rationnelle fermée ω sur X une fonction f_ω bien définie à addition près d'une constante et localement analytique sur $X(K) \setminus \{\text{pôles de } \omega\}$, de sorte que :*

- (i) $df_\omega = \omega$
- (ii) $f_{\lambda\omega + \lambda'\omega'} = \lambda f_\omega + \lambda' f_{\omega'}$ ($\lambda, \lambda' \in K$)
- (iii) $f_\omega = f$ si $\omega = df$ où $f \in K(X)_{log}$
- (iv) si $g : X' \rightarrow X$ est un morphisme de variétés lisses sur $\text{Spec}(K)$, $f_{g^*\omega} = g^* f_\omega$ ($= f_\omega \circ g$).

1.1.2. *Description de la preuve.* Nous donnons ici la preuve de Colmez (voir 1.1.3 pour Zarhin). Elle se coupe en deux : 1) en utilisant les morphismes d'Albanese on montre qu'il suffit de définir une intégration sur les variétés abéliennes, 2) on construit effectivement cette intégration sur les variétés abéliennes.

Commençons par 1). Rappelons qu'à toute variété X lisse sur K , on peut associer de manière fonctorielle sa variété d'Albanese $Alb(X)$ (= une variété abélienne) encore définie sur K (lorsque X est propre et lisse sur \mathbf{C} , $Alb(X)$ est isomorphe en tant que tore complexe à $H_{dR}^1(X)^*/H_1(X(\mathbf{C}), \mathbf{Z})$). Tout point $P \in X(K)$ donne lieu à un morphisme

unique $\iota_P : X \rightarrow \text{Alb}(X)$ défini sur K tel que $\iota_P(P) = 0$. Un tel morphisme s'appelle un morphisme d'Albanese (pour tout cela, voir [Se1] ou [La] ou ([Colm2], I.5)). De plus, si X est propre, ι_P induit un isomorphisme $H_{dR}^1(\text{Alb}(X)) \xrightarrow{\sim} H_{dR}^1(X)$.

Lemme 1.1.2.1. — Soient $\iota_P : X \rightarrow \text{Alb}(X)$ un morphisme d'Albanese et ω une 1-forme rationnelle fermée sur X , alors il existe une 1-forme rationnelle fermée η sur $\text{Alb}(X)$, des fonctions rationnelles f, f_1, \dots, f_n sur X et des constantes $\lambda_1, \dots, \lambda_n$ dans K telles que :

$$\omega = \iota_P^* \eta + df + \sum_{i=1}^n \lambda_i \frac{df_i}{f_i}.$$

Indication de preuve. — Soit $\tilde{H}^1(X)$ (resp. $\tilde{H}^1(\text{Alb}(X))$) le quotient des 1-formes rationnelles fermées par $dK(X)_{\log}$ (resp. $dK(\text{Alb}(X))_{\log}$), il suffit de montrer que ι_P induit un isomorphisme $\tilde{H}^1(\text{Alb}(X)) \xrightarrow{\sim} \tilde{H}^1(X)$ et il suffit pour cela d'étendre les scalaires à une clôture algébrique \bar{K} de K . En remarquant qu'alors, si \bar{X} est une compactification lisse de X (qui existe par la résolution des singularités d'Hironaka), on a $\tilde{H}^1(X) \simeq \tilde{H}^1(\bar{X})$ (car $\tilde{H}^1(X)$ ne dépend que du corps des fractions de X) et $\text{Alb}(X) \simeq \text{Alb}(\bar{X})$, on est ramené au cas d'une variété X propre et lisse sur \bar{K} . On a dans ce cas un diagramme commutatif de suites exactes avec des isomorphismes à gauche et à droite ([Colm2], I.1.16) :

$$\begin{array}{ccccccc} 0 & \rightarrow & H_{dR}^1(X) & \rightarrow & \tilde{H}^1(X) & \rightarrow & \bar{K} \otimes_{\mathbf{Z}} \text{Pic}^0(X) \rightarrow 0 \\ & & \wr \uparrow & & \uparrow & & \uparrow \wr \\ 0 & \rightarrow & H_{dR}^1(\text{Alb}(X)) & \rightarrow & \tilde{H}^1(\text{Alb}(X)) & \rightarrow & \bar{K} \otimes_{\mathbf{Z}} \text{Pic}^0(\text{Alb}(X)) \rightarrow 0 \end{array}$$

où les flèches horizontales de droite sont les applications qui à une forme différentielle associe son résidu, un diviseur algébriquement équivalent à 0 (voir [BLR] pour Pic^0). On en déduit le résultat. \square

Proposition 1.1.2.2. — Supposons que l'énoncé (1.1.1.5) soit vrai pour les variétés abéliennes sur K (i.e. en ne considérant en (iv) que des morphismes entre variétés abéliennes), alors il est vrai pour toutes les variétés algébriques lisses sur K .

Preuve. — Pour toute 1-forme rationnelle fermée η sur une variété abélienne, on a donc une primitive f_η bien définie à une constante près. Soit ω une 1-forme rationnelle fermée sur une variété X lisse sur K , $\iota_P : X \rightarrow \text{Alb}(X)$ un morphisme d'Albanese et η, f, f_i, λ_i comme en (1.1.2.1). Si f_ω comme en (1.1.1.5) existe, alors $f_\omega = \iota_P^* f_\eta + f + \sum \lambda_i \log(f_i)$. Si $\omega = \iota_P^* \eta' + df' + \sum \lambda'_i \frac{df'_i}{f'_i}$ est une autre écriture, on a $\iota_P^*(\eta - \eta') \in dK(X)_{\log}$, d'où $\eta - \eta' \in dK(\text{Alb}(X))_{\log}$ puisque $\tilde{H}^1(X) \simeq \tilde{H}^1(\text{Alb}(X))$ (cf. preuve de 1.1.2.1), et $f_\eta - f_{\eta'} \in K(\text{Alb}(X))_{\log}$ d'après la propriété (iii). Soit $f'_\omega = \iota_P^* f_{\eta'} + f' + \sum \lambda'_i \log(f'_i) : f_\omega - f_{\omega'} \in K(X)_{\log}$ et $d(f_\omega - f_{\omega'}) = 0$, donc $f_\omega - f_{\omega'}$ est constante sur un ouvert p -adique, donc partout (1.1.1.4). La primitive f_ω modulo les constantes est donc indépendante des choix. Quitte à changer les écritures de ω , on vérifie qu'elle s'étend bien en une fonction localement analytique sur $X(K) \setminus \{\text{pôles de } \omega\}$. Il reste à montrer la propriété (iv), mais si $g : X' \rightarrow X$ est un morphisme défini sur

$K, P \in X(K), Q \in X'(K)$, il existe un (unique) morphisme $Alb(g) : Alb(X') \rightarrow Alb(X)$ tel que le diagramme :

$$\begin{array}{ccc} X' & \xrightarrow{g} & X \\ \iota_P \downarrow & & \downarrow \iota_Q \\ Alb(X') & \xrightarrow{Alb(g)} & Alb(X) \end{array}$$

soit commutatif (cela résulte de la propriété universelle de la variété d'Albanese [Se1]). La propriété (iv) découle facilement de la construction de f_ω ci-dessus et du fait que $f_\eta \circ Alb(g) = f_{Alb(g)^*\eta}$ si η est une 1-forme rationnelle fermée sur $Alb(X)$. \square

Il faut maintenant définir ces primitives sur les variétés abéliennes. Si X est une variété abélienne, on note, pour $I \subset \{1, 2\}$, $m_I : X^3 \rightarrow X$ l'application qui à $(x, h_1, h_2) \in X^3$ associe $x \oplus (\oplus_{i \in I} h_i)$ où \oplus est la loi d'addition sur X (plus exactement sur le faisceau en groupes représenté). Si α est une fonction, ou une 1-forme, rationnelle sur X , on pose $\Delta^{[2]}\alpha = m_{\{1,2\}}^*\alpha - m_{\{1\}}^*\alpha - m_{\{2\}}^*\alpha + m_\emptyset^*\alpha$. Les deux résultats suivants vont jouer un rôle clef :

Théorème 1.1.2.3. — *Soit ω une 1-forme rationnelle fermée sur X , alors $\Delta^{[2]}\omega \in dK(X^3)_{log}$.*

Théorème 1.1.2.4. — (i) *Les sous-groupes ouverts de $X(K)$ forment un système fondamental de voisinages de l'origine pour la topologie p -adique.*

(ii) *Si V est un sous-groupe ouvert de $X(K)$, alors $X(K)/V$ est un groupe de torsion.*

Le théorème (1.1.2.3) est une des incarnations du théorème du carré ([Mu1],6 cor.4). C'est un énoncé algébrique qui n'a rien de spécifique à la topologie p -adique. Pour une preuve sur \overline{K} sous cette forme via le principe de Lefschetz, voir ([Colm2],I.3.6), le résultat sur K s'en déduisant en prenant les invariants sous $Gal(\overline{K}/K)$. Le théorème (1.1.2.4) est lui par contre tout à fait particulier à la topologie p -adique. Un des arguments clef de sa preuve est que les boules ouvertes comme en (1.1.1.1) sont des groupes pour l'addition sur K^d , ce qui est rarement le cas des boules ouvertes de \mathbf{C}^d ! Voir ([Colm2],II.1.9) ou ([Cole4],4.1) pour des preuves de (1.1.2.4).

Si une primitive f_ω de ω satisfaisant les conditions du théorème existe, on doit avoir $f_{\Delta^{[2]}\omega} \in K(X^3)_{log}$ et $\Delta^{[2]}f_\omega = f_{\Delta^{[2]}\omega}$ d'après les conditions (iii) et (iv) en (1.1.1.5), ce qui entraîne $\Delta^{[2]}f_\omega \in K(X^3)_{log}$.

Théorème 1.1.2.5. — *Soient ω une 1-forme rationnelle fermée sur X et U_ω l'ouvert de Zariski de X complémentaire des pôles de ω , il existe à addition près d'une constante une unique fonction f_ω localement analytique sur $U_\omega(K)$ telle que :*

- 1) $df_\omega = \omega$
- 2) $\Delta^{[2]}f_\omega \in K(X^3)_{log}$.

Preuve. — En utilisant $\Delta^{[2]}T_a^* = T_{(a,0,0)}^*\Delta^{[2]}$ (où $a \in X(K)$ et T_a est la translation par a sur $X(K)$), on se ramène au cas où ω est définie en 0.

Unicité : S'il y a deux primitives, leur différence D_ω satisfait $dD_\omega = 0$, donc D_ω est localement constant et on a aussi $d(\Delta^{[2]}D_\omega) = 0$. Mais $\Delta^{[2]}D_\omega \in K(X^3)_{log}$, d'où $\Delta^{[2]}D_\omega = 0$ par (1.1.1.4) car $\Delta^{[2]}D_\omega|_{X \times \{0\} \times \{0\}} = 0$. Pour $h_1, h_2 \in U_\omega(K)$ tels que $h_1 \oplus h_2 \in U_\omega(K)$, on a donc, quitte à remplacer D_ω par $D_\omega - D_\omega(0)$, $D_\omega(h_1 \oplus h_2) = D_\omega(h_1) + D_\omega(h_2)$. Cela entraîne que D_ω se prolonge par continuité en un morphisme de groupes sur tout $X(K)$: si V est un sous-groupe ouvert de $U_\omega(K)$ sur lequel D_ω est défini et $x \in X(K)$, d'après (1.1.2.4,ii) il existe $n_x \in \mathbf{N}$ tel que $[n_x]x = x \oplus \dots \oplus x$ (n_x fois) est dans V et on pose $D_\omega(x) = \frac{1}{n_x}D_\omega([n_x]x)$. Comme D_ω est localement constant, on peut choisir un tel V sur lequel D_ω est constant. Mais $D_\omega|_V$ est alors nul puisque c'est un homomorphisme de groupes, donc D_ω est nul partout par la formule ci-dessus.

Existence : D'après (1.1.2.3), il existe $g \in K(X^3)_{log}$ tel que $\Delta^{[2]}\omega = dg$ et on peut supposer g nulle sur $X(K) \times \{0\} \times \{0\}$ puisque $\Delta^{[2]}\omega|_{X \times \{0\} \times \{0\}} = 0$.

Première étape : Soit V un sous-groupe ouvert de $X(K)$ contenu dans $U_\omega(K)$ sur lequel ω est analytique. On peut prendre V isomorphe à une boule ouverte $B(0, \delta_-)^d$ comme en (1.1.1.1). Soit f_ω l'unique primitive analytique de ω sur V nulle en 0. On a $d(\Delta^{[2]}f_\omega - g|_{V^3}) = \Delta^{[2]}\omega - \Delta^{[2]}\omega = 0$ donc $\Delta^{[2]}f_\omega = g|_{V^3}$ car ces deux fonctions sont analytiques sur V^3 et nulles sur $V \times \{0\} \times \{0\}$.

Deuxième étape : On prolonge f_ω à $U_\omega(K)$. Notons V_n l'ouvert p -adique de $X(K)$ formé des x tels que $[n]x \in V$ et $[k]x$ n'est pas un pôle de ω pour tout $k \in \{1, \dots, n\}$. On étend f_ω à V_n en posant :

$$f_\omega(x) = \frac{1}{n} \left(f_\omega([n]x) - \sum_{k=1}^{n-1} g(0, [k]x, x) \right).$$

En utilisant $f_\omega(x \oplus y) - f_\omega(x) - f_\omega(y) = g(0, x, y)$ sur V , on vérifie que cette équation est en particulier satisfaite sur V . De plus, $f_\omega(x)$ est indépendant de n tel que $x \in V_n$ et on a $\Delta^{[2]}f_\omega = g$ là où les deux termes sont définis. En effet, ces assertions se ramènent facilement à des égalités fonctionnelles sur g seulement, qui sont automatiquement satisfaites sur V et on peut utiliser (1.1.1.4) puisque $g \in K(X^3)_{log}$ pour en déduire qu'elles sont vraies partout. On obtient une fonction f_ω localement analytique sur $\cup_n V_n$ qui est un ouvert p -adique dense de $X(K)$ contenu dans $U_\omega(K)$ et telle que $\Delta^{[2]}f_\omega = g$ partout où ça a un sens. Quitte à traduire 0 en un point $a \in U_\omega(K)$ et à utiliser $T_{(\ominus a, 0, 0)}^* \Delta^{[2]}T_a^* = \Delta^{[2]}$, le même raisonnement appliqué à $T_a^*\omega$ fournit une autre fonction $f_{\omega, a}$ localement analytique sur un ouvert p -adique dense contenant 0 et telle que $\Delta^{[2]}(f_\omega - T_{\ominus a}^*f_{\omega, a}) = 0$. Soit b un point où les deux fonctions f_ω et $T_{\ominus a}^*f_{\omega, a}$ sont définies, la fonction $h(x) = (f_\omega - T_{\ominus a}^*f_{\omega, a})(b \oplus x) - T_{\ominus a}^*f_{\omega, a}(b) + f_\omega(b)$ satisfait $h(x \oplus y) = h(x) + h(y)$ partout où elle est définie, et cela entraîne qu'elle se prolonge par continuité à $X(K)$ tout

entier (cf. la partie sur l'unicité). Donc f_ω se prolonge par continuité en a , c'est-à-dire finalement à tout $U_\omega(K)$.

Troisième étape : Soit f_ω comme dans la deuxième étape et $\eta = df_\omega - \omega$, on a $\Delta^{[2]}\eta = 0$, d'où $(\Delta^{[2]}\eta)|_{\{0\} \times \{a\} \times X} = 0$ i.e. $T_a^*\eta = \eta$. Donc η est une 1-forme localement analytique invariante par translation et nulle sur V , donc nulle partout. \square

Si P et Q sont deux points de $X(K)$ où ω est définie, posons $\int_P^Q \omega = f_\omega(Q) - f_\omega(P)$.

Corollaire 1.1.2.6. — Avec les notations précédentes :

- (i) $\int_P^Q (\lambda\omega + \lambda'\omega') = \lambda \int_P^Q \omega + \lambda' \int_P^Q \omega'$ ($\lambda, \lambda' \in K$)
- (ii) $\int_P^Q \omega = f(Q) - f(P)$ si $\omega = df$ où $f \in K(X)_{\log}$
- (iii) si $g : X' \rightarrow X$ est un morphisme de variétés abéliennes sur $\text{Spec}(K)$, $\int_P^Q g^*\omega = \int_{g(P)}^{g(Q)} \omega$.

Preuve. — Les énoncés (i) à (iii) résultent de l'unicité modulo cte de f_ω et de $g^*\Delta^{[2]} = \Delta^{[2]}g^*$. \square

Avec (1.1.2.2), cela achève la preuve du théorème (1.1.1.5).

Remarque 1.1.2.7. — On peut préciser le comportement des primitives f_ω au voisinage des pôles de ω : c'est la somme d'une fonction localement méromorphe et d'une combinaison linéaire de logarithmes de fonctions localement méromorphes.

Remarque 1.1.2.8. — Si X est une variété abélienne et $\omega \in H^0(X, \Omega_{X/K}^1)$, ω est une 1-forme fermée invariante par translation sur X ([Mu1], II.4.iii) et f_ω est alors un homomorphisme de groupes de $X(K)$ dans K . Selon la terminologie de ([Bou], III.7), une telle primitive est appelée un *logarithme de X* .

1.1.3. *Le point de vue de Zarhin.* Dans [Za1] et [Za2], Zarhin a construit indépendamment des primitives à toutes les 1-formes rationnelles fermées sur les variétés lisses qui satisfont toutes les propriétés en (1.1.1.5), et donc coïncident avec les primitives de Colmez. Les méthodes de Zarhin et Colmez sont assez proches. Dans un premier temps, Zarhin montre qu'il y a une manière canonique d'associer une primitive f_ω à une 1-forme ω invariante sur un groupe algébrique commutatif sur K . Dans un deuxième temps, il utilise que toute 1-forme rationnelle fermée sur une variété lisse, vue sur son ouvert de définition, est l'image inverse d'une 1-forme invariante sur un groupe algébrique et il prend l'image inverse de la primitive définie sur ce groupe.

1.2. La construction de Coleman

Dans cette partie, on suppose pour simplifier $K = \mathbf{C}_p$.

1.2.1. *Quelques mots d'introduction sur la géométrie rigide.* On renvoie à [Ber], [Ga] ou [Ra1] pour des introductions efficaces et concises à la géométrie rigide. Mais d'abord, pourquoi la géométrie rigide pour définir une intégration p -adique? Essentiellement pour deux raisons. La première est que les fonctions rigides localement constantes sur un espace analytique rigide connexe sont automatiquement constantes. Ceci joue un rôle clef dans les démonstrations ci-dessous. La deuxième est que les morphismes de Frobenius sur les schémas en caractéristique p se relèvent localement sur les *complétés p -adiques* des modèles lisses en caractéristique 0 (lorsque de tels modèles existent), mais pas sur les modèles eux-mêmes. On obtient ainsi des morphismes de Frobenius sur les fibres génériques au sens de Raynaud de ces complétés ([Ber],0.1.2), qui sont seulement des espaces rigides. Or, dans la théorie de Coleman, ces morphismes de Frobenius jouent un rôle analogue au théorème du carré dans l'intégrale de Colmez.

Rappelons qu'une algèbre de Tate (sur \mathbf{C}_p) est un quotient de $\mathbf{C}_p\{z_1, \dots, z_d\} = \{\sum_{\underline{\alpha}=(\alpha_1, \dots, \alpha_d)} a_{\underline{\alpha}} z_1^{\alpha_1} \dots z_d^{\alpha_d}, a_{\underline{\alpha}} \rightarrow 0 \text{ si } |\underline{\alpha}| \rightarrow +\infty\}$ pour un d convenable. Si A est une algèbre de Tate, on note $\text{Spm}(A) = \text{Spec}(A)(\mathbf{C}_p)$ l'ensemble des idéaux maximaux de A : les $\text{Spm}(A)$ jouent en géométrie rigide le rôle que jouent les schémas affines en géométrie algébrique. On veut les munir d'une "topologie" plus fine que la topologie de Zariski, mais pour laquelle les seules fonctions localement constantes sont les constantes (du moins si $\text{Spec}(A)$ est connexe). Il n'est pas difficile de trouver des sous-ensembles de $\text{Spm}(A)$ dont on a envie qu'ils soient des ouverts, par exemple $\text{Spm}(A\{z\}/(zf - c))$ où $f \in A$ et $c \in \mathbf{C}_p^*$, mais la topologie engendrée par ces ouverts redonne en général la topologie p -adique totalement discontinue, dont on ne veut pas. L'idée est alors de "sélectionner" certains ouverts et certains recouvrements, que l'on appelle admissibles, et remplacer la notion de topologie par la notion plus générale de topologie de Grothendieck. En munissant l'ensemble $\text{Spm}(A)$ de cette topologie de Grothendieck et du faisceau structural rigide (dont les sections globales sur $\text{Spm}(A)$ ne sont autres que A), on obtient l'espace analytique rigide affinoïde associé à $\text{Spm}(A)$ (voir [Ber],0.1). Les espaces analytiques rigides généraux sont ensuite définis comme les ensembles munis d'une topologie de Grothendieck et d'un faisceau de \mathbf{C}_p -algèbres qui sont localement isomorphes à un espace rigide affinoïde. On définit des morphismes d'espaces rigides, des différentielles rigides (en particulier des 1-formes), etc... : voir ([Ber],0).

Exemple 1.2.1.1. — Si $A = \mathbf{C}_p\{z\}$, $\text{Spm}(A) \simeq \{z \in \mathbf{C}_p \mid |z| \leq 1\}$, si $A = \mathbf{C}_p\{z_1, z_2\}/(z_1 z_2 - p)$, $\text{Spm}(A) \simeq \{z \in \mathbf{C}_p \mid \frac{1}{p} \leq |z| \leq 1\}$.

Si X est un schéma sur $\text{Spec}(\mathbf{C}_p)$, disons localement de type fini, on peut lui associer un espace analytique rigide X^{rig} dont l'ensemble sous-jacent est formé des points fermés $X(\mathbf{C}_p)$ de X . Tout ouvert de Zariski U dans X fournit par exemple un ouvert admissible $U(\mathbf{C}_p)$ pour la topologie de Grothendieck sur $X(\mathbf{C}_p)$. Mais d'autres ouverts jouent un rôle crucial : les tubes. Supposons pour simplifier X propre et soit \mathcal{X} un modèle propre et plat

sur $\text{Spec}(\mathcal{O}_{\mathbf{C}_p})$ de fibre spéciale Y sur $\text{Spec}(\overline{\mathbf{F}}_p)$. Le critère valuatif de propreté nous dit $X(\mathbf{C}_p) \simeq \mathcal{X}(\mathcal{O}_{\mathbf{C}_p})$, on a donc une flèche de spécialisation $sp : X(\mathbf{C}_p) \rightarrow Y(\overline{\mathbf{F}}_p)$. Soient Z un sous-schéma de Y , et $]Z[= sp^{-1}(Z(\overline{\mathbf{F}}_p))$: on peut montrer que $]Z[$ est un ouvert admissible de $X^{rig} = X(\mathbf{C}_p)$, qu'on appelle le tube de Z dans X^{rig} ([Ber],1.1.1).

Exemple 1.2.1.2. — Si Z est un ouvert de Zariski affine de Y , $]Z[$ est un ouvert affinoïde de X^{rig} ([Ber],0.2.2.1). Si $Z = \{y\}$ est un point fermé de Y et si de plus \mathcal{X} est lisse, $]Z[$ est isomorphe à la boule ouverte $B(0, 1_-)^d = \{(z_1, \dots, z_d) \in K^d \mid |z_i| < 1, 1 \leq i \leq d\}$ où d est la dimension de Y (même genre d'arguments que pour (1.1.1.1)).

Si X est un schéma sur \mathbf{C}_p , on note X^{rig} l'espace analytique rigide associé. Si X est un espace rigide sur \mathbf{C}_p , on note \mathcal{O}_X le faisceau des fonctions rigides sur X .

1.2.2. *Intégration des formes de seconde espèce.* Soit X une variété algébrique sur $\text{Spec}(\mathbf{C}_p)$ admettant un modèle propre et lisse \mathcal{X} sur $\text{Spec}(\mathcal{O}_{\mathbf{C}_p})$ et notons Y sa fibre spéciale sur $\overline{\mathbf{F}}_p$. Comme Y est de type fini sur $\overline{\mathbf{F}}_p$, Y provient d'un modèle défini sur une extension finie (non unique) \mathbf{F}_{p^r} de \mathbf{F}_p . On appelle morphisme de Frobenius sur Y tout morphisme de $\overline{\mathbf{F}}_p$ -schémas $Y \rightarrow Y$ provenant par extension des scalaires pour un certain r de la puissance $r^{ième}$ du Frobenius absolu sur un tel modèle. Si ϕ et ϕ' sont deux morphismes de Frobenius sur Y , il est clair qu'on a $\phi^s = \phi'^t$ pour des entiers s, t convenables. Par ailleurs, il résulte de [BO] qu'on a un isomorphisme :

$$H_{dR}^1(X) \simeq \mathbf{C}_p \otimes H_{cris}^1(Y/W(\overline{\mathbf{F}}_p))$$

où $W(\overline{\mathbf{F}}_p)$ désigne les vecteurs de Witt à coefficients dans $\overline{\mathbf{F}}_p$ et H_{cris}^1 le H^1 cristallin. Tout morphisme de Frobenius ϕ sur Y induit donc par functorialité un morphisme \mathbf{C}_p -linéaire $\phi^* : H_{dR}^1(X) \rightarrow H_{dR}^1(X)$ dont on note P_ϕ le polynôme caractéristique. C'est une conséquence du théorème de pureté de Deligne qu'aucune des racines de P_ϕ n'est une racine de l'unité (voir [KM]).

Avant d'énoncer (et de démontrer) le théorème principal de Coleman, disons quelques mots sur la méthode. On se rappelle que, dans le cas où X est une variété abélienne, la stratégie de Colmez était de remarquer que puisque $m_{\{1,2\}}^* \omega - m_{\{1\}}^* \omega - m_{\{2\}}^* \omega + m_\emptyset^* \omega = dg_\omega$ où ω est une 1-forme de seconde espèce quelconque sur X et g_ω une fonction *rationnelle* sur X^3 (cf. 1.1.2.3), il est naturel de chercher une primitive localement analytique f_ω telle que $m_{\{1,2\}}^* f_\omega - m_{\{1\}}^* f_\omega - m_{\{2\}}^* f_\omega + m_\emptyset^* f_\omega$ soit une fonction *rationnelle* sur X^3 . Revenons à la situation précédente et supposons pour simplifier qu'il existe un morphisme $\phi : X^{rig} \rightarrow X^{rig}$ qui relève un morphisme de Frobenius sur Y (via la flèche de spécialisation) et induit ϕ^* sur $H_{dR}^1(X^{rig}) = H_{dR}^1(X)$. Pour toute 1-forme de seconde espèce ω sur X , vue sur X^{rig} , on a donc $P_\phi(\phi^*)(\omega) = dg_\omega$ où g_ω est une fonction *rigide* définie sur un ouvert de Zariski de X^{rig} . La stratégie de Coleman (antérieure) consiste alors à chercher une primitive localement analytique f_ω telle que $P_\phi(\phi^*)(f_\omega)$ soit une fonction *rigide*. Bien

sûr, en général, ϕ n'est défini que localement sur X^{rig} et il faut recoller...

Si U est un ouvert de Zariski dans Y , on note comme précédemment $]U[\subset X(\mathbf{C}_p)$ l'ensemble des points fermés de X qui se spécialisent sur Y en des points fermés de U . Un résultat de Coleman ([Cole3], 1.1) nous dit que si U est affine, tout morphisme de Frobenius $\phi_U : U \rightarrow U$ se relève (de façon non unique) en un morphisme d'espaces rigides $\phi_{]U[} :]U[\rightarrow]U[$. Fixons ω une 1-forme rationnelle de seconde espèce sur X . On peut montrer qu'il existe un recouvrement affine fini Y_i de Y tel que $\omega|_{]Y_i[} = \omega_i + df_i|_{]Y_i[}$ où $\omega_i \in \Omega^1_{]Y_i[/math>/ \mathbf{C}_p} est une 1-forme rigide fermée et $f_i \in \mathbf{C}_p(X)$. Faisons le choix :$

1) d'un tel recouvrement, 2) d'une écriture $\omega_i + df_i|_{]Y_i[}$ sur chaque $]Y_i[$, 3) d'un morphisme de Frobenius ϕ sur Y qui préserve les Y_i , 4) de relevés ϕ_i de ϕ sur les affinoïdes $]Y_i[$. Rappelons que $\mathcal{O}_{]Y_i[}$ désigne le faisceau structural rigide sur $]Y_i[$.

Théorème 1.2.2.1. — *Avec les notations ci-dessus, soit U_ω l'ouvert de Zariski de X complémentaire des pôles de ω ; il existe à addition près d'une constante une unique fonction f_ω localement analytique sur $U_\omega(\mathbf{C}_p)$ telle que :*

$$1) df_\omega = \omega$$

2) pour tout i , $(f_\omega - f_i)|_{]Y_i[}$ se prolonge en une fonction localement analytique sur $]Y_i[$ et $P_\phi(\phi_i^*)((f_\omega - f_i)|_{]Y_i[}) \in \Gamma(]Y_i[, \mathcal{O}_{]Y_i[})$.

De plus, f_ω est indépendante des choix 1) à 4) ci-dessus.

Preuve. — Notons pour alléger $\mathcal{O}(]Y_i[) = \Gamma(]Y_i[, \mathcal{O}_{]Y_i[})$. Le schéma de preuve est le suivant : dans une première étape, on montre que les conditions $dg_i = \omega_i$ et $P_\phi(\phi_i^*)(g_i) \in \mathcal{O}(]Y_i[)$ déterminent une unique fonction $g_i = "(f_\omega - f_i)|_{]Y_i["$ localement analytique sur $]Y_i[$ modulo une constante ; dans une seconde étape, on montre que $g_i + f_i$ et $g_j + f_j$ coïncident sur $]Y_i[\cap]Y_j[$ modulo une constante et dans la dernière étape, on montre que l'on peut ajuster les constantes pour vraiment recoller les $g_i + f_i$.

Première étape : détermination de $(f_\omega - f_i)|_{]Y_i[}$.

La flèche de functorialité $H_{dR}^1(X) \rightarrow H_{dR}^1(]Y_i[)$ envoie $P_\phi(\phi^*)(\text{classe de } \omega)$ sur la classe de $P_\phi(\phi_i^*)(\omega_i)$. Cette classe est donc nulle par choix de P_ϕ , i.e., $P_\phi(\phi_i^*)(\omega_i) \in d\mathcal{O}(]Y_i[)$. Montrons qu'il existe une unique fonction g_i localement analytique sur $]Y_i[$ (à addition près d'une constante) telle que :

$$\begin{aligned} dg_i &= \omega_i \\ P_\phi(\phi_i^*)(g_i) &\in \mathcal{O}(]Y_i[). \end{aligned}$$

Quitte à multiplier P_ϕ par une constante, on a $P_\phi(T) = a_0 + a_1T + \dots + a_{n-1}T^{n-1} + T^n \in \mathbf{C}_p[T]$ et $P_\phi(\phi_i^*)(g_i) = \sum_k a_k(g_i \circ \phi_i^k)$. Si $\Omega_i = (\omega_i, \phi_i^*\omega_i, \dots, (\phi_i^{n-1})^*\omega_i)$, on voit en posant $G_i = (g_i, g_i \circ \phi_i, \dots, g_i \circ \phi_i^{n-1})$ qu'il suffit de montrer qu'il existe une unique fonction (à constante près) localement analytique $G_i :]Y_i[\rightarrow \mathbf{C}_p^n$ telle que $dG_i = \Omega_i$ et

$G_i \circ \phi_i - MG_i \in \mathcal{O}(]Y_i])^n$ où :

$$M = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -a_0 & -a_1 & \dots & \dots & -a_{n-1} \end{pmatrix}.$$

Unicité : S'il y en a deux, leur différence D_i vérifie $dD_i = 0$, donc est localement constante, et $D_i \circ \phi_i - MD_i$ aussi. Mais $D_i \circ \phi_i - MD_i \in \mathcal{O}(]Y_i])^n$ et comme $]Y_i[$ est connexe (i.e. ne peut s'écrire $]Y_i[= U \cup U'$ où U et U' sont deux ouverts admissibles non vides d'intersection vide), $D_i \circ \phi_i - MD_i$ est constant de valeur C_i . Par ailleurs, comme 1 n'est pas racine de P_ϕ , $1 - M \in GL_n(\mathbf{C}_p)$ et un calcul donnent, pour $k \geq 1$:

$$(1) \quad D_i \circ \phi_i^k - M^k D_i = (1 - M^k)(1 - M)^{-1} C_i.$$

Soient $y_i \in Y_i(\overline{\mathbf{F}}_p)$ et $]y_i[= sp^{-1}(y_i) \subset]Y_i[$ le tube correspondant, isomorphe (en tant qu'espace rigide) à une boule ouverte (cf. 1.2.1.2). Comme y_i est défini sur une extension finie de \mathbf{F}_p , il existe $m \in \mathbf{N}$ tel que $\phi_i^m(]y_i[) \subset]y_i[$. De plus, pour tout $x \in]y_i[$, $\phi_i^{ml}(x)$ converge dans $]y_i[$ quand l tend vers $+\infty$ vers un ϵ_{y_i} indépendant de x ([Dw], lemme 3.0) et on a $\phi_i^m(\epsilon_{y_i}) = \epsilon_{y_i}$. En faisant $m = k$ dans (1) appliqué au point ϵ_{y_i} , on obtient :

$$(2) \quad D_i(\epsilon_{y_i}) = (1 - M)^{-1} C_i.$$

Par ailleurs, comme D_i est localement constante, $D_i(\phi_i^{ml_x}(x)) = D_i(\epsilon_{y_i})$ pour $l_x \gg 0$. En faisant $k = ml_x$ dans (1) appliqué au point x , on obtient en utilisant (2) :

$$(3) \quad M^{ml_x}(D_i(x) - (1 - M)^{-1} C_i) = 0.$$

Comme \mathbf{C}_p est algébriquement clos, $\phi_i^m :]y_i[\rightarrow]y_i[$ est en fait surjectif et tout $x \in]y_i[$ est dans l'image de ϕ_i^{ml} pour tout l . En faisant $k = ml$ dans (1), un calcul donne $D_i(x) - (1 - M)^{-1} C_i \in M^{ml}(\mathbf{C}_p^n)$ pour tout l . Donc $D_i(x) - (1 - M)^{-1} C_i \in \text{Ker}(M^{ml_x}) \cap \bigcap_{l \in \mathbf{N}} \text{Im}(M^{ml}) = \{0\}$ (algèbre linéaire élémentaire), ce qui montre que D_i est bien une fonction constante.

Existence : Puisque $P_\phi(\phi_i^*(\omega_i)) \in d\mathcal{O}(]Y_i])$, on a $H_i \in \mathcal{O}(]Y_i])^n$ tel que $\phi_i^* \Omega_i - M \Omega_i = dH_i$. Soient $y_i \in Y_i(\overline{\mathbf{F}}_p)$ et $\epsilon_{y_i} \in]y_i[$ comme précédemment, comme Ω_i est fermée et $]y_i[^n$ est une boule ouverte, il existe une unique fonction G_{y_i} analytique sur $]y_i[^n$ telle que $dG_{y_i} = \Omega_i|_{]y_i[}$ et $G_{y_i}(\epsilon_{y_i}) = (1 - M^m)^{-1} \sum_{l=0}^{m-1} M^l H_i(\phi_i^{m-l-1}(\epsilon_{y_i}))$ (i.e. on fixe cette valeur en ϵ_{y_i}). Soit G_i l'unique fonction localement analytique sur $]Y_i[$ telle que $G_i|_{]y_i[} = G_{y_i}$ pour $y_i \in Y_i(\overline{\mathbf{F}}_p)$

(les $]y_i[$ sont disjoints deux à deux). On a $dG_i = \Omega_i$ et un calcul montre que :

$$\begin{aligned}(G_i \circ \phi_i - MG_i)(\epsilon_{y_i}) &= H_i(\epsilon_{y_i}) \\ d(G_i \circ \phi_i - MG_i) &= dH_i.\end{aligned}$$

La première égalité ayant lieu pour tout $y_i \in Y_i(\overline{\mathbf{F}}_p)$, on en déduit $G_i \circ \phi_i - MG_i = H_i \in \mathcal{O}(]Y_i[)^n$.

Deuxième étape : Comparaison des fonctions sur $]Y_i[\cap]Y_j[$.

Montrons que pour tous i, j , $(f_i + g_i)|_{]Y_i[\cap]Y_j[} = (f_j + g_j)|_{]Y_j[\cap]Y_i[} + c_{ij}$ où $c_{ij} \in \mathbf{C}_p$. En remarquant que ϕ_i et ϕ_j laissent stable $]Y_i[\cap]Y_j[=]Y_i \times_Y Y_j[$, on voit qu'il suffit de montrer que 1) $f_i + g_i$ est indépendant (modulo cte) de l'écriture $\omega|_{]Y_i[} = \omega_i + df_i$ et 2) $f_i + g_i$ est indépendant (modulo cte) du relevé ϕ_i choisi. Pour 1), soit $\omega|_{]Y_i[} = \omega'_i + df'_i$ une autre écriture et g'_i la primitive de ω'_i construite comme dans la première étape. Alors $\omega_i - \omega'_i = d(g_i - g'_i)$ et $P_\phi(\phi_i^*)(g_i - g'_i) \in \mathcal{O}(]Y_i[)$. Mais on a aussi $\omega_i - \omega'_i = d(f'_i - f_i)$, donc $f'_i - f_i$ est défini sur tout $]Y_i[$ et $P_\phi(\phi_i^*)(f'_i - f_i) \in \mathcal{O}(]Y_i[)$, d'où $f'_i - f_i = g_i - g'_i$ modulo constante par l'unicité dans la première étape appliquée à la forme $\omega_i - \omega'_i$. Pour 2), Coleman montre ([Cole3], 1.2a) que si ϕ_i et ϕ'_i sont deux relevés du même ϕ , alors $P_\phi(\phi_i^*)(g_i) - P_{\phi'}(\phi_i'^*)(g_i) \in \mathcal{O}(]Y_i[)$ d'où $P_{\phi'}(\phi_i'^*)(g_i) \in \mathcal{O}(]Y_i[)$ ce qui entraîne 2) par unicité.

Troisième étape : recollement.

Reste donc à recoller les $f_i + g_i$, i.e. à trouver des constantes $c_i \in \mathbf{C}_p$ telles que $c_i - c_j = c_{ij}$ pour tout i, j , et on définit alors f_ω comme l'unique fonction telle que $f_\omega|_{]Y_i[} = f_i + g_i + c_i$. Mais puisque le schéma Y est connexe, $\cap_i Y_i \neq \emptyset$, donc $\cap_i]Y_i[\neq \emptyset$ et il est alors élémentaire que l'on peut ajuster les constantes c_i .

Montrons enfin que f_ω ne dépend pas du choix du morphisme ϕ sur Y . Il suffit pour cela de montrer que remplacer ϕ par ϕ^s ne change pas f_ω . Comme f_ω ne dépend pas des autres choix, on garde le même recouvrement, la même écriture et on remplace juste ϕ_i par ϕ_i^s sur $]Y_i[$. On vérifie que $P_{\phi^s}(T^s) = \prod_{\zeta^s=1} P_\phi(\zeta T)$. Donc $P_\phi(\phi_i^s)(g_i) \in \mathcal{O}(]Y_i[)$ entraîne $P_{\phi^s}((\phi_i^s)^*)(g_i) \in \mathcal{O}(]Y_i[)$ et le résultat découle encore de l'unicité de g_i dans la première étape. \square

Comme dans la première partie, on pose, si P et Q sont deux points de $X(\mathbf{C}_p)$ où ω est définie : $\int_P^Q \omega = f_\omega(Q) - f_\omega(P)$.

Corollaire 1.2.2.2. — *Avec les notations précédentes :*

- (i) $\int_P^Q (\lambda\omega + \lambda'\omega') = \lambda \int_P^Q \omega + \lambda' \int_P^Q \omega'$ ($\lambda, \lambda' \in \mathbf{C}_p$)
- (ii) $\int_P^Q \omega = f(Q) - f(P)$ si $\omega = df$ où $f \in \mathbf{C}_p(X)$
- (iii) si $g : X' \rightarrow X$ est un morphisme de variétés propres sur $\text{Spec}(\mathbf{C}_p)$ ayant bonne réduction, $\int_P^Q g^*\omega = \int_{g(P)}^{g(Q)} \omega$.

Preuve. — Les énoncés (i) et (ii) résultent de l'unicité modulo cte de f_ω . L'énoncé (iii) est facile si g provient d'un morphisme $\mathcal{X}' \rightarrow \mathcal{X}$ entre des modèles propres et lisses. Sinon, soient $Alb(\mathcal{X})$ et $Alb(\mathcal{X}')$ les modèles de Néron des variétés d'Albanese $Alb(X)$ et $Alb(X')$ associées à X et X' ([BLR]). Des propriétés du modèle de Néron on déduit un diagramme commutatif :

$$\begin{array}{ccc} X' & \xrightarrow{g} & X \\ \downarrow & & \downarrow \\ \mathcal{X}' & & \mathcal{X} \\ \downarrow & & \downarrow \\ Alb(\mathcal{X}') & \longrightarrow & Alb(\mathcal{X}) \end{array}$$

d'où (iii) en remarquant que toute 1-forme de seconde espèce sur X provient, modulo une forme exacte, d'une 1-forme de seconde espèce sur $Alb(X)$ (car $H_{dR}^1(X) \simeq H_{dR}^1(Alb(X))$). \square

Corollaire 1.2.2.3. — *L'intégrale $\int_P^Q \omega$ ne dépend pas du modèle choisi et coïncide avec l'intégrale définie par Colmez.*

Preuve. — Utiliser (ii) et (iii) dans (1.2.2.2), le fait que les variétés auxiliaires utilisées par Colmez pour définir son intégrale (i.e. $Alb(X)$ et $Alb(X)^3$) ont toutes bonne réduction lorsque X a bonne réduction et l'unicité en (1.1.2.5). \square

Remarque 1.2.2.4. — Si σ est un automorphisme continu de \mathbf{C}_p , de l'unicité de f_ω résulte aussi le fait que $\sigma(\int_P^Q \omega) = \int_{\sigma(P)}^{\sigma(Q)} \sigma^* \omega$ où $\sigma : \mathbf{C}_p \otimes_\sigma X \rightarrow X$. En particulier, si X provient d'une variété définie sur une extension finie K de \mathbf{Q}_p , si $\sigma \in Gal(\overline{\mathbf{Q}_p}/K)$, $P, Q \in X(K)$ et ω est défini sur K , on a bien $\int_P^Q \omega \in K$.

2. APPLICATIONS

2.1. L'accouplement des périodes p -adiques

Dans cette partie, K est une extension finie de \mathbf{Q}_p .

Nous allons appliquer ce qui précède, en suivant Colmez, aux périodes p -adiques des variétés abéliennes. Voir [II] et la remarque (2.1.3.4) pour l'historique de ce sujet, cas particulier des conjectures de Fontaine. C'est Coleman qui le premier a eu l'idée dans [Cole2] d'utiliser l'intégration p -adique pour redéfinir les périodes p -adiques des variétés abéliennes avec bonne réduction, ou plutôt leur image dans \mathbf{C}_p .

Comme les intégrales p -adiques ne sont pas multivaluées, on s'attendrait normalement à ce qu'il n'existe pas de périodes en p -adique ! Et effectivement, si on veut des périodes à valeurs dans \mathbf{C}_p associées aux 1-formes de $H^0(X, \Omega_{X/K}^1)$ (comme sur les complexes), on

les trouve *a posteriori* toutes nulles (cf. 2.1.2.2). Le miracle est qu'il existe néanmoins des périodes mais à un niveau "supérieur", c'est-à-dire à valeurs dans une $\overline{\mathbf{Q}}_p$ -algèbre B_{dR}^+ , introduite par Fontaine, qui se surjecte seulement sur \mathbf{C}_p .

2.1.1. *Quelques rappels sur B_{dR}^+ .* L'anneau B_{dR}^+ a déjà été introduit dans ce séminaire ([II]), mais nous aurons besoin des détails de sa construction. Les références pour ce qui suit sont [Fo1], [Fo2].

Soit R l'ensemble des suites $x = (x^{(0)}, x^{(1)}, \dots, x^{(n)}, \dots)$ d'éléments de $\mathcal{O}_{\mathbf{C}_p}$ telles que $(x^{(n+1)})^p = x^{(n)}$. On munit R des lois \cdot et $+$ en posant $x \cdot y = (x^{(n)}y^{(n)})_{n \in \mathbf{N}}$ et $x + y = (s^{(n)})_{n \in \mathbf{N}}$ où :

$$s^{(n)} = \lim_{m \rightarrow +\infty} (x^{(n+m)} + y^{(n+m)})^{p^m}$$

les hypothèses faisant que cela converge bien dans $\mathcal{O}_{\mathbf{C}_p}$. On montre que ces lois font de R un anneau commutatif intègre d'élément unité $1 = (1, 1, \dots)$. De plus :

$$p \cdot 1 = \lim_{m \rightarrow +\infty} \underbrace{(1 + \dots + 1)}_{p \text{ fois}}^{p^m} = 0$$

donc R est un anneau de caractéristique p et le Frobenius $x = (x^{(n)}) \mapsto x^p = ((x^{(n)})^p)$ sur R est clairement un isomorphisme, i.e., R est parfait. Il est de plus muni d'une action naturelle de $\text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ via l'action sur $\mathcal{O}_{\mathbf{C}_p}$ et d'une valuation $\text{val}(x) = \text{val}(x^{(0)})$ pour laquelle il est séparé et complet de corps résiduel $R/\{x | \text{val}(x) > 0\} \simeq \overline{\mathbf{F}}_p$.

Comme R est parfait, il est tentant de considérer les vecteurs de Witt $W(R)$ à coefficients dans R ([Se2]) : tout élément de $W(R)$ s'écrit de façon unique $\sum_{n=0}^{+\infty} p^n [x_n]$ où $x_n \in R$ et $[x_n]$ est son représentant multiplicatif dans $W(R)$. On montre alors qu'on a une surjection d'anneaux :

$$\begin{aligned} \theta : W(R) &\rightarrow \mathcal{O}_{\mathbf{C}_p} \\ \sum_{n=0}^{+\infty} p^n [x_n] &\mapsto \sum_{n=0}^{+\infty} p^n x_n^{(0)} \end{aligned}$$

dont le noyau est un idéal principal engendré par $p - \underline{[p]}$ où $\underline{[p]} = (p^{(n)}) \in R$ est un élément tel que $p^{(0)} = p$. On remarque que $W(R)$ est complet pour la topologie $(p, \text{Ker}(\theta))$ -adique = $(p, \underline{[p]})$ -adique.

Définition 2.1.1.1. — *L'anneau B_{dR}^+ est le complété de $W(R)[\frac{1}{p}]$ par rapport à l'idéal $\text{Ker}(\theta) = (p - \underline{[p]})$.*

La surjection θ se prolonge en une surjection encore notée $\theta : B_{dR}^+ \rightarrow \mathbf{C}_p$. L'anneau B_{dR}^+ est de valuation discrète complet d'idéal maximal $\text{Ker}(\theta) = (p - \underline{[p]})B_{dR}^+$ et de corps résiduel $B_{dR}^+/\text{Ker}(\theta) \simeq \mathbf{C}_p$, c'est-à-dire qu'il s'identifie *non canoniquement* à $\mathbf{C}_p[[T]]$. On peut

considérer plusieurs topologies sur B_{dR}^+ . Nous munissons ici B_{dR}^+ de celle pour laquelle les $p^m W(R) + (\text{Ker}(\theta))^k$ forment une base de voisinage de 0 quand m et k décrivent \mathbf{N} , c'est-à-dire que $(x_n)_n$ tend vers 0 si et seulement si, pour tout $M \in \mathbf{N}$, $x_n \in p^M W(R) + (\text{Ker}(\theta))^M$ pour $n \gg 0$. Pour cette topologie, B_{dR}^+ est séparé et complet. Il existe une action naturelle et continue de $\text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ sur B_{dR}^+ via l'action sur R , qui commute à θ . Il n'existe pas de section $\mathbf{C}_p \hookrightarrow B_{dR}^+$ de θ qui soit compatible à l'action de Galois, mais on peut montrer que $\overline{\mathbf{Q}}_p$ s'identifie canoniquement à la clôture algébrique de \mathbf{Q}_p dans B_{dR}^+ et que le diagramme :

$$\begin{array}{ccc} \overline{\mathbf{Q}}_p & \hookrightarrow & B_{dR}^+ \\ \parallel & & \downarrow \theta \\ \overline{\mathbf{Q}}_p & \hookrightarrow & \mathbf{C}_p \end{array}$$

est commutatif et compatible à $\text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$. En fait, si on munit $\overline{\mathbf{Q}}_p$ de la topologie induite par celle de B_{dR}^+ (qui n'est pas la topologie p -adique), Colmez a montré que B_{dR}^+ n'est autre que le complété de $\overline{\mathbf{Q}}_p$ pour cette topologie. Autrement dit, $\overline{\mathbf{Q}}_p$ est dense dans B_{dR}^+ (cf. appendice de [Fo2]).

2.1.2. Accouplement des périodes p -adiques : énoncé du théorème. Soit X une variété lisse sur K de dimension d . On note encore $\theta : X(B_{dR}^+) \rightarrow X(\mathbf{C}_p)$ l'application induite par $\theta : B_{dR}^+ \rightarrow \mathbf{C}_p$ et on remarque qu'elle est surjective (choisir une section $\mathbf{C}_p \hookrightarrow B_{dR}^+$ de θ). Par ailleurs, comme $\overline{\mathbf{Q}}_p \hookrightarrow B_{dR}^+$, $X(\overline{\mathbf{Q}}_p) \hookrightarrow X(B_{dR}^+)$.

Lemme 2.1.2.1. — *Soit f une fonction localement analytique sur $X(\overline{\mathbf{Q}}_p)$ qui se prolonge en une fonction localement analytique sur $X(\mathbf{Q}_p)$, alors f se prolonge aussi canoniquement à $X(B_{dR}^+)$.*

Preuve. — Soit $x \in X(B_{dR}^+)$, on peut trouver un ouvert de Zariski $U_x = \text{Spec}(\overline{\mathbf{Q}}_p[x_1, \dots, x_n, z_1, \dots, z_d]/(f_1, \dots, f_n))$ de $X \otimes_K \overline{\mathbf{Q}}_p$ tel que la projection sur z_1, \dots, z_d induit un isomorphisme entre un sous-ensemble V de $U_x(B_{dR}^+)$ contenant x et $\{(z_1, \dots, z_d) \in (B_{dR}^+)^d \mid |\theta(z_i)| < \delta, 1 \leq i \leq d\}$ pour un δ suffisamment petit dans $\mathbf{R}^+ - \{0\}$. Comme $X(\overline{\mathbf{Q}}_p) \cap V \simeq \{(z_1, \dots, z_d) \in \overline{\mathbf{Q}}_p^d \mid |z_i| < \delta, 1 \leq i \leq d\}$, et comme f se prolonge en une fonction analytique au voisinage de $\theta(x) \in X(\mathbf{C}_p)$, on peut supposer f analytique sur $X(\overline{\mathbf{Q}}_p) \cap V$ quitte à se restreindre à un ouvert de V contenant x . En remarquant que si $x \in B_{dR}^+$, $(x^n)_{n \in \mathbf{N}}$ converge dans B_{dR}^+ si et seulement si $(\theta(x)^n)_{n \in \mathbf{N}}$ converge dans \mathbf{C}_p , on voit que la série formelle en z_1, \dots, z_d définissant f sur $X(\overline{\mathbf{Q}}_p) \cap V$ converge dans B_{dR}^+ sur tout V . \square

Si ω est une 1-forme rationnelle fermée sur X et f_ω une des primitives construites en (1.1.1.5), on note encore f_ω la prolongée à $X(B_{dR}^+)$ (à valeurs maintenant dans B_{dR}^+) et $\int_P^Q \omega = f_\omega(Q) - f_\omega(P)$ si $P, Q \in X(B_{dR}^+)$. Avant d'énoncer le théorème sur l'accouplement des périodes p -adiques des variétés abéliennes, donnons un exemple simple (mais sortant *stricto sensu* du cadre des variétés abéliennes).

Exemple 2.1.2.2. — Supposons $X = \text{Spec}(\mathbf{C}[z, \frac{1}{z}]) = \mathbf{G}_m$. Soient γ un générateur de $H_1(X(\mathbf{C}), \mathbf{Z}) = H_1(\mathbf{C}^*, \mathbf{Z})$, $\omega = \frac{dz}{z} \in H^0(X, \Omega_{X/\mathbf{C}}^1)$ et $\varepsilon_n = e^{2i\pi/p^n} \in X(\mathbf{C})$. On sait que :

$$\int_{\gamma} \omega = p^n \int_1^{\varepsilon_n} \omega = p^n \int_0^{\frac{2\pi}{p^n}} \frac{de^{i\theta}}{e^{i\theta}} = 2i\pi.$$

C'est la période classique. Supposons maintenant $X = \text{Spec}(K[z, \frac{1}{z}])$ et soient $\gamma = (\varepsilon_n)_n$ un générateur de $T_p(X) = \varprojlim \mu_{p^n}(\overline{\mathbf{Q}}_p)$ et $\omega = \frac{dz}{z} \in H^0(X, \Omega_{X/K}^1)$. On a $f_\omega = \log$ et, par définition :

$$p^n \int_1^{\varepsilon_n} \omega = p^n \log(\varepsilon_n) = \log(\varepsilon_n^{p^n}) = 0 !$$

Considérons maintenant $\tilde{\varepsilon}_n = [(\varepsilon_{m+n})_{m \in \mathbf{N}}] \in X(B_{dR}^+)$ et soit $t = \log(\tilde{\varepsilon}_0) = -\sum_{n=1}^{+\infty} \frac{(1-\tilde{\varepsilon}_0)^n}{n} \in B_{dR}^+ - \{0\}$. Bien sûr, $\tilde{\varepsilon}_n \neq \varepsilon_n$ mais $\theta(\tilde{\varepsilon}_n) = \varepsilon_n$. De plus :

$$p^n \int_1^{\tilde{\varepsilon}_n} \omega = p^n \log(\tilde{\varepsilon}_n) = \log(\tilde{\varepsilon}_0) = t.$$

Plus généralement, si $x \in B_{dR}^+$, soit $m_k(x)$ le plus grand entier de \mathbf{Z} tel que $x \in p^{m_k(x)}W(R) + (\text{Ker}(\theta))^{k+1}$ et disons qu'une suite $(x_n)_n$ de points de B_{dR}^+ est bornée si la suite $(m_k(x_n))_n$ est bornée pour tout k . On peut voir que la suite $(\varepsilon_n)_n$ ci-dessus vue dans B_{dR}^+ n'est pas une suite bornée, par contre la suite $(\tilde{\varepsilon}_n)_n$ est évidemment bornée (tous les $m_k(\tilde{\varepsilon}_n)$ sont nuls). Soit $(\tilde{\varepsilon}'_n)_n$ une autre suite bornée quelconque de B_{dR}^+ vérifiant seulement $\theta(\tilde{\varepsilon}'_n) = \varepsilon_n$, alors on peut montrer que la suite $p^n \int_1^{\tilde{\varepsilon}'_n} \omega = p^n \log(\tilde{\varepsilon}'_n)$ converge toujours vers t dans B_{dR}^+ . Cet élément t , introduit par Fontaine, peut être vu comme l'analogue p -adique de $2i\pi$.

Ce qui suit va consister à transposer l'exemple ci-dessus au cadre des variétés abéliennes. Rappelons qu'une 1-forme de seconde espèce sur X est par définition une 1-forme rationnelle fermée sur X dont le résidu est nul, i.e. qui ne contient pas de termes en $\frac{df}{f}$ dans une de ses écritures, et que $H_{dR}^1(X)$ s'identifie au quotient des 1-formes de seconde espèce sur X par les différentielles des fonctions rationnelles. On suppose maintenant que X est une variété abélienne. Soient ω une 1-forme de seconde espèce sur X et $\gamma = (\gamma_n)_n \in T_p(X) = \varprojlim X[p^n](\overline{\mathbf{Q}}_p)$ où $X[p^n](\overline{\mathbf{Q}}_p) = \{x \in X(\overline{\mathbf{Q}}_p) \mid [p^n]x = 0\}$. En vertu de l'exemple précédent, on a envie de considérer $\lim_{n \rightarrow +\infty} p^n \int_0^{\tilde{\gamma}_n} \omega$ pour des "bons" relevés $\tilde{\gamma}_n$ de γ_n dans $X(B_{dR}^+)$ par θ .

Définition 2.1.2.3. — Si Y est une variété lisse sur K , une suite $(\tilde{\gamma}_n)_n$ de points de $Y(B_{dR}^+)$ est dite bornée dans $Y(B_{dR}^+)$ si pour tout recouvrement fini $(U_i)_{i \in I}$ de Y par des ouverts de Zariski, il existe une décomposition $\mathbf{N} = \cup_{i \in I} I_i$ telle que $\{\tilde{\gamma}_n, n \in I_i\} \subset U_i(B_{dR}^+)$ et telle que pour toute fonction $f_i \in \Gamma(U_i, \mathcal{O}_{U_i})$, la suite $\{f_i(\tilde{\gamma}_n), n \in I_i\}$ est bornée dans B_{dR}^+ (cf. 2.1.2.2).

Si U est un ouvert de Zariski de X , une suite bornée dans $U(B_{dR}^+)$ est bornée dans $X(B_{dR}^+)$ mais la réciproque est fautive (prendre une suite qui s'accumule autour du fermé

complémentaire de U dans X). On peut maintenant énoncer le théorème. On choisit une suite bornée $(\tilde{\gamma}_n)_n$ de relevés de $(\gamma_n)_n$ et on espère que la suite $p^n \int_0^{\tilde{\gamma}_n} \omega$ converge dans B_{dR}^+ vers une limite qui n'est pas trop souvent nulle... Mais il faut modifier 0 et $\tilde{\gamma}_n$ pour être sûr qu'ils ne s'approchent pas trop des pôles de ω dans $X(B_{dR}^+)$.

Théorème 2.1.2.4. — *Soient X une variété abélienne sur K , ω une 1-forme rationnelle de seconde espèce sur X , U_ω l'ouvert de Zariski de X complémentaire des pôles de ω et $\gamma = (\gamma_n)_n \in T_p(X)$.*

(i) *Il existe une suite $(\tilde{\gamma}_n)_n$ bornée dans $X(B_{dR}^+)$ et une suite $(\delta_n)_n$ bornée dans $U_\omega(B_{dR}^+)$ telles que $\theta(\tilde{\gamma}_n) = \gamma_n$ et la suite $(\delta_n \oplus \tilde{\gamma}_n)_n$ est bornée dans $U_\omega(B_{dR}^+)$.*

(ii) *Si $(\tilde{\gamma}_n)_n$ et $(\delta_n)_n$ sont de telles suites, la suite $p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega$ converge dans B_{dR}^+ vers un élément $\int_\gamma \omega$ qui ne dépend que de γ et de la classe de ω dans $H_{dR}^1(X)$.*

(iii) *L'application de $H_{dR}^1(X) \times T_p(X)$ dans B_{dR}^+ qui à (ω, γ) associe $\int_\gamma \omega$ est bilinéaire, commute à l'action de Galois et envoie $H^0(X, \Omega_{X/K}^1) \times T_p(X)$ dans $(p - [p])B_{dR}^+ = \text{Ker}(\theta)$.*

(iv) *L'accouplement "période" ci-dessus est non dégénéré, i.e. si $(\omega_1, \dots, \omega_{2d})$ fournit une base de $H_{dR}^1(X)$ et $(\gamma_1, \dots, \gamma_{2d})$ une base de $T_p(X)$, la matrice $(\int_{\gamma_j} \omega_i)_{1 \leq i, j \leq 2d}$ appartient à $GL_{2d}(B_{dR})$ où $B_{dR} = \text{Frac}(B_{dR}^+)$.*

2.1.3. *Esquisse de preuve et relations de Riemann p -adiques.* Nous ne disons rien sur la preuve de (i) qui est purement technique (voir [Colm2], II.3.1).

Passons à (ii). Remarquons d'abord que si $(x_n)_n$ est une suite de B_{dR}^+ telle que $(px_n - x_{n-1})_n$ est bornée, alors $p^{n-1}(px_n - x_{n-1})$ tend vers 0 dans B_{dR}^+ quand $n \rightarrow +\infty$ et donc $p^n x_n$ converge dans B_{dR}^+ . Oublions dans un premier temps les δ_n et supposons $[p]\tilde{\gamma}_n = \tilde{\gamma}_{n-1}$. Pour la convergence, il suffit de voir que la suite $y_n = p \int_0^{\tilde{\gamma}_n} \omega - \int_0^{\tilde{\gamma}_{n-1}} \omega$ est bornée dans B_{dR}^+ . Mais $y_n = (pf_\omega - [p]^* f_\omega)(\tilde{\gamma}_n) + \text{cte}$ et $pf_\omega - [p]^* f_\omega \in K(X)$ (car $[p]$ sur X induit la multiplication par p sur $H_{dR}^1(X)$). Si la suite $\tilde{\gamma}_n$ est bornée dans l'ouvert où $pf_\omega - [p]^* f_\omega$ est défini, alors par définition, $(y_n)_n$ est bornée dans B_{dR}^+ . Si elle n'est pas bornée dans cet ouvert, il faut la modifier avec des δ'_n pour qu'elle le devienne. Dans le cas général, on décompose $y_n = p \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega - \int_{\delta_{n-1}}^{\delta_{n-1} \oplus \tilde{\gamma}_{n-1}} \omega$ sous la forme $y_n = y_{1,n} + y_{2,n} + y_{3,n} + y_{4,n}$ où :

$$\begin{aligned} y_{1,n} &= p \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega - p \int_{\delta'_n}^{\delta'_n \oplus \tilde{\gamma}_n} \omega = p(\Delta^{[2]} f_\omega)(\delta'_n, \delta_n \ominus \delta'_n, \tilde{\gamma}_n) \\ y_{2,n} &= p \int_{\delta'_n \oplus \tilde{\gamma}_n}^{\delta'_n \oplus \tilde{\gamma}_n} \omega - \int_{[p]\delta'_n \oplus [p]\tilde{\gamma}_n}^{\delta'_n \oplus \tilde{\gamma}_n} \omega = (pf_\omega - [p]^* f_\omega)(\delta'_n \oplus \tilde{\gamma}_n) - (pf_\omega - [p]^* f_\omega)(\tilde{\gamma}_n) \\ y_{3,n} &= \int_{[p]\delta'_n \oplus [p]\tilde{\gamma}_n}^{\delta'_n \oplus \tilde{\gamma}_n} \omega - \int_{\delta_{n-1}}^{\delta_{n-1} \oplus [p]\tilde{\gamma}_n} \omega = (\Delta^{[2]} f_\omega)(\delta_{n-1}, [p]\delta'_n \ominus \delta_{n-1}, [p]\tilde{\gamma}_n) \\ y_{4,n} &= \int_{\delta_{n-1}}^{\delta_{n-1} \oplus [p]\tilde{\gamma}_n} \omega - \int_{\delta_{n-1}}^{\delta_{n-1} \oplus \tilde{\gamma}_{n-1}} \omega = f_\omega(\delta_{n-1} \oplus [p]\tilde{\gamma}_n) - f_\omega(\delta_{n-1} \oplus \tilde{\gamma}_{n-1}) \end{aligned}$$

et où δ'_n est une suite bornée dans $U_\omega(B_{dR}^+) \cap ([p]^* U_\omega)(B_{dR}^+)$ telle que $\delta'_n \oplus \tilde{\gamma}_n$ est bornée dans ce même ouvert (il en existe). On peut déduire que les suites $(y_{1,n})_n$, $(y_{2,n})_n$ et $(y_{3,n})_n$

sont bornées du fait que les fonctions $\Delta^{[2]}f_\omega$ et $pf_\omega - [p]^*f_\omega$ sont rationnelles. Quant à $(y_{4,n})_n$, on montre qu'elle est bornée en utilisant $\theta(\delta_{n-1} \oplus [p]\tilde{\gamma}_n) = \theta(\delta_{n-1} \oplus \tilde{\gamma}_{n-1})$. Donc $p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega$ converge dans B_{dR}^+ . Si $(\delta_n)_n$ est remplacé par $(\delta''_n)_n$ satisfaisant les propriétés (i) en (1.1.1.5) (sans changer $\tilde{\gamma}_n$), alors la suite $(\delta'''_n)_n$ définie par $\delta'''_n = \delta_n$ si n est pair et $\delta'''_n = \delta''_n$ si n est impair vérifie encore ces propriétés, d'où on déduit que la limite est indépendante de $(\delta_n)_n$. Si on choisit un autre relevé borné $(\tilde{\gamma}'_n)_n$, la suite $(\delta_n \oplus \tilde{\gamma}'_n)_n$ est encore bornée dans $U_\omega(B_{dR}^+)$ ([Colm2],II.3.1) et le même raisonnement en "mélangeant" les suites $(\tilde{\gamma}_n)_n$ et $(\tilde{\gamma}'_n)_n$ montre que la limite est indépendante du relevé borné de $(\gamma_n)_n$ choisi. Enfin, si $\omega = df_\omega$ avec $f_\omega \in K(X)$, la suite $\int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega = f_\omega(\delta_n \oplus \tilde{\gamma}_n) - f_\omega(\delta_n)$ est par définition bornée ce qui entraîne que $p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega$ converge vers 0 dans B_{dR}^+ . Ceci achève (ii).

Passons au (iii). La linéarité par rapport à ω vient de la linéarité de l'intégrale. Si $((\gamma_n)_n, (\gamma'_n)_n) \in T_p(X) \times T_p(X)$, on prend $\tilde{\gamma}_n \oplus \tilde{\gamma}'_n$ comme relevé de $\gamma_n \oplus \gamma'_n$, qui forme encore une suite bornée dans $X(B_{dR}^+)$. On peut trouver $(\delta_n)_n$ bornée dans $U_\omega(B_{dR}^+)$ telle que les trois suites $(\delta_n \oplus \tilde{\gamma}_n)_n$, $(\delta_n \oplus \tilde{\gamma}'_n)_n$ et $(\delta_n \oplus \tilde{\gamma}_n \oplus \tilde{\gamma}'_n)_n$ y sont encore bornées. On écrit :

$$p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n \oplus \tilde{\gamma}'_n} \omega = p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega + p^n \int_{\delta_n \oplus \tilde{\gamma}_n}^{\delta_n \oplus \tilde{\gamma}_n \oplus \tilde{\gamma}'_n} \omega$$

et on remarque que la suite $(\delta_n \oplus \tilde{\gamma}_n)_n$ vérifie les propriétés (i) par rapport à $(\tilde{\gamma}'_n)_n$. Par (ii), cela entraîne que $p^n \int_{\delta_n \oplus \tilde{\gamma}_n}^{\delta_n \oplus \tilde{\gamma}_n \oplus \tilde{\gamma}'_n} \omega$ converge dans B_{dR}^+ vers $\int_{\gamma'} \omega$, d'où la linéarité par rapport à γ . Enfin, si $\omega \in H^0(X, \Omega_{X/K}^1)$, la primitive f_ω en (1.1.1.5) nulle en 0 est un homomorphisme de groupes (cf. 1.1.2.8) donc $p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega = p^n f_\omega(\tilde{\gamma}_n) = f_\omega([p]^n \tilde{\gamma}_n)$. Or $\theta(f_\omega([p]^n \tilde{\gamma}_n)) = f_\omega([p]^n \gamma_n) = f_\omega(0) = 0$, d'où $p^n \int_{\delta_n}^{\delta_n \oplus \tilde{\gamma}_n} \omega \in \text{Ker}(\theta)$.

Le plus délicat est (iv), qui découle de l'analogie p -adique des relations de Riemann entre les périodes. Rappelons que si U est un ouvert de Zariski de X , on a un isomorphisme $\Gamma(U, \mathcal{O}_U) \otimes_K H^0(X, \Omega_{X/K}^1) \xrightarrow{\sim} H^0(U, \Omega_{U/K}^1)$ (car X est abélienne). Soit $(\omega_1, \dots, \omega_d)$ une base sur K de $H^0(X, \Omega_{X/K}^1)$, en passant à $U^{an} = U(K)$, on voit que si G est une fonction localement analytique sur $U(K)$, on a $dG = \sum_{i=1}^d G_i \omega_i$ où les G_i sont localement analytiques sur $U(K)$.

Proposition 2.1.3.1. — ([Colm2],II.1.19) *Soit D un diviseur de X . Il existe une fonction G_D localement analytique sur $X(K) - D(K)$ et ayant une singularité logarithmique le long de D telle que $\Delta^*G_D = G_D(x \oplus y \oplus z) - G_D(x \oplus y) - G_D(x \oplus z) - G_D(y \oplus z) + G_D(x) + G_D(y) + G_D(z)$ soit le logarithme d'une fonction rationnelle sur X^3 . De plus, si $dG_D = \sum_{i=1}^d G_{D,i} \omega_i$, $\omega_{D,i} = dG_{D,i}$ est une 1-forme de seconde espèce sur X pour tout i .*

À la base de cette proposition est le théorème du cube qui dit que l'image inverse Δ^*D de D sur X^3 par $\Delta : X^3 \rightarrow X$ est le diviseur d'une fonction rationnelle. D'autre part, rappelons qu'à D est associé l'accouplement de Weil $\varepsilon_D : T_p(X) \times T_p(X) \rightarrow T_p(\mathbf{G}_m)$ (cf.

([Mu1],IV.20) par exemple). Si $(\gamma, \gamma') \in T_p(X) \times T_p(X)$, $\log[\varepsilon_D(\gamma, \gamma')] \in \mathbf{Z}_p t$ (voir 2.1.2.2 pour t) et l'accouplement :

$$\log[\varepsilon_D] : T_p(X) \times T_p(X) \rightarrow \mathbf{Z}_p t$$

est bilinéaire et antisymétrique.

Théorème 2.1.3.2. — (*Relations de Riemann p -adiques*) Soit $(\gamma, \gamma') \in T_p(X) \times T_p(X)$, alors :

$$\log[\varepsilon_D(\gamma, \gamma')] = \sum_{i=1}^d \left(\int_{\gamma} \omega_i \int_{\gamma'} \omega_{D,i} - \int_{\gamma'} \omega_i \int_{\gamma} \omega_{D,i} \right).$$

Voir ([Colm2],II.3.5) pour la preuve. En conséquence, on voit que si $\gamma_1, \dots, \gamma_{2d}$ est une base de $T_p(X)$ et si on note $\omega_{d+i} = \omega_{D,i}$ ($i \in \{1, \dots, d\}$), on a l'égalité matricielle :

$$\left(\log[\varepsilon_D(\gamma_i, \gamma_j)] \right)_{1 \leq i, j \leq 2d} = {}^t \left(\int_{\gamma_j} \omega_i \right)_{1 \leq i, j \leq 2d} \cdot \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix} \cdot \left(\int_{\gamma_j} \omega_i \right)_{1 \leq i, j \leq 2d}$$

où i est le numéro de la ligne et j celui de la colonne. Comme on peut toujours choisir D tel que l'accouplement de Weil soit non dégénéré, i.e. tel que :

$$\det \left(\log[\varepsilon_D(\gamma_i, \gamma_j)] \right)_{1 \leq i, j \leq 2d} \in \mathbf{Q}_p^\times t \subset B_{dR}^\times$$

cela montre que, pour un tel D , $(\omega_1, \dots, \omega_{2d})$ fournit une base de $H_{dR}^1(X)$ et $\left(\int_{\gamma_j} \omega_i \right)_{1 \leq i, j \leq 2d} \in GL_{2d}(B_{dR})$.

Remarque 2.1.3.3. — Les fonctions G_D méritent le nom de fonctions de Green p -adiques. Les 1-formes $\omega_{D,i}$ sont déjà définies dans [Cole2]. De (2.1.2.4), on peut déduire des isomorphismes $B_{dR} \otimes_K H_{dR}^i(X) \xrightarrow{\sim} B_{dR} \otimes_{\mathbf{Q}_p} H_{\text{ét}}^i(X_{\overline{K}}, \mathbf{Q}_p)$ pour tout i (où $H^i(X) = \bigwedge^i H^1(X)$) et les relations de Riemann p -adiques sont une autre façon de dire que pour $i = 2$ l'isomorphisme envoie la classe du diviseur D dans $H_{dR}^2(X)$ sur sa classe dans $H_{\text{ét}}^2(X_{\overline{K}}, \mathbf{Q}_p)$.

Remarque 2.1.3.4. — Les premiers théorèmes de comparaison entre $H_{dR}^1(X)$ et $H_{\text{ét}}^1(X_{\overline{K}}, \mathbf{Q}_p)$ pour une variété abélienne X sur K sont dus à Tate dans le cas de bonne réduction ([Ta1]) et Raynaud dans le cas général ([Bog]), qui démontrent la célèbre décomposition de Hodge-Tate :

$$\mathbf{C}_p \otimes_{\mathbf{Q}_p} H_{\text{ét}}^1(X_{\overline{K}}, \mathbf{Q}_p) \simeq \left(\mathbf{C}_p(-1) \otimes_K H^0(X, \Omega_{X/K}^1) \right) \oplus \left(\mathbf{C}_p \otimes_K H^1(X, \mathcal{O}_X) \right).$$

Cette décomposition a ensuite été redémontrée par Fontaine dans le cas général ([Fo3]), puis par Coleman dans le cas de bonne réduction ([Cole2]) via son intégration p -adique exposée au (1.2). L'énoncé avec B_{dR}^+ comme en (2.1.2.4) est plus fort (la décomposition ci-dessus s'en déduit facilement) et a d'abord été obtenu par Fontaine et Messing (cf. [Wi]). C'est un cas particulier de conjectures en tout degré de cohomologie dues à Fontaine (appendice de [Fo1]) qui sont maintenant démontrées en toute généralité (travaux

de Fontaine-Messing, Faltings, Kato, Tsuji, etc...) : voir l'exposé d'Illusie [Il] dans ce séminaire et [Ts] pour plus de détails. Mais les preuves, où il n'est pas du tout question d'intégrer des formes différentielles, sont souvent longues et techniques. La preuve de Colmez pour le H^1 , fidèle à la stratégie classique sur les complexes, n'en est que plus élégante.

2.2. Les polylogarithmes p -adiques

Dans cette partie, $K = \mathbf{C}_p$.

Dans le cas des courbes avec bonne réduction, Coleman a développé une variante de sa construction en (1.2) (les “basic wide open spaces” en anglais) qui lui permet d'intégrer toutes les formes différentielles rationnelles, mais aussi les formes rigides et même, par itération, des formes localement analytiques pas forcément algébriques ou rigides, ce que ne permet pas l'intégrale de Colmez. Nous présentons ci-dessous un cas particulier de cette construction que nous utilisons ensuite pour définir les polylogarithmes p -adiques.

2.2.1. Itération de l'intégration sur les courbes. Soient donc X une courbe propre et lisse sur $\text{Spec}(\mathbf{C}_p)$ qui admet un modèle \mathcal{X} propre et lisse sur $\text{Spec}(\mathcal{O}_{\mathbf{C}_p})$ et Y la fibre spéciale sur $\text{Spec}(\overline{\mathbf{F}}_p)$.

Soient $S \subset X(\mathbf{C}_p)$ un ensemble fini de points fermés, $U = X - S$ et \overline{S} l'image de S dans $Y(\overline{\mathbf{F}}_p)$. Soient $U^{rig} = U(\mathbf{C}_p)$ (= espace analytique rigide associé à U) et $W = sp^{-1}(Y(\overline{\mathbf{F}}_p) - \overline{S}) \subset U^{rig}$ où sp est la flèche de spécialisation. On suppose que l'application $S \rightarrow \overline{S}$ est injective. L'ouvert U^{rig} est alors un cas particulier de ce que Coleman appelle un “basic wide open space” ([CS],2.1) et W est un “affinoïde sous-jacent” (cf. 1.2.1.2). On a dans ce cas :

$$U^{rig} - W = \cup_{s \in S} C_s$$

où C_s est une couronne ouverte autour de $s \in S$ isomorphe à $\{z \in \mathbf{C}_p \mid 0 < |z| < 1\}$. Via cet isomorphisme, on pose pour $r \in]0, 1[$:

$$U_r = W \cup \left(\bigcup_{s \in S} C_{s,r} \right) \subset U$$

où $C_{s,r} \subset C_s$ correspond à $\{z \in \mathbf{C}_p \mid r < |z| < 1\} \subset \{z \in \mathbf{C}_p \mid 0 < |z| < 1\}$. Quand r tend vers 1, les U_r forment une famille décroissante de voisinages de W . Ils dépendent des isomorphismes choisis, mais si $(U'_{r'})_{r'}$ est une famille obtenue à partir d'autres choix de coordonnées locales sur les C_s , et si $r' \in]0, 1[$, il existe r suffisamment proche de 1 tel que $U_r \subset U_{r'}$, et réciproquement. Cela suffit pour que les résultats ci-dessous soient indépendants de la famille $(U_r)_r$.

Fixons un morphisme de Frobenius sur Y (cf. 1.2.2) qui envoie \overline{S} dans \overline{S} .

Proposition 2.2.1.1. — (i) Pour r suffisamment proche de 1, il existe un morphisme rigide $\phi : U_r \rightarrow U^{rig}$ tel que $\phi(W) \subset W$, $\phi(C_s \cap U_r) \subset C_s$ ($s \in S$) et qui commute via la flèche de spécialisation avec le morphisme de Frobenius fixé sur $Y(\overline{\mathbf{F}}_p)$.

(ii) Si N est un entier fixé et $n \in \{0, \dots, N\}$, il existe r suffisamment proche de 1 tel que $\phi^n = \phi \circ \dots \circ \phi$ (n fois) est encore bien défini de U_r dans U^{rig} .

Voir ([CS], 2.2) pour la preuve. Si n est un entier fixé et si $1 - r$ est suffisamment petit, on peut donc définir par functorialité : $(\phi^*)^n = (\phi^n)^* : \Omega_{U^{rig}/\mathbf{C}_p}^1 \rightarrow \Omega_{U_r/\mathbf{C}_p}^1$ où $\Omega_{U^{rig}/\mathbf{C}_p}^1$ désigne les 1-formes (rigides) sur U^{rig} (resp. avec U_r).

Proposition 2.2.1.2. — Il existe un polynôme $P_1(T) \in \mathbf{C}_p[T]$ dont aucune des racines n'est une racine de l'unité tel que, pour tout $\omega \in \Omega_{U^{rig}/\mathbf{C}_p}^1$, tout ϕ comme en (2.2.1.1) et tout r suffisamment proche de 1, $P_1(\phi^*)(\omega) \in d\Gamma(U_r, \mathcal{O}_{U_r})$.

Indication de preuve. — L'inclusion d'espaces rigides $U_r \hookrightarrow U^{rig}$ induit un isomorphisme $H_{dR}^1(U^{rig}) \xrightarrow{\sim} H_{dR}^1(U_r)$ où $H_{dR}^1(U^{rig}) = \Omega_{U^{rig}/\mathbf{C}_p}^1/d\Gamma(U^{rig}, \mathcal{O}_{U^{rig}})$ (resp. avec U_r). Le morphisme $\phi : U_r \rightarrow U^{rig}$ donne donc un morphisme $\phi^* : H_{dR}^1(U^{rig}) \rightarrow H_{dR}^1(U_r)$. Le résultat vient du fait que le polynôme caractéristique de ϕ^* s'identifie au numérateur de la fonction ζ de la courbe affine $Y - \overline{S}$, dont les racines sont des nombres de Weil, i.e. de valeur absolue complexe (pour tout plongement dans \mathbf{C}) une puissance non nulle de p . \square

Pour tout $s \in S$, soit $\Gamma(C_s, \mathcal{O}_{C_s})_{log} = \Gamma(C_s, \mathcal{O}_{C_s})[\log(f), f \in \Gamma(C_s, \mathcal{O}_{C_s})^\times]$.

Théorème 2.2.1.3. — Soient ϕ , P_1 et r comme ci-dessus et ω dans $\Omega_{U^{rig}/\mathbf{C}_p}^1$, il existe à addition près d'une constante une unique fonction f_ω localement analytique sur $U^{rig} = U(\mathbf{C}_p)$ telle que :

- 1) $df_\omega = \omega$
- 2) $P_1(\phi^*)(f_\omega) \in \Gamma(U_r, \mathcal{O}_{U_r})$
- 3) pour tout $s \in S$, $f_\omega|_{C_s} \in \Gamma(C_s, \mathcal{O}_{C_s})_{log}$.

De plus f_ω est indépendant de r , ϕ , P_1 et du morphisme de Frobenius sur Y .

Preuve. — Soit $g \in \Gamma(U_r, \mathcal{O}_{U_r})$ tel que $P_1(\phi^*)(\omega) = dg$. En procédant comme dans la première étape de (1.2.2.1), mais avec l'affinoïde W au lieu de l'affinoïde $]Y_i[$, on a une unique (à cte près) fonction f_W localement analytique sur W telle que $df_W = \omega|_W$ et $P_1(\phi^*)(f_W) = g|_W \in \Gamma(W, \mathcal{O}_W)$. Sur C_s , il est facile de voir que tout 1-forme de $\Omega_{C_s/\mathbf{C}_p}^1$ s'intègre de façon unique (mod. cte) dans $\Gamma(C_s, \mathcal{O}_{C_s})_{log}$: soit f_s une telle primitive de $\omega|_{C_s}$. Soit f l'unique fonction localement analytique sur U^{rig} telle que $f|_W = f_W$ et $f|_{C_s} = f_s$ pour tout $s \in S$: il faut ajuster les constantes d'intégration pour avoir $P_1(\phi^*)(f) \in \Gamma(U_r, \mathcal{O}_{U_r})$. Soit $c_s = (P_1(\phi^*)(f_s) - g)|_{C_s \cap U_r} \in \Gamma(C_s \cap U_r, \mathcal{O}_{C_s \cap U_r})_{log}$ (on utilise ici $\phi(C_s \cap U_r) \subset C_s$), on a $dc_s = 0$ sur $C_s \cap U_r = C_{s,r}$ qui est connexe, donc c_s est constant. Soit f_ω tel que $f_\omega|_W = f_W$ et $f_\omega|_{C_s} = f_s - \frac{c_s}{P_1(1)}$: on vérifie que $P_1(\phi^*)(f_\omega) = g \in \Gamma(U_r, \mathcal{O}_{U_r})$, et que

g détermine f_ω de façon unique. On renvoie à ([CS],2.3,2.4) pour l'indépendance par rapport à P_1 et au relevé ϕ (même méthode pour ce dernier qu'en (1.2.2.1)). L'indépendance par rapport au morphisme de Frobenius sur Y se prouve comme dans (1.2.2.1). \square

Remarquons que si ω est une forme rationnelle quelconque sur X (i.e. pas forcément de seconde espèce), $S \subset X(\mathbf{C}_p)$ l'ensemble des points où ω a un pôle, et si *de plus* l'application de spécialisation est injective sur S , (2.2.1.3) fournit une primitive de ω bien définie modulo constante. On peut montrer que cette primitive coïncide avec celle en (1.1.1.5), en particulier $f_\omega = \log(f)$ si $\omega = \frac{df}{f}$ avec f non nul sur U ([CS],2.5.1).

Nous allons itérer ce processus. Notons $\mathcal{O}_0(U^{rig}) = \Gamma(U^{rig}, \mathcal{O}_{U^{rig}})$ et $\mathcal{O}_0(U_r) = \Gamma(U_r, \mathcal{O}_{U_r})$ pour $0 < r < 1$. On définit les sous-espaces suivants des fonctions localement analytiques sur U^{rig} ou U_r :

$$\begin{aligned} \mathcal{O}_1(U^{rig}) &= \mathcal{O}_0(U^{rig}) + \sum_{\omega \in \Omega_{U^{rig}/\mathbf{C}_p}^1} \mathcal{O}_0(U^{rig})f_\omega \\ \mathcal{O}_1(U_r) &= \mathcal{O}_0(U_r) + \sum_{\omega \in \Omega_{U^{rig}/\mathbf{C}_p}^1} \mathcal{O}_0(U_r)f_\omega|_{U_r} \end{aligned}$$

où f_ω est "la" primitive en (2.2.1.3) et r est suffisamment proche de 1.

Proposition 2.2.1.4. — (i) Soit $f \in \mathcal{O}_1(U_r)$ (resp. $\mathcal{O}_1(U^{rig})$) tel que $df = 0$, alors f est constante sur U_r (resp. U^{rig}).

(ii) Il existe un polynôme $P_2(T) \in \mathbf{C}_p[T]$ dont aucune des racines n'est une racine de l'unité tel que, pour tout $\omega \in \mathcal{O}_1(U^{rig}) \otimes_{\mathcal{O}_0(U^{rig})} \Omega_{U^{rig}/\mathbf{C}_p}^1$, tout ϕ comme en (2.2.1.1) et tout r suffisamment proche de 1, $P_2(\phi^*)(\omega) \in d\mathcal{O}_1(U_r)$.

Indication de preuve. — On renvoie à ([CS],2.4.4) pour (i). Pour (ii), soit $H_{dR}^1(\mathcal{O}_1(U^{rig})) = (\mathcal{O}_1(U^{rig}) \otimes \Omega_{U^{rig}/\mathbf{C}_p}^1)/d\mathcal{O}_1(U^{rig})$, l'application :

$$\begin{aligned} \Omega_{U^{rig}/\mathbf{C}_p}^1 \times \Omega_{U^{rig}/\mathbf{C}_p}^1 &\rightarrow H_{dR}^1(\mathcal{O}_1(U^{rig})) \\ (\omega, \omega') &\mapsto f_\omega \omega' \text{ mod. } d\mathcal{O}_1(U^{rig}) \end{aligned}$$

est bien définie et Coleman et de Shalit montrent ([CS],2.4.6) qu'elle induit un isomorphisme compatible à ϕ^* :

$$H_{dR}^1(U^{rig}) \otimes_{\mathbf{C}_p} H_{dR}^1(U^{rig}) \xrightarrow{\sim} H_{dR}^1(\mathcal{O}_1(U^{rig})).$$

Les valeurs propres de ϕ^* sur $H_{dR}^1(\mathcal{O}_1(U^{rig}))$ sont donc des produits de nombres de Weil (cf. preuve de (2.2.1.2)) : ce ne sont en particulier jamais des racines de l'unité et on peut prendre pour P_2 le polynôme caractéristique de ϕ^* sur $H_{dR}^1(\mathcal{O}_1(U^{rig}))$. \square

Il est alors formel à partir de (2.2.1.4) de généraliser la preuve de (2.2.1.3) (et ce qu'elle utilise de (1.2.2.1)) au cas où $\omega \in \mathcal{O}_1(U^{rig}) \otimes \Omega_{U^{rig}/\mathbf{C}_p}^1$ en remplaçant $P_1(\phi^*)$ par $P_2(\phi^*)$

et $\Gamma(U_r, \mathcal{O}_{U_r}) = \mathcal{O}_0(U_r)$ par $\mathcal{O}_1(U_r)$. On définit de manière analogue $\mathcal{O}_2(U^{rig})$ comme le $\mathcal{O}_1(U^{rig})$ -module engendré par les constantes et les primitives qu'on vient d'obtenir des 1-formes de $\mathcal{O}_1(U^{rig}) \otimes \Omega_{U^{rig}/\mathbf{C}_p}^1$, et $\mathcal{O}_2(U_r)$ comme le $\mathcal{O}_1(U_r)$ -module engendré par les constantes et les restrictions à U_r de ces primitives. La proposition (2.2.1.4) est encore valable avec $\mathcal{O}_2(U_r)$ ou $\mathcal{O}_2(U^{rig})$ et ainsi de suite... Par récurrence, on obtient finalement des suites croissantes canoniques d'espaces de fonctions localement analytiques : $\mathcal{O}_k(U^{rig}) \subset \mathcal{O}_{k+1}(U^{rig})$ et $\mathcal{O}_k(U_r) \subset \mathcal{O}_{k+1}(U_r)$ et des polynômes $P_k(T)$ (non uniques mais dont les racines ne sont pas des racines de l'unité) tels qu'on ait, si r est suffisamment proche de 1 :

Théorème 2.2.1.5. — *Soient $k \in \mathbf{N}$ et $\omega \in \mathcal{O}_k(U^{rig}) \otimes_{\mathcal{O}_0(U^{rig})} \Omega_{U^{rig}/\mathbf{C}_p}^1$, il existe à addition près d'une constante une unique fonction $f_\omega \in \mathcal{O}_{k+1}(U^{rig})$ telle que :*

- 1) $df_\omega = \omega$
- 2) $P_{k+1}(\phi^*)(f_\omega) \in \mathcal{O}_k(U_r)$
- 3) pour tout $s \in S$, $f_\omega|_{C_s} \in \Gamma(C_s, \mathcal{O}_{C_s})_{log}$.

Bien sûr, la récurrence ci-dessus peut se formaliser et se généraliser en dégageant les propriétés des $\mathcal{O}_k(U^{rig})$: c'est ce qui est fait dans [Cole1] et [CS] où sont introduits des “ F -cristaux logarithmiques”.

2.2.2. *Construction des polylogarithmes p -adiques.* Pour une introduction aux polylogarithmes complexes, on renvoie à l'exposé d'Oesterlé dans ce séminaire ([Oe]). Soit k un entier, rappelons que les polylogarithmes classiques $\ell_k(z)$ (plus exactement leur détermination principale) sont les fonctions holomorphes sur $\mathbf{C} - [1, +\infty[$ données sur la boule unité ouverte par la formule :

$$\ell_k(z) = \sum_{n=1}^{+\infty} \frac{z^n}{n^k}.$$

On a $\ell_0(z) = \frac{z}{z-1}$, $\ell_1(z) = -\log(1-z)$ et $d\ell_k(z) = \ell_{k-1}(z) \frac{dz}{z}$. Nous allons définir leurs analogues p -adiques, que nous noterons encore ℓ_k , vu qu'il ne sera (presque) plus question des polylogarithmes complexes dans la suite.

Avec les notations de la partie précédente, considérons :

$$\begin{aligned} X &= \mathbf{P}^1 \\ S &= \{1, \infty\} \subset \mathbf{P}^1(\mathbf{C}_p). \end{aligned}$$

On a :

$$\begin{aligned}
 U^{rig} &= X(\mathbf{C}_p) - S = \mathbf{C}_p - \{1\} \\
 W &= \{z \in \mathbf{C}_p \mid |z| \leq 1, |z-1| \geq 1\} \\
 C_1 &= \{z \in \mathbf{C}_p \mid 0 < |z-1| < 1\} \\
 C_\infty &= \{z \in \mathbf{C}_p \mid |z| > 1\} \\
 U_r &= \{z \in \mathbf{C}_p \mid |z| < \frac{1}{r}, |z-1| > r\} \quad (0 < r < 1).
 \end{aligned}$$

Remarquons que $\mathcal{O}_0(U^{rig}) = \mathcal{O}_0(\mathbf{C}_p - \{1\})$ est l'ensemble des séries de Laurent $\sum_{n=-\infty}^{+\infty} a_n(z-1)^n$ qui convergent sur $\mathbf{C}_p - \{1\}$ et $\Omega_{U^{rig}/\mathbf{C}_p}^1 = \mathcal{O}_0(U^{rig})dz$. On choisit comme morphisme de Frobenius $\phi(z) = z^p$ qui préserve bien S et envoie U_r dans U^{rig} et $U_r \cap C_s$ dans C_s ($s \in S$) lorsque $\frac{1}{p^{p-1}} < r < 1$ (car $\{\zeta \mid \zeta^p = 1\} \cap U_r = \emptyset$ pour de tels r). D'après le paragraphe précédent, on dispose des espaces de fonctions localement analytiques $\mathcal{O}_k(\mathbf{C}_p - \{1\})$ pour $k \in \mathbf{N}$.

Corollaire 2.2.2.1. — *Il existe une unique suite de fonctions $\ell_k(z) \in \mathcal{O}_k(\mathbf{C}_p - \{1\})$ telle que :*

$$\begin{aligned}
 \ell_0(z) &= \frac{z}{z-1} \\
 \ell_k(0) &= 0 \\
 d\ell_k(z) &= \frac{\ell_{k-1}(z)}{z} dz.
 \end{aligned}$$

De plus, $\ell_1(z) = -\log(1-z)$ et, si $|z| < 1$, $\ell_k(z) = \sum_{n=1}^{+\infty} \frac{z^n}{n^k}$.

Preuve. — L'existence et l'unicité sont conséquences de (2.2.1.5). Un examen des preuves de (1.2.2.1) et (2.2.1.3) montre que les fonctions de $\mathcal{O}_k(\mathbf{C}_p - \{1\})$ sont toutes analytiques sur la boule unité ouverte $\{z \in \mathbf{C}_p \mid |z| < 1\} \subset W$. Or, la fonction $\sum_{n=1}^{+\infty} \frac{z^n}{n^k}$ est aussi analytique sur cette boule et satisfait les propriétés de l'énoncé. On a donc $\ell_k(z) = \sum_{n=1}^{+\infty} \frac{z^n}{n^k}$ pour $|z| < 1$ par unicité (mod. cte) de la primitive analytique des 1-formes analytiques sur la boule ouverte. Vérifions que $P_1(T) = p - T$ satisfait $P_1(\phi^*)(\omega) \in d\mathcal{O}_0(U_r)$ pour tout $\omega \in \mathcal{O}_0(U^{rig})dz$ et tout $r \in]\frac{1}{p^{p-1}}, 1[$. On voit qu'il suffit de vérifier $P_1(\phi^*)(\frac{dz}{z-1}) \in d\mathcal{O}_0(U_r)$. Mais on a :

$$P_1(\phi^*)\left(\frac{dz}{z-1}\right) = p\left(\frac{1}{z-1} - \frac{z^{p-1}}{z^p-1} \text{Big}\right)dz$$

et un calcul montre que cette 1-forme se prolonge au point ∞ (faire $z \leftrightarrow 1/z$), donc à $\mathbf{P}^1(\mathbf{C}_p) - \{z \in \mathbf{C}_p \mid |z-1| \leq r\}$ qui contient U_r et est isomorphe à une boule ouverte. Or,

sur une boule ouverte, toute 1-forme analytique fermée s'intègre, et donc en particulier $P_1(\phi^*)(\frac{dz}{z-1})$, i.e. $P_1(\phi^*)(\frac{dz}{z-1}) \in d\mathcal{O}_0(U_r)$. Pour montrer que $\ell_1(z) = -\log(1-z)$, il suffit d'avoir $P_1(\phi^*)(-\log(1-z)) \in \mathcal{O}_0(U_r)$. Mais :

$$P_1(\phi^*)(-\log(1-z)) = \log \left(1 - \left(1 - \frac{1-z^p}{(1-z)^p} \right) \right)$$

et pour $\frac{1}{p^{p-1}} < r < |z-1|$, on vérifie que $\left| 1 - \frac{1-z^p}{(1-z)^p} \right| < 1$, donc :

$$P_1(\phi^*)(-\log(1-z)) = - \sum_{n \geq 1} \frac{\left(1 - \frac{1-z^p}{(1-z)^p} \right)^n}{n} \in \mathcal{O}_0(U_r). \quad \square$$

Remarque 2.2.2.2. — On peut prendre pour $P_k(T)$ les polynômes $p^k - T$ pour tout k ([Cole1],5.2).

Voici une application des polylogarithmes p -adiques. Soient χ un caractère de Dirichlet non trivial et $L_p(s, \chi)$ la fonction L p -adique de Kubota-Leopoldt, obtenue par interpolation des valeurs aux entiers *négatifs* de la fonction $L(s, \chi)$ complexe de Dirichlet ([Iw],3). Il est naturel de se demander quelles sont alors les valeurs de $L_p(s, \chi)$ aux entiers *positifs*. La réponse est fournie par le théorème suivant, dû à Coleman ([Cole1]) :

Théorème 2.2.2.3. — *Supposons $p \neq 2$. Soient $\chi : (\mathbf{Z}/d\mathbf{Z})^\times \rightarrow \mathbf{C}_p^\times$ un caractère de Dirichlet de conducteur $d > 1$ et ω le caractère de Dirichlet "représentant de Teichmüller" : $(\mathbf{Z}/p\mathbf{Z})^\times \rightarrow \mathbf{C}_p^\times$. Pour tout $k \geq 1$, on a :*

$$L_p(k, \chi\omega^{1-k}) = \left(1 - \frac{\chi(p)}{p} \right) \frac{g(\chi, \zeta)}{d} \sum_{a=1}^{d-1} \bar{\chi}(a) \ell_k(\zeta^{-a})$$

où $\chi\omega^{1-k}$ est le caractère de Dirichlet produit des caractères χ et ω^{1-k} , ζ une racine primitive $d^{\text{ième}}$ quelconque de 1 et $g(\chi, \zeta)$ la somme de Gauss définie par $g(\chi, \zeta) = \sum_{a=1}^{d-1} \chi(a)\zeta^a$.

On a supposé $p \neq 2$ seulement pour simplifier l'exposition. Le cas $k = 1$, qui n'utilise que le logarithme p -adique, est dû à Leopoldt ([Iw],5). On remarquera l'analogie avec la formule classique :

$$L(k, \chi) = \frac{g(\chi, \zeta)}{d} \sum_{a=1}^{d-1} \bar{\chi}(a) \ell_k(\zeta^{-a})$$

(qui découle simplement de l'identité pour tout $n : d\chi(n) = \sum_{1 \leq a, b \leq d-1} \chi(a)\bar{\chi}(b)\zeta^{a-bn}$) où ℓ_k désigne ici bien sûr le polylogarithme complexe.

Remarque 2.2.2.4. — Si l'on ne s'intéresse qu'aux polylogarithmes p -adiques, il existe une façon plus directe de les obtenir (voir [St]). Mais la méthode précédente (qui est celle originelle de Coleman) a l'avantage de s'étendre à des situations plus générales (voir *Régulateurs p -adiques* ci-dessous).

2.3. Autres applications

Nous passons en revue très rapidement certaines des autres applications de l'intégration p -adique.

Points de torsion sur les courbes

Dans [Cole3] et [Cole5], Coleman démontre, en utilisant l'intégration p -adique, la conjecture de Manin-Mumford, précédemment montrée par Raynaud ([Ra2]), qui dit que toute courbe de genre ≥ 2 contenue dans une variété abélienne sur un corps de caractéristique 0 n'a qu'un nombre fini de points de torsion. On se ramène au cas d'un corps de nombres et le point de départ de la preuve de Coleman est que si X est une courbe sur une extension finie K de \mathbf{Q}_p avec bonne réduction et $\iota : X \rightarrow J$ un morphisme d'Albanese, $\iota(Q) - \iota(P)$ est de torsion dans J si et seulement si $\int_P^Q \iota^* \omega = 0$ sur X pour tout $\omega \in H^0(J, \Omega_{J/\mathbf{C}_p}^1)$.

Régulateurs p -adiques

Dans [CS], Coleman et de Shalit généralisent l'intégration en (2.2.1) à des ouverts rigides un peu plus généraux que U^{rig} appelés "basic wide open spaces". Par une méthode tout à fait analogue à celle décrite en (2.2.1), cela leur permet d'intégrer sans restriction les logarithmes de toutes les fonctions rationnelles sur les courbes avec bonne réduction (et même sur certaines courbes à réduction semi-stable) et de définir ainsi des régulateurs p -adiques, généralisations des dilogarithmes p -adiques. Comme en (2.2.2.3), ils les utilisent pour calculer des valeurs spéciales de la fonction L p -adique associée à une courbe elliptique sur \mathbf{Q} avec multiplication complexe et bonne réduction en p .

Loi de réciprocité et hauteurs p -adiques

Soit X une courbe algébrique propre et lisse sur \mathbf{C}_p . Rappelons qu'une forme de troisième espèce sur X est une forme rationnelle ayant au plus des pôles simples et tels que les résidus en ces pôles soient des entiers. Si ω est une telle forme, le diviseur des résidus $Res(\omega) = \sum_P n_P P$ où P parcourt $X(\mathbf{C}_p)$ et n_P est le résidu en P est de degré 0 (i.e. $\sum n_P = 0$).

Pour toute forme rationnelle ω' dont les pôles évitent $Res(\omega)$, $\int_{Res(\omega)} \omega' = \sum_P n_P f_{\omega'}(P)$

ne dépend donc pas de la primitive $f_{\omega'}$ choisie en (1.1.1.5). Si η et η' sont deux formes de seconde espèce sur X , notons $\eta \cup \eta'$ la somme des résidus de $f_\eta \eta'$ où f_η est une primitive en (1.1.1.5) et où on calcule le résidu en un point à partir des développements analytiques de f_η et η' dans un voisinage p -adique de ce point (c'est indépendant des choix car η' a tous ses résidus nuls).

Soit J la jacobienne de X et \tilde{J} l'extension universelle de J par un groupe additif ([Me]). Les points de $\tilde{J}(\mathbf{C}_p)$ représentent les formes de troisième espèce sur X modulo les différentielles logarithmiques des fonctions rationnelles et l'algèbre de Lie de \tilde{J} est isomorphe à $H_{dR}^1(X)$. On a en particulier une application logarithme $Log_{\tilde{J}} : \tilde{J}(\mathbf{C}_p) \rightarrow H_{dR}^1(X)$.

Théorème 2.3.1. — (*Loi de réciprocité*) Si ω et ω' sont deux formes différentielles de troisième espèce dont les diviseurs des résidus sont étrangers, alors :

$$\int_{\text{Res}(\omega)} \omega' - \int_{\text{Res}(\omega')} \omega = \text{Log}_{\bar{J}}(\omega) \cup \text{Log}_{\bar{J}}(\omega').$$

Ce théorème a d'abord été démontré par Coleman dans le cas où X a bonne réduction ([Cole4]), puis par Colmez dans le cas général ([Colm2]).

Soit maintenant X une courbe propre et lisse sur un corps de nombres et J sa jacobienne. À partir des fonctions de Green p -adiques (cf. 2.1.3.1) Colmez définit un accouplement de hauteur bilinéaire et symétrique : $J(\overline{\mathbf{Q}}) \times J(\overline{\mathbf{Q}}) \rightarrow \prod_{p \leq \infty} \mathbf{Q}_p$. La symétrie découle de (2.3.1). Projeté sur \mathbf{R} , cet accouplement coïncide avec l'accouplement de Néron-Tate. Projeté sur \mathbf{Q}_p aux places p où X a bonne réduction, il avait déjà été défini par Coleman et Gross ([CG]). À noter que le fait de pouvoir considérer $\log(p)$ comme une *variable* est important dans la définition de l'accouplement ci-dessus.

Cohomologie de de Rham et réduction semi-stable

Si X est une courbe propre et lisse sur une extension finie K de \mathbf{Q}_p ayant réduction semi-stable (au sens où elle admet un modèle \mathcal{X} propre, plat et régulier sur les entiers de K dont la fibre spéciale est un diviseur à croisements normaux dans \mathcal{X}), et si de plus les composantes irréductibles de la fibre spéciale sont lisses, leurs tubes dans X^{rig} (cf. 1.2.1) sont des “basic wide open spaces” sur lesquels on peut intégrer les formes différentielles. En utilisant l'intégration sur ces tubes, Coleman et Iovita donnent dans [CI] une construction purement géométrique de la K_0 -structure de Fontaine-Jannsen D définie par Hyodo-Kato ([HK]) sur $H_{dR}^1(X)$ (i.e. $H_{dR}^1(X) = K \otimes_{K_0} D$) où K_0 est l'extension maximale de \mathbf{Q}_p non ramifiée dans K , ainsi que des opérateurs ϕ et N définis sur D (Frobenius et monodromie). Voir aussi les résultats de Le Stum à ce sujet ([Le]).

RÉFÉRENCES

- [Ber] Berthelot P. – *Cohomologie rigide et cohomologie rigide à supports propres, première partie*, prépublication de l'Institut de Recherche Mathématique de Rennes, 1996.
- [Bes] Besser A. – *A generalization of Coleman's p-adic integration theory*, prépublication du Mathematische Institut der Universität Münster, 1999.
- [BO] Berthelot P. – Ogus A., *F-isocrystals and de Rham cohomology I*, Inv. Math. 72, 1983, 159-199.
- [Bog] Bogomolov F. – *Sur l'algébricité des représentations l-adiques*, C.R.A.S. 290, 1980, 701-703.

- [BLR] Bosch S. – Lütkebohmert W., Raynaud M., *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete 21, Springer-Verlag, 1990.
- [Bou] Bourbaki N. – *Groupes et algèbres de Lie, Chapitres 2 et 3*, Hermann, 1972.
- [Colm1] Colmez P. – *Périodes p -adiques des variétés abéliennes*, Math. Ann. 292, 1992, 629-644.
- [Colm2] Colmez P. – *Intégration sur les variétés p -adiques*, Astérisque 248, Soc. Math. de France, 1998.
- [Cole1] Coleman R. – *Dilogarithms, regulators and p -adic L -functions*, Inv. Math. 69, 1982, 171-208.
- [Cole2] Coleman R. – *Hodge-Tate periods and p -adic abelian integrals*, Inv. Math. 78, 1984, 351-379.
- [Cole3] Coleman R. – *Torsion points on curves and p -adic abelian integrals*, Annals of Math. 121, 1985, 111-168.
- [Cole4] Coleman R. – *Reciprocity laws on curves*, Comp. Math. 72, 1989, 205-235.
- [Cole5] Coleman R. – *Ramified torsion points*, Duke Math. J. 54, 1987, 615-640.
- [CG] Coleman R., Gross B. – *p -adic heights on curves*, Adv. Math. 17, 1989, 73-81.
- [CI] Coleman R., Iovita A. – *The Frobenius and Monodromy operators for Curves and Abelian Varieties*, Duke Math. J. 97, 1999, 171-215.
- [CS] Coleman R., de Shalit E. – *p -adic regulators on curves and special values of p -adic L -functions*, Inv. Math. 93, 1988, 239-266.
- [Dw] Dwork B. – *Normalized period matrices I*, Annals of Math. 94, 1971, 337-388.
- [Fo1] Fontaine J.-M. – *Sur certains types de représentations p -adiques du groupe de Galois d'un corps local; construction d'un anneau de Barsotti-Tate*, Annals of maths 115, 1982, 529-577.
- [Fo2] Fontaine J.-M. – *Le corps des périodes p -adiques*, Astérisque 223, Soc. Math. de France, 1994, 59-111.
- [Fo3] Fontaine J.-M. – *Formes différentielles et modules de Tate des variétés abéliennes sur les corps locaux*, Inv. Math. 65, 1982, 379-409.
- [Ga] Garuti M. – *Géométrie rigide et géométrie formelle*, à paraître aux Proceedings de la conférence de Luminy sur le groupe fondamental des courbes en caractéristique $p > 0$, décembre 1998.
- [II] Illusie L. – *Cohomologie de de Rham et cohomologie étale p -adique (d'après G. Faltings, J.-M. Fontaine et al.)*, Sémin. Bourbaki, Exposé n° 726, Astérisque 189-190, 1990, 325-374.

- [Iw] Iwasawa K. – *Lectures on p -adic L -functions*, Annals of Math. Studies 74, Princeton University Press, 1972.
- [KM] Katz N., Messing W. – *Some consequences of the Riemann hypothesis for varieties over finite fields*, Inv. Math. 23, 1974, 73-77.
- [La] Lang S. – *Abelian varieties*, Interscience, New-York, 1959.
- [Le] Le Stum B. – *La structure de Hyodo-Kato pour les courbes*, Rend. Sem. Mat. Univ. Padova 94, 1995, 279-301.
- [Me] Messing W. – *The universal extension of an abelian variety by a vector group*, Symp. Math. XI (Rome 1972), Academic Press, 1973, 359-372.
- [Mu1] Mumford D. – *Abelian varieties*, Oxford University Press, 1970.
- [Mu2] Mumford D. – *Analytic construction of degenerating curves over complete local fields*, Compos. Math. 24, 1972, 129-174.
- [Oe] Oesterlé J. – *Polylogarithmes*, Sémin. Bourbaki, Exposé n° 762, Astérisque 216, 1993, 49-67.
- [Ra1] Raynaud M. – *Géométrie analytique rigide, d'après Tate, Kiehl,...*, Mémoire de la S.M.F. 39-40, 1974, 319-327.
- [Ra2] Raynaud M. – *Courbes sur une variété abélienne et points de torsion*, Inv. Math. 71, 1983, 207-233.
- [Ra3] Raynaud M. – *Variétés abéliennes et géométrie rigide*, Proceedings du Congrès International de Nice, 1970, 473-477.
- [Se1] Serre J.-P. – *Morphismes universels et variétés d'Albanese*, Séminaire Chevalley 1958/1959, exposé 10.
- [Se2] Serre J.-P. – *Corps locaux*, 3^{ième} édition, Hermann, 1968.
- [St] Stevens G. – *p -adic polylogarithms*, notes d'une conférence à Schloss Ringberg am Tegernsee, 4 septembre 1998.
- [Ta1] Tate J. – *p -divisible groups*, Proc. of a conf. on local fields, Driebergen, 1967, 158-183.
- [Ta2] Tate J. – *Rigid analytic spaces*, Inv. Math. 12, 1971, 257-289.
- [Ts] Tsuji T. – *p -adic-étale cohomology and crystalline cohomology in the semi-stable reduction case*, à paraître à Inv. Math.
- [Wi] Wintenberger J.-P. – *Théorème de comparaison p -adique pour les schémas abéliens I : construction de l'accouplement de périodes*, Astérisque 223, Soc. Math. de France, 1994, 349-397.
- [Za1] Zarhin Y. – *Local heights and abelian varieties*, Groupe d'étude sur les problèmes diophantiens 1988/1989, Publ. math. n° 90 de l'université Pierre et Marie Curie.

- [Za2] Zarhin Y. – *p-adic abelian integrals and commutative Lie groups*, Algebraic geometry 4, J. Math. Sci. 81, 1996, 2744-2750.

Christophe BREUIL

Mathématiques, Bât. 425

U.R.A. 752 du C.N.R.S.

Université Paris-Sud

Bâtiment 425

F-91405 ORSAY cedex

E-mail : `breuil@math.u-psud.fr`

RECENT ADVANCES IN THE THEORY OF HOLONOMY

by Robert BRYANT

1. INTRODUCTION

1.0. OUTLINE. This report is organized as follows:

1. Introduction
2. Riemannian Holonomy
3. Torsion-free non-metric connections: the irreducible case

In a short lecture of this nature, it is impossible to describe the history of the subject in any depth, but the reader can find more information on the Riemannian (and pseudo-Riemannian) case by consulting [Bes], [Sa], and the forthcoming, much-anticipated [Jo3], especially for the exceptional cases. For the non-metric case, aside from the survey [Br3], the expository papers [MS2] and [Schw] provide a valuable account of both the representation theoretic and twistor theoretic approaches to the study of holonomy.

1.1. HISTORICAL REMARKS. According to the Oxford English Dictionary, it was Heinrich Hertz in 1899 who introduced the words *holonomic* and *nonholonomic* to describe a property of velocity constraints in mechanical systems.

Velocity constraints are *holonomic* if they force a curve in state space to stay in a proper subspace. As an example, the condition $\mathbf{p} \cdot d\mathbf{p} = 0$ for a vector particle $\mathbf{p} \in \mathbf{R}^n$ forces \mathbf{p} to have constant length, while the constraint $\mathbf{p} \wedge d\mathbf{p} = 0$ forces \mathbf{p} to move on a line.

Nonholonomic constraints, on the other hand, imply no such ‘finite’ constraints. A classical example is that of a ball rolling on a table without slipping or twisting. The state space is $B = \text{SO}(3) \times \mathbf{R}^2$, where the $\text{SO}(3)$ records the orientation of the ball and the \mathbf{R}^2 records its contact point on the plane. The rolling constraint is expressed as the set of differential equations

$$(0.1) \quad \alpha := a^{-1} da + a^{-1} \begin{pmatrix} 0 & 0 & -dx \\ 0 & 0 & -dy \\ dx & dy & 0 \end{pmatrix} a = 0$$

for a curve $(a(t); x(t), y(t))$ in B . The curves in B satisfying this constraint are those tangent to the 2-plane field $D = \ker \alpha$ that is transverse to the fibers of the

projection $\mathrm{SO}(3) \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$. It is not difficult to show that any two points of B can be joined by a curve tangent to D .

Such constraints and their geometry have long been of considerable interest in the calculus of variations and control theory. For recent results, see the foundational work on Carnot-Carathéodory geometries by Gromov [Gr].

1.2. THE HOLONOMY GROUP. Élie Cartan [Ca2] introduced the *holonomy group* in the context of differential geometry. It measures the failure of the parallel translation associated to a connection to be holonomic.

The data are a connected manifold M , a Lie group H with Lie algebra \mathfrak{h} , a principal right H -bundle $\pi : B \rightarrow M$, and a connection θ on B , i.e., θ is an \mathfrak{h} -valued 1-form on B that pulls back to each π -fiber to be the canonical left invariant 1-form on H and that satisfies the equivariance relation $R_h^*(\theta) = \mathrm{Ad}(h^{-1})(\theta)$. (The nonholonomic example described above with $(M, H, B, \theta) = (\mathbb{R}^2, \mathrm{SO}(3), \mathrm{SO}(3) \times \mathbb{R}^2, \alpha)$ is an example.)

A piecewise C^1 curve $\gamma : [0, 1] \rightarrow B$ is said to be θ -horizontal or θ -parallel if $\gamma^*(\theta) = 0$. The θ -holonomy $B_u^\theta \subset B$ of $u \in B$ is defined to be the set of values of $\gamma(1)$ as $\gamma : [0, 1] \rightarrow B$ ranges over the θ -horizontal curves with $\gamma(0) = u$. The $\mathrm{Ad}(H)$ -equivariance of θ (coupled with the connectedness of M) implies that there is a subgroup $H_u^\theta \subset H$ so that B_u^θ is an H_u^θ -subbundle of B and that $H_{u \cdot h}^\theta = h^{-1} H_u^\theta h$ for $h \in H$. Consequently, the conjugacy class of $H_u^\theta \subset H$ depends only on θ . The group H_u^θ (or, more informally, its conjugacy class in H) is called the *holonomy* of θ .

It is a theorem of Borel and Lichnerowicz [BL] that H_u^θ is a Lie subgroup of H . By a theorem of Ambrose and Singer [AS], the Lie algebra \mathfrak{h}_u^θ of H_u^θ is spanned by the set

$$\{ \Theta(x, y) \mid x, y \in T_v B, v \in B_u^\theta \}$$

where $\Theta = d\theta + \frac{1}{2}[\theta, \theta]$ is the *curvature* of θ . The identity component $(H_u^\theta)^0 \subset H_u^\theta$ is known as the *restricted holonomy* of θ . There is a well-defined surjective homomorphism $\rho^\theta : \pi_1(M, \pi(u)) \rightarrow H_u^\theta / (H_u^\theta)^0$ that satisfies $\rho^\theta([\pi \circ \gamma]) = \gamma(1)(H_u^\theta)^0$ for every θ -horizontal curve γ with $\gamma(0) = u$ and $\gamma(1) \in u \cdot H$.

Using these results, it can be shown [KNo] that, when $\dim M > 1$, a Lie subgroup $G \subset H$ can be the holonomy group of a connection on B if and only if B admits a structural reduction to a G -bundle. Thus, the set of possible holonomies of connections on B is determined topologically.

1.3. H -STRUCTURES AND TORSION. A common source of connections in geometry is that of H -structures on manifolds. Suppose $\dim M = n$ and let \mathfrak{m} be a reference vector space of dimension n . An (\mathfrak{m} -valued) *coframe* at $x \in M$ is a linear isomorphism $u : T_x M \rightarrow \mathfrak{m}$. The set $F(M, \mathfrak{m})$ of \mathfrak{m} -valued coframes at the points of M is naturally a principal right $\mathrm{GL}(\mathfrak{m})$ -bundle over M , with basepoint projection $\pi : F(M, \mathfrak{m}) \rightarrow M$. There is a tautological \mathfrak{m} -valued 1-form ω on $F(M, \mathfrak{m})$ defined by the formula $\omega(v) = u(\pi'(v))$ for all $v \in T_u F(M, \mathfrak{m})$.

Let $H \subset \mathrm{GL}(\mathfrak{m})$ be a subgroup. An H -structure on M is an H -subbundle B such that $B \subset F(M, \mathfrak{m})$. When H is a closed subgroup of $\mathrm{GL}(\mathfrak{m})$ (the only case I will consider today), the set of H -structures on M is the set of sections of the

bundle $F(M, \mathfrak{m})/H \rightarrow M$. The problem of determining whether there exists an H -structure on M is a purely topological one.

Most of the familiar geometric structures on M can be described as H -structures. For example, when $H = O(Q) \subset GL(\mathfrak{m})$ is the group of linear transformations preserving a quadratic form Q of type (p, q) on \mathfrak{m} , a choice of H -structure on M is equivalent to a choice of pseudo-Riemannian metric of type (p, q) on M . When $H = Sp(S) \subset GL(\mathfrak{m})$ is the group of linear transformations preserving a nondegenerate skewsymmetric bilinear form S on \mathfrak{m} , a choice of H -structure on M is equivalent to a choice of a nondegenerate 2-form σ on M , i.e., an almost symplectic structure.

If $\pi : B \rightarrow M$ is an H -structure on M , the tautological form ω pulled back to B will also be denoted ω when there is no chance of confusion. If θ is a connection on B , then the *first structure equation of Cartan* says that there exists an H -equivariant function $T : B \rightarrow \text{Hom}(\Lambda^2(\mathfrak{m}), \mathfrak{m}) \simeq \mathfrak{m} \otimes \Lambda^2(\mathfrak{m}^*)$ so that

$$(0.2) \quad d\omega + \theta \wedge \omega = \frac{1}{2} T(\omega \wedge \omega).$$

The function T is the *torsion function* of θ , and θ is *torsion-free* if $T = 0$. Any connection θ' on B is $\theta + p(\omega)$ where $p : B \rightarrow \text{Hom}(\mathfrak{m}, \mathfrak{h})$ is an H -equivariant function. Its torsion function is $T' = T + \delta(p)$, where $\delta : \mathfrak{h} \otimes \mathfrak{m}^* \rightarrow \mathfrak{m} \otimes \Lambda^2(\mathfrak{m}^*)$ is defined via the inclusion $\mathfrak{h} \subset \mathfrak{m} \otimes \mathfrak{m}^*$ and the skewsymmetrizing map $\mathfrak{m} \otimes \mathfrak{m}^* \otimes \mathfrak{m}^* \rightarrow \mathfrak{m} \otimes \Lambda^2(\mathfrak{m}^*)$.

Evidently, the reduced map $\bar{T} : B \rightarrow \text{coker}(\delta)$ is independent of the choice of θ and is the obstruction to choosing a torsion-free connection on B . For example, when $H = Sp(S)$ as above, it is not difficult to see that $\text{coker}(\delta) \simeq \Lambda^3(\mathfrak{m}^*)$ and that the reduced torsion represents the exterior derivative $d\sigma$ of the associated almost symplectic form $\sigma \in \Omega^2(M)$. Thus, an $Sp(S)$ -structure admits a torsion-free connection if and only if the almost symplectic structure on M is actually symplectic. Furthermore, $\ker(\delta) \simeq S^3(\mathfrak{m}^*)$, so that the torsion function T does not determine a unique connection θ . By contrast, when $H = O(Q)$, the map δ is an isomorphism, which is merely the fundamental lemma of Riemannian geometry: Every pseudo-Riemannian metric on M possesses a unique compatible, torsion-free connection.

The reduced torsion is the first order local invariant for H -structures on M and H -structures satisfying $\bar{T} = 0$ are usually referred to as *torsion-free* or *1-flat*. Many (but not all) of the H -structures that have received the most attention in differential geometry are 1-flat. All pseudo-Riemannian metric structures are 1-flat, an almost symplectic structure is 1-flat if and only if it is symplectic (Darboux' Theorem), an almost complex structure is 1-flat if and only if it is complex (Newlander-Nirenberg Theorem), an Hermitian structure is 1-flat if and only if it is Kähler. Contact structures and Carnot-Carathéodory structures, on the other hand, are definitely not 1-flat, as the nondegeneracy of the reduced torsion is an essential part of the geometry.

1.4. CRITERIA FOR HOLONOMY IN THE TORSION-FREE CASE. The condition $\bar{T} = 0$ is a $\text{Diff}(M)$ -invariant, first order equation for H -structures on M . Given a torsion-free H -structure $B \subset F(M, \mathfrak{m})$, one can ask about the possibilities for the holonomy group of a torsion-free connection θ on B . When $\ker(\delta) = 0$, a very common situation, the torsion-free connection, if it exists, will be unique, so that it makes sense to speak of the holonomy of B itself.

The vanishing of the torsion implies nontrivial restrictions on the holonomy. Assuming $T = 0$ and differentiating (0.2) yields $\Theta \wedge \omega = 0$. Thus, the *curvature function* $R : B \rightarrow \mathfrak{h} \otimes \Lambda^2(\mathfrak{m}^*)$ of θ , for which the *second structure equation of Cartan*

$$(0.3) \quad \Theta = d\theta + \frac{1}{2}[\theta, \theta] = \frac{1}{2}R(\omega \wedge \omega)$$

holds, takes values in the kernel $K(\mathfrak{h})$ of the map $\delta : \mathfrak{h} \otimes \Lambda^2(\mathfrak{m}^*) \rightarrow \mathfrak{m} \otimes \Lambda^3(\mathfrak{m}^*)$ defined by the same methods¹ as the previous δ . In particular, if there is a proper subalgebra $\mathfrak{g} \subset \mathfrak{h}$ so that $K(\mathfrak{h}) \subseteq \mathfrak{g} \otimes \Lambda^2(\mathfrak{m}^*)$, then, by the Ambrose-Singer holonomy theorem, the (restricted) holonomy of θ must lie in the connected Lie group $G \subset H$ whose Lie algebra is \mathfrak{g} .

The intersection of the subspaces $\mathfrak{s} \subset \mathfrak{h}$ that satisfy $K(\mathfrak{h}) \subseteq \mathfrak{s} \otimes \Lambda^2(\mathfrak{m}^*)$ is an ideal $\mathfrak{g} \subset \mathfrak{h}$. Thus, a necessary condition that there exist a torsion-free connection with holonomy $H \subset \mathrm{GL}(\mathfrak{m})$ is that $K(\mathfrak{h}) \neq K(\mathfrak{g})$ for any proper ideal $\mathfrak{g} \subset \mathfrak{h}$. This is usually referred to as *Berger's first criterion* [Ber, Br3].

This criterion is very restrictive: If \mathfrak{h} is semi-simple, there are, up to equivalence, only a finite number of representations $\mathfrak{h} \hookrightarrow \mathfrak{gl}(\mathfrak{m})$ without trivial summands satisfying it. As a simple example, if $\mathfrak{h} \simeq \mathfrak{sl}(2, \mathbb{R})$, and $V_k \simeq S^k(V_1)$ denotes the irreducible $\mathfrak{sl}(2, \mathbb{R})$ -representation of dimension $k+1$, then the $\mathfrak{sl}(2, \mathbb{R})$ -representations \mathfrak{m} without V_0 -summands that satisfy Berger's first criterion are $V_1, V_1 \oplus V_1, V_2, V_3$, and V_4 .

Symmetric examples. One large class of examples where Berger's first criterion is satisfied is provided by the following construction: Suppose that there is a surjective skewsymmetric pairing $\mathfrak{m} \times \mathfrak{m} \rightarrow \mathfrak{h}$ so that, together with the Lie algebra bracket on \mathfrak{h} and the \mathfrak{h} -module pairing $\mathfrak{h} \times \mathfrak{m} \rightarrow \mathfrak{m}$, it defines a Lie algebra on $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$. Then the pair $(\mathfrak{g}, \mathfrak{h})$ is a symmetric pair of Lie algebras [KN]. Let G be the simply connected Lie group with Lie algebra \mathfrak{g} and let $\tilde{H} \subset G$ be the (necessarily closed) connected subgroup corresponding to the subalgebra $\mathfrak{h} \subset \mathfrak{g}$ and let $H = \mathrm{Ad}_M(\tilde{H}) \subset \mathrm{GL}(\mathfrak{m})$ be its almost faithful image. Then $M = G/\tilde{H}$ is an affine symmetric space in a canonical way, and the coset projection $G \rightarrow M$ covers a torsion-free H -structure on M with connection whose holonomy is H (by the assumption $[\mathfrak{m}, \mathfrak{m}] = \mathfrak{h}$ and the Ambrose-Singer holonomy theorem.) The classification of the symmetric Lie algebra pairs $(\mathfrak{g}, \mathfrak{h})$ is a (rather involved) algebra problem. It was solved by Berger [Ber2] in the case that \mathfrak{g} is semi-simple or when \mathfrak{h} acts irreducibly on $\mathfrak{m} \simeq \mathfrak{g}/\mathfrak{h}$.

The case where θ is a locally symmetric connection with holonomy H is characterized on the H -structure B by the condition that $R : B \rightarrow K(\mathfrak{h})$ be constant. In fact, differentiating (0.3) yields $0 = (dR + \theta.R)(\omega \wedge \omega)$ (where $\theta.R$ is the result of the \mathfrak{h} -module pairing $\mathfrak{h} \times K(\mathfrak{h}) \rightarrow K(\mathfrak{h})$). Equivalently, $dR = -\theta.R + R'(\omega)$ where $R' : B \rightarrow K(\mathfrak{h}) \otimes \mathfrak{m}^*$ must take values in the kernel $K^1(\mathfrak{h})$ of the natural map $\delta : K(\mathfrak{h}) \otimes \mathfrak{m}^* \rightarrow \mathfrak{h} \otimes \Lambda^3(\mathfrak{m}^*)$. Now, R' vanishes identically exactly when R is parallel, which is exactly when the pair (B, θ) defines a locally symmetric affine

¹ The maps I am denoting by δ (as well as the ones to follow) are part of the *Spencer complex* associated to the inclusion $\mathfrak{h} \subset \mathfrak{m} \otimes \mathfrak{m}^*$, see [Br3]. These maps are \mathfrak{h} -module maps and so the various kernels and cokernels to be introduced are \mathfrak{h} -modules as well.

structure on M . Thus, if H can be the holonomy of a torsion-free affine connection that is not locally symmetric, then $K^1(\mathfrak{h}) \neq 0$. This is *Berger's second criterion*.

For example, when $\mathfrak{h} = \mathfrak{sl}(2, \mathbb{R})$ and $\mathfrak{m} = V_4$, one has $K^1(\mathfrak{h}) = 0$. Thus, this 5-dimensional representation of $\mathrm{SL}(2, \mathbb{R})$ could occur as holonomy of a torsion-free connection on M^5 only when that connection is locally symmetric. In fact, it occurs as the holonomy of the symmetric spaces $\mathrm{SL}(3, \mathbb{R})/\mathrm{SO}(2, 1)$ and $\mathrm{SU}(2, 1)/\mathrm{SO}(2, 1)$ and in no other way. The other four cases with $\mathfrak{h} = \mathfrak{sl}(2, \mathbb{R})$ satisfying Berger's first criterion also satisfy Berger's second criterion. Moreover, each does occur as the holonomy of a torsion-free connection that is not locally symmetric.

1.5. CLASSIFICATION. In the case where \mathfrak{m} is an irreducible \mathfrak{h} -module (implying that \mathfrak{h} is reductive), Berger's fundamental works [Ber1, Ber2] went a long way towards classifying the subalgebras $\mathfrak{h} \subset \mathfrak{gl}(\mathfrak{m})$ that satisfy his first and second criteria. The methods involved heavy use of representation theory and ultimately reduced to a painstaking, elaborate case analysis. This list was refined and completed by the combined work of several people: Alekseevskii [Al1], myself, and most recently and importantly, the combined work of Chi, Merkulov, and Schwachhöfer. I will discuss this further in §3.

Berger's work provided a (partial) list of candidates for the irreducibly acting holonomy groups of torsion-free connections that are not locally symmetric. He divided the list into two parts: The first part consists of $\mathfrak{h} \subset \mathfrak{gl}(\mathfrak{m})$ that lie in some $\mathfrak{so}(Q)$ for some nondegenerate quadratic form Q on \mathfrak{m} , so that the associated H -structure defines a pseudo-Riemannian structure on M . For these cases, the injectivity of the map $\delta : \mathfrak{h} \otimes \mathfrak{m}^* \rightarrow \mathfrak{m} \otimes \Lambda^2(\mathfrak{m}^*)$ implies that the torsion-free connection is unique, so that it makes sense to speak of the holonomy of the underlying H -structure itself. I will refer to this part as the *metric list*. The second part consists of $\mathfrak{h} \subset \mathfrak{gl}(\mathfrak{m})$ that do not lie in any $\mathfrak{so}(Q)$ and hence will be referred to as the *nonmetric list*. For many of the subalgebras on the nonmetric list, the map δ is not injective, so that the associated H -structure does not determine the torsion-free connection.

These two lists are Tables I and II, essentially. I have taken the liberty of modifying Berger's lists slightly, dropping the entries in the original list that did not actually satisfy Berger's two criteria ($\mathrm{SO}^*(2n) \simeq \mathrm{SO}(n, \mathbb{H})$, which does not satisfy the first criterion², and the $\mathrm{Spin}(9, \mathbb{C})$ -type entries, which do not satisfy the second criterion³) and including $\mathrm{Sp}(p, \mathbb{R}) \cdot \mathrm{SL}(2, \mathbb{R})$, the inadvertently omitted 'split form' of the quaternionic Kähler case. In the nonmetric case, I have amplified the list somewhat by making explicit the various real forms as well as the fact that, except for $\mathrm{CO}(p, q)$, each of the entries on Berger's nonmetric list represents only the semi-simple part of H , to which one can add an arbitrary subgroup of the (abelian) commuting subgroup to make the full group H .

Exotic Holonomies. The nonmetric list supplied by Berger was never claimed to be a complete list, though it was supposed to have omitted at most a finite number of possibilities. The full list of omissions, comprising Tables III and IV and nowadays

² An observation due to R. McLean.

³ Observed independently by Alekseevskii [Al1] and Brown and Gray [BG].

I. Pseudo-Riemannian, Irreducible Holonomies in \mathbb{R}^n

n	H	Local Generality*
$p+q \geq 2$ $2p$	$SO(p, q)$ $SO(p, \mathbb{C})$	$\frac{1}{2}n(n-1)$ of n $\frac{1}{2}p(p-1)^c$ of p^c
$2(p+q) \geq 4$ $2(p+q) \geq 4$	$U(p, q)$ $SU(p, q)$	1 of n 2 of $n-1$
$4(p+q) \geq 8$	$Sp(p, q)$	$2(p+q)$ of $(2p+2q+1)$
$4(p+q) \geq 8$ $4p \geq 8$ $8p \geq 16$	$Sp(p, q) \cdot Sp(1)$ $Sp(p, \mathbb{R}) \cdot SL(2, \mathbb{R})$ $Sp(p, \mathbb{C}) \cdot SL(2, \mathbb{C})$	$2(p+q)$ of $(2p+2q+1)$ $2p$ of $(2p+1)$ $2p^c$ of $(2p+1)^c$
7	G_2	6 of 6
7	G'_2	6 of 6
14	G_2^c	6^c of 6^c
8	$Spin(7)$	12 of 7
8	$Spin(4, 3)$	12 of 7
16	$Spin(7, \mathbb{C})$	12^c of 7^c

*Counted modulo diffeomorphism. The notation “ d of ℓ ” means “ d functions of ℓ variables” and a superscript \mathbb{C} means ‘holomorphic’.

referred to as the *exotic list*, was recently compiled by a combination of the efforts of Chi, Merkulov, Schwachhöfer, and myself. This will be reported on in §3, along with the reasons for the division into two lists.

1.6. LOCAL EXISTENCE. Berger’s lists (suitably modified) provide possibilities for irreducibly acting holonomy groups, but to verify that these possibilities actually can occur requires methods beyond representation theory. Most of the methods that have been employed can be grouped into a small number of categories:

Explicit construction. This is the simplest method, when it is available. For the metric list, there are locally symmetric examples with every holonomy except the special Kähler cases, where the holonomy is $SU(p, q) \subset GL(\mathbb{C}^{p+q})$; the hyper-Kähler cases, where the holonomy is $Sp(p, q) \subset GL(\mathbb{H}^{p+q})$; and the ‘exceptional’ holonomies, which comprise the groups $G_2^c \subset GL(\mathbb{C}^7)$ and $Spin(7, \mathbb{C}) \subset GL(\mathbb{C}^8)$ and certain of their real forms. Of course, one would like to know that the locally symmetric examples are not the only ones.

Sometimes constructing examples is easy: The generic pseudo-Riemannian metric has holonomy is $SO(p, q)$.

In other cases, simple underlying geometric structures can be used as a starting point. For example, all complex structures are flat, and the general (pseudo-)Kähler metric can be described in the standard background complex structure by means of a (pseudo-)Kähler potential. For the generic choice of such a potential, the holonomy will be $U(p, q) \subset SO(2p, 2q)$. Another example is the special Kähler case, where one

II. The ‘Classical’ Non-Metric, Irreducible Holonomies

H^\pm	m	Restrictions ^o
$G_{\mathcal{R}} \cdot \mathrm{SL}(n, \mathbb{R})$	\mathbb{R}^n	$n \geq 2$
$G_c \cdot \mathrm{SL}(n, \mathbb{C})$	\mathbb{C}^n	$n \geq 1$
$G_{\mathcal{R}} \cdot \mathrm{SL}(n, \mathbb{H})$	\mathbb{H}^n	$n \geq 1$
$\mathrm{Sp}(n, \mathbb{R})$	\mathbb{R}^{2n}	$n \geq 2$
$\mathrm{Sp}(n, \mathbb{C})$	\mathbb{C}^{2n}	$n \geq 2$
$\mathbb{R}^+ \cdot \mathrm{Sp}(2, \mathbb{R})$	\mathbb{R}^4	
$\mathbb{C}^* \cdot \mathrm{Sp}(2, \mathbb{C})$	\mathbb{C}^4	
$\mathrm{CO}(p, q)$	\mathbb{R}^{p+q}	$p + q \geq 3$
$G_c \cdot \mathrm{SO}(n, \mathbb{C})$	\mathbb{C}^n	$n \geq 3$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{R}) \cdot \mathrm{SL}(q, \mathbb{R})$	$\mathbb{R}^{pq} \simeq \mathbb{R}^p \otimes_{\mathbb{R}} \mathbb{R}^q$	$p \geq q \geq 2, (p, q) \neq (2, 2)$
$G_c \cdot \mathrm{SL}(p, \mathbb{C}) \cdot \mathrm{SL}(q, \mathbb{C})$	$\mathbb{C}^{pq} \simeq \mathbb{C}^p \otimes_{\mathbb{C}} \mathbb{C}^q$	$p \geq q \geq 2, (p, q) \neq (2, 2)$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{H}) \cdot \mathrm{SL}(q, \mathbb{H})$	$\mathbb{R}^{4pq} \simeq \mathbb{H}^p \otimes_{\mathbb{H}} \mathbb{H}^q$	$p \geq q \geq 1, (p, q) \neq (1, 1)$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{C})$	$\mathbb{R}^{p^2} \simeq (\mathbb{C}^p \otimes_{\mathbb{C}} \overline{\mathbb{C}^p})^{\mathcal{R}}$	$p \geq 3$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{R})$	$\mathbb{R}^{p(p+1)/2} \simeq \mathbb{S}_{\mathcal{R}}^2(\mathbb{R}^p)$	$p \geq 3$
$G_c \cdot \mathrm{SL}(p, \mathbb{C})$	$\mathbb{C}^{p(p+1)/2} \simeq \mathbb{S}_c^2(\mathbb{C}^p)$	$p \geq 3$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{H})$	$\mathbb{R}^{p(2p+1)} \simeq \mathbb{S}_{\mathcal{H}}^2(\mathbb{H}^p)$	$p \geq 2$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{R})$	$\mathbb{R}^{p(p-1)/2} \simeq \Lambda_{\mathcal{R}}^2(\mathbb{R}^p)$	$p \geq 5$
$G_c \cdot \mathrm{SL}(p, \mathbb{C})$	$\mathbb{C}^{p(p-1)/2} \simeq \Lambda_c^2(\mathbb{C}^p)$	$p \geq 5$
$G_{\mathcal{R}} \cdot \mathrm{SL}(p, \mathbb{H})$	$\mathbb{R}^{p(2p-1)} \simeq \Lambda_{\mathcal{H}}^2(\mathbb{H}^p)$	$p \geq 3$

[±] $G_{\mathcal{F}}$ is any connected subgroup of \mathbb{F}^* .

^o To avoid repetition or reducibility.

can start with a background complex structure with a specified holomorphic volume form and then find a Kähler metric preserving this volume form by requiring that the Kähler potential satisfy a single second order, elliptic equation.

One can also find examples by looking for those with a large symmetry group. For the hyper-Kähler case, one can start with the complex symplectic structure on the cotangent bundle of certain Hermitian symmetric spaces and look for a Kähler potential compatible with this complex symplectic structure that is invariant under the action of the isometry group, thereby reducing the problem to solving an ordinary differential equation. This was Calabi’s method for constructing a hyper-Kähler metric on $T^*\mathbb{C}\mathbb{P}^n$, the first known example in general dimensions. In the quaternionic Kähler case, where the holonomy is $\mathrm{Sp}(p, q) \cdot \mathrm{Sp}(1) \subset \mathrm{GL}(\mathbb{R}^{4(p+q)})$, Alekseevskii found homogeneous nonsymmetric examples on certain solvable Lie groups. Even for the exceptional holonomies, there are explicit examples of cohomogeneity one [Br1], [BS].

Examples in the hyper-Kähler and quaternionic Kähler cases can also be constructed by the method of *reduction*, which takes advantage of descriptions of these

III. Exotic Conformal Holonomies

IV. Exotic Symplectic Holonomies

H^\pm	\mathfrak{m}	H	\mathfrak{m}
$\mathbb{R}^+ \cdot \mathrm{SL}(2, \mathbb{R})$ $\mathbb{C}^* \cdot \mathrm{SL}(2, \mathbb{C})$	$\mathbb{R}^4 \simeq \mathbb{S}^3(\mathbb{R}^2)$ $\mathbb{C}^4 \simeq \mathbb{S}^3(\mathbb{C}^2)$	$\mathrm{SL}(2, \mathbb{R})$ $\mathrm{SL}(2, \mathbb{C})$	$\mathbb{R}^4 \simeq \mathbb{S}^3(\mathbb{R}^2)$ $\mathbb{C}^4 \simeq \mathbb{S}^3(\mathbb{C}^2)$
$G_c \cdot \mathrm{SL}(2, \mathbb{R})^a$ $G_c \cdot \mathrm{Sp}(1)^b$	$\mathbb{C}^2 \simeq \mathbb{C} \otimes_{\mathbb{R}} \mathbb{R}^2$ $\mathbb{C}^2 \simeq \mathbb{C} \otimes_{\mathbb{C}} \mathbb{H}$	$\mathrm{SL}(2, \mathbb{R}) \cdot \mathrm{SO}(p, q)^c$ $\mathrm{SL}(2, \mathbb{C}) \cdot \mathrm{SO}(n, \mathbb{C})^d$ $\mathrm{Sp}(1) \cdot \mathrm{SO}(n, \mathbb{H})^e$	$\mathbb{R}^2 \otimes \mathbb{R}^{p+q}$ $\mathbb{C}^2 \otimes \mathbb{C}^n$ \mathbb{H}^n
$G_{\mathbb{R}} \cdot \mathrm{Spin}(5, 5)$ $G_{\mathbb{R}} \cdot \mathrm{Spin}(1, 9)$ $G_c \cdot \mathrm{Spin}(10, \mathbb{C})$	\mathbb{R}^{16} \mathbb{R}^{16} \mathbb{C}^{16}	$\mathrm{Sp}(3, \mathbb{R})$ $\mathrm{Sp}(3, \mathbb{C})$	$\mathbb{R}^{14} \simeq \Lambda_0^3(\mathbb{R}^6)$ $\mathbb{C}^{14} \simeq \Lambda_0^3(\mathbb{C}^6)$
$G_{\mathbb{R}} \cdot \mathbb{E}_6^1$ $G_{\mathbb{R}} \cdot \mathbb{E}_6^4$ $G_c \cdot \mathbb{E}_6^c$	\mathbb{R}^{27} \mathbb{R}^{27} \mathbb{C}^{27}	$\mathrm{SL}(6, \mathbb{R})$ $\mathrm{SU}(1, 5)$ $\mathrm{SU}(3, 3)$ $\mathrm{SL}(6, \mathbb{C})$	$\mathbb{R}^{20} \simeq \Lambda^3(\mathbb{R}^6)$ $\mathbb{R}^{20} \simeq \Lambda^3(\mathbb{C}^6)^\tau$ $\mathbb{R}^{20} \simeq \Lambda^3(\mathbb{C}^6)^\tau$ $\mathbb{C}^{20} \simeq \Lambda^3(\mathbb{C}^6)$
		$\mathrm{Spin}(2, 10)$ $\mathrm{Spin}(6, 6)$ $\mathrm{Spin}(12, \mathbb{C})$	\mathbb{R}^{32} \mathbb{R}^{32} \mathbb{C}^{32}
		\mathbb{E}_7^5 \mathbb{E}_7^7 \mathbb{E}_7^c	\mathbb{R}^{56} \mathbb{R}^{56} \mathbb{C}^{56}

$^\pm G_{\mathbb{F}}$ is any connected subgroup of \mathbb{F}^*

Restrictions:

$^a G_c \not\subseteq \mathbb{R}^*$ (for irreducibility).

$^b G_c \not\subseteq S^1$ (to be nonmetric).

$^c p + q \geq 3$ (for irreducibility).

$^d n \geq 3$ (for irreducibility).

$^e n \geq 2$ (to be nonmetric).

structures in terms of multi-symplectic geometry, generalizing the well-known method of Marsden-Weinstein reduction in symplectic geometry so as to handle the multi-symplectic case. For an account, see [Bes, Addendum E].

Twistor Methods. After Penrose's description of the self-dual metrics in dimension 4, Hitchin and Salamon [Sa], among others, were able to generalize this method to describe the hyper-Kähler and quaternionic Kähler metrics in terms of natural holomorphic geometric structures on the moduli space of rational curves with certain simple normal bundles in a complex manifold.

In fact, it was the study [Br2] of the moduli space of rational curves on a complex surface with normal bundle $\mathcal{O}(3)$ that turned up the first known examples (the first two entries in each of Tables III and IV) of omissions⁴ from Berger's nonmetric list. Moreover, in the holomorphic category, it was shown that any connection with one of these holonomies could be constructed as the natural connection on the four-dimensional component of the moduli space of Legendrian rational curves in a complex contact three-fold.

Following this, Merkulov [Me1] showed that this approach could be generalized to cover the geometry of the moduli space of Legendrian deformations of certain

⁴ These examples were referred to as 'exotic' in [Br2] and the term has been adopted to describe any nonmetric subgroup $H \subset \mathrm{GL}(\mathfrak{m})$ that satisfies Berger's criteria but that does not appear on Berger's original nonmetric list.

complex homogeneous spaces and began to discover more exotic examples. This and its further developments will be reported on in §3.

Exterior differential systems. Another approach is to treat the equation $\bar{T} = 0$ directly as a system of PDE for sections of the bundle $F(M, \mathfrak{m})/H \rightarrow M$. In nearly all cases, this method leads to the study of an overdetermined system of PDE, so that Cartan-Kähler machinery must be brought to bear. For definitions and results regarding Cartan-Kähler theory, the reader can consult [BCG].

The general approach can be summarized as follows: Consider the structure equations derived so far for a torsion-free connection θ on an H -structure $B \subset F(M, \mathfrak{m})$,

$$(0.4) \quad \begin{aligned} d\omega &= -\theta \wedge \omega \\ d\theta &= -\frac{1}{2}[\theta, \theta] + \frac{1}{2}R(\omega \wedge \omega) \\ dR &= -\theta.R + R'(\omega) \end{aligned}$$

where $R : B \rightarrow K(\mathfrak{h})$ and $R' : B \rightarrow K^1(\mathfrak{h}) \subset K(\mathfrak{h}) \otimes \mathfrak{m}^*$ are as defined before. There are two things that need to be checked in order to be able to apply Cartan's general existence theorem for coframings at this level: First, the inclusion $K^1(\mathfrak{h}) \subset K(\mathfrak{h}) \otimes \mathfrak{m}^*$ should be an involutive tableau in Cartan's sense. Second, there should be a quadratic map $Q : K(\mathfrak{h}) \rightarrow K^1(\mathfrak{h}) \otimes \mathfrak{m}^*$ so that the exterior derivative of the third structure equation in (0.4) can be written in the form $(dR' + \theta.R' - Q(R)(\omega))(\omega) = 0$. (This is the familiar 'vanishing torsion' condition in exterior differential systems.)

When these two conditions are satisfied, Cartan's existence theorem asserts that, up to local diffeomorphism, the real analytic torsion-free connections with holonomy lying in $H \subset \text{GL}(\mathfrak{m})$ depend on s_q functions of q variables, where s_q is the last nonzero Cartan character of $K^1(\mathfrak{h}) \subset K(\mathfrak{h}) \otimes \mathfrak{m}^*$. Moreover, for any $(R_0, R'_0) \in K(\mathfrak{h}) \times K^1(\mathfrak{h})$, there exists a torsion-free connection θ on an H -structure $B \subset F(\mathfrak{m}, \mathfrak{m})$ and a $u_0 \in B$ for which the curvature functions R and R' satisfy $R(u_0) = R_0$ and $R'(u_0) = R'_0$.

When one can choose the element $R_0 \in K(\mathfrak{h}) \subset \mathfrak{h} \otimes \Lambda^2(\mathfrak{m}^*)$ so that it is surjective⁵ as a map $R_0 : \Lambda^2(\mathfrak{m}) \rightarrow \mathfrak{h}$, it will then follow from the Ambrose-Singer holonomy theorem that the holonomy of θ contains the identity component of H . If, moreover, one can choose R'_0 to be nonzero, such a connection θ will not be locally symmetric.

This analysis applies successfully to each of the entries of Tables I, II, and III. By contrast, for each of the entries of Table IV, the tableau $K^1(\mathfrak{h}) \subset K(\mathfrak{h}) \otimes \mathfrak{m}^*$ is not involutive. Further discussion of this point will be given in §3.

Generally, this method is good only for local analysis, but it has the distinct advantage that it not only proves existence of connections with a given holonomy, but provides their 'degree of generality' in Cartan's sense. For example, Table I gives the degree of generality of each of the possible pseudo-Riemannian, irreducible holonomies. For a similar discussion of the nonmetric list, see the survey [Br3], where various simplifications of the general argument are introduced to shorten the exposition.

⁵ Actually, one only needs that the image $R_0(\Lambda^2(\mathfrak{m})) \subset \mathfrak{h}$ generates \mathfrak{h} as an algebra.

This method was first used to prove the existence of metrics with holonomy G_2 and $\text{Spin}(7)$ and is still the only method that constructs the general local solution and describes its degree of generality. This is also the only known method for analyzing Entries 3 and 4 of Table III.

Poisson Constructions. The examples H in Table IV are subgroups of $\text{Sp}(S) \subset \text{GL}(\mathfrak{m})$ for a nondegenerate skewsymmetric bilinear form S on \mathfrak{m} . Hence the corresponding H -structures (when they exist) have an underlying symplectic structure.

For the first two examples from Table IV, each torsion-free connection (M, B, θ) of these types was analyzed and reconstructed in [Br2] from its derived curvature map $J = (R, R') : B \rightarrow K(\mathfrak{h}) \oplus K^1(\mathfrak{h})$. This reconstruction involved a number of seemingly miraculous identities, but since only these two examples were known, it did not seem worthwhile to look for an interpretation of these identities. However, when Chi, Merkulov, and Schwachhöfer [CMS] found other exotic symplectic examples, they noticed that this reconstruction technique generalized and they were able to explain it in the context of Poisson geometry in a very beautiful way. This, too, will be reported on in §3.

1.7. COMPACT RIEMANNIAN EXAMPLES. The history of compact Riemannian manifolds with reduced holonomy groups is long and complex, so I will not attempt a full account here. For more details on the cases I only mention, the reader can consult the relevant chapters of [Bes] and the references cited therein.

Kähler manifolds. This subject has the longest history, predating even Berger's classification. Every smooth algebraic variety carries a Kähler structure and this accounts for their importance in algebraic geometry.

Special Kähler manifolds. The major milestone here is, of course, Yau's solution in the mid 1970s of the Calabi conjecture, showing that every compact Kähler manifold with trivial canonical bundle carries a special Kähler structure. For this reason, compact manifolds endowed with such a structure are usually referred to as Calabi-Yau manifolds.

Hyper-Kähler manifolds. In the early 1980s, Beauville and Mukai were each able to construct compact, simply connected Kähler manifolds M^{4p} that carried a nondegenerate complex symplectic form (Fujiki had constructed examples for $p = 2$ slightly earlier). These necessarily had trivial canonical bundle, so by Yau's solution of the Calabi conjecture, these carried special Kähler structures. By an argument of Bochner, the complex symplectic form had to be parallel with respect to this special Kähler structure, which forced the holonomy to lie in $\text{Sp}(p)$. Further arguments showed that these examples were not products of lower dimensional complex manifolds and this implied that the holonomy had to actually be $\text{Sp}(p)$. These were the first known compact examples.

Quaternion Kähler manifolds. In this case, all known compact examples are locally symmetric, but we know of no reason why this should be true, except for dimension 8 [PoS]. Of course, a great deal is known about the possible geometry of such examples, see [Sa].

G_2 and Spin(7) manifolds. The remarkable recent work of Joyce [Jo1,2] establishes the existence of compact manifolds with these holonomies. This will be reported on in §2.

2. RIEMANNIAN HOLONOMY

In this section, g will denote a smooth Riemannian metric on a connected, smooth manifold M^n . The reference space \mathfrak{m} will be taken to be \mathbb{R}^n with its standard inner product, and $O(\mathfrak{m}) \subset GL(\mathfrak{m})$ will denote its orthogonal group. The $O(\mathfrak{m})$ -structure $B \subset F(M, \mathfrak{m})$ consisting of the coframes $u : T_x M \rightarrow \mathfrak{m}$ that are isometries of vector spaces is the orthonormal coframe bundle and the Levi-Civita connection on B will be denoted θ . This is, of course, the unique torsion-free $O(\mathfrak{m})$ -connection on B . When there is no danger of confusion, I will simply write H_u and B_u instead of H_u^θ and B_u^θ .

One thing that makes the Riemannian case simpler to deal with than others is the *de Rham Splitting Theorem* [Bes, KNo], which occurs in a local form and a global form. The local form asserts that if $(H_u)^\mathfrak{0}$ acts reducibly on \mathfrak{m} , say, preserving irreducible orthogonal subspaces $\mathfrak{m}_i \subset \mathfrak{m}$ for $1 \leq i \leq k$, then $(H_u)^\mathfrak{0}$ is the direct product of its subgroups $(H_u)_i^\mathfrak{0}$, where $(H_u)_i^\mathfrak{0}$ is the subgroup that acts trivially on \mathfrak{m}_j for $j \neq i$. Moreover, the metric g locally splits as a product in a corresponding way⁶. The global form asserts that if, in addition, M is simply connected and the metric g is complete, then M globally splits as a Riemannian product

$$(M, g) = (M_1, g_1) \times \cdots \times (M_k, g_k).$$

so that $(H_u)_i^\mathfrak{0} = (H_u)_i$ is the holonomy of (M_i, g_i) .

2.1. NON-CLOSED HOLONOMY. While the remainder of this report deals only with connected holonomy groups, I cannot pass up the opportunity to mention a recent result of particular interest in Riemannian holonomy. It had been a question for some time whether the holonomy of a compact Riemannian manifold must necessarily be compact. Using the de Rham Splitting Theorem and Berger's holonomy classification in the irreducible Riemannian case, one sees that this is so if M is simply connected. Thus, for any Riemannian manifold, the restricted holonomy group is compact, so it becomes a question of whether the fundamental group can cause the holonomy group to have an infinite number of components even when M is compact.

Very recently, B. Wilking [Wi] has shown that this can indeed occur. He has produced an example of a compact manifold whose holonomy group does have an infinite number of components. His example is of the form $M^5 = \Gamma \backslash (\mathbb{R}^2 \times N^3)$ where Γ is a subgroup of the isometry group of $\mathbb{R}^2 \times N^3$ that acts properly discontinuously and cocompactly.

⁶ What is also true, but not obvious, is that each of the groups $(H_u)_i^\mathfrak{0} \subset SO(\mathfrak{m}_i)$ is the holonomy of some Riemannian metric, even when it is not the holonomy of the corresponding local factor of g . See [Bes, Theorem 10.108]

2.2. COMPACT MANIFOLDS WITH EXCEPTIONAL HOLONOMY. The most remarkable development in Riemannian holonomy in recent years has been the spectacular construction by Dominic Joyce of compact 7-manifolds with holonomy G_2 and compact 8-manifolds with holonomy $\text{Spin}(7)$. His constructions are full of new ideas and, while it is not difficult to outline these ideas, their successful execution turns out to require very careful, subtle estimates. I will not attempt to explain these, but refer the reader to the original sources [Jo1,2] and to the forthcoming book [Jo3]. Also, I will concentrate on the G_2 case, as the $\text{Spin}(7)$ case follows the same spirit, but the details are different.

The fundamental 3-form. Let $\mathfrak{m} = \mathbb{R}^7$. It has been known for some time [Br1] that there is an open $\text{GL}(\mathfrak{m})$ -orbit $\Lambda_+^3(\mathfrak{m}^*)$ in the 3-forms on \mathfrak{m} so that the stabilizer of any element $\phi \in \Lambda_+^3(\mathfrak{m}^*)$ is a compact connected simple Lie group of dimension 14 and which is therefore isomorphic to G_2 . Consequently, for any 7-manifold M , there is an open subbundle $\Lambda_+^3(T^*M) \subset \Lambda^3(T^*M)$ so that the G_2 -structures on M are in one-to-one correspondence with the sections $\Omega_+^3(TM)$ of this bundle. Such a section will exist if and only if M is orientable and spinnable [LM], so I assume this from now on.

Associated to any section $\sigma \in \Omega_+^3(M)$, there is a canonical metric g_σ and orientation, inducing a well-defined Hodge star operator $*_\sigma : \Omega^p(M) \rightarrow \Omega^{7-p}(M)$. If σ is parallel with respect to the Levi-Civita connection of g_σ then the holonomy of g_σ will be a subgroup of G_2 and the G_2 -structure associated to σ will be torsion-free. By a result of Gray, this G_2 -structure is torsion-free if and only if both σ and $*_\sigma\sigma$ are closed. Conversely, if a Riemannian metric g on M^7 has holonomy a subgroup of G_2 , then there will be a g -parallel section $\sigma \in \Omega_+^3(M)$ for which $g = g_\sigma$.

A metric with holonomy a subgroup of G_2 is known to be Ricci flat [Bes, 10.64], so if M is compact, Bochner vanishing shows that its harmonic 1-forms and its Killing fields must be parallel. Thus, in the compact case, $b_1(M) = 0$ if the holonomy actually equals all of G_2 . Combining this information with the Cheeger-Gromoll splitting theorem [Bes, 6.67], one finds that a compact manifold with holonomy G_2 must have finite fundamental group, so one can, without loss of generality, assume that M is simply connected, which I do from now on. Conversely, a compact simply connected Riemannian 7-manifold whose holonomy is a subgroup of G_2 must actually equal G_2 since any proper subgroup of G_2 that can be a holonomy group fixes a nontrivial subspace.

One can now consider the moduli space \mathcal{M} consisting of the sections $\sigma \in \Omega_+^3(M)$ satisfying $d\sigma = d*_\sigma\sigma = 0$ modulo diffeomorphisms of M isotopic to the identity. There is an obvious ‘Torelli’ map $\tau : \mathcal{M} \rightarrow H_{\text{dR}}^3(M)$ defined by $\tau([\sigma]) = [\sigma]_{\text{dR}}$. Joyce’s first striking result [Jo1] is the analog of the local Torelli theorem, namely that τ is locally one-to-one and onto, in fact, a local diffeomorphism in the natural smooth structure on \mathcal{M} . Thus, the moduli space is said to be ‘unobstructed’.

Next, Joyce proves a remarkable existence theorem: If $\sigma \in \Omega_+^3(M)$ is closed, then there is a constant C that depends on the norm of the curvature of the metric g_σ , its volume, and its injectivity radius, so that, if $|d(*_\sigma\sigma)|_\sigma < C$, then there exists an exact 3-form ϕ so that $\sigma + \phi$ lies in $\Omega_+^3(M)$ and is closed and coclosed. In other words, a closed 3-form in $\Omega_+^3(M)$ that is ‘close enough’ to being coclosed can be perturbed

to a cohomologous 3-form in $\Omega_+^3(M)$ that is both closed and coclosed.

Thus, to prove the existence of a compact Riemannian 7-manifold with holonomy G_2 , it suffices to construct a simply connected 7-manifold endowed with a closed 3-form that satisfies such a ‘close enough’ estimate.

Joyce’s idea for doing this is extremely clever: He starts with the flat G_2 -structure σ_0 on the 7-torus $T^7 = \mathbf{R}^7/\mathbf{Z}^7$ and passes to a simply connected quotient orbifold $X = \Gamma \backslash T$ where Γ is a finite group of σ_0 -symmetries. This provides a flat G_2 -orbifold whose singular locus is a finite number of 3-tori T^3 , each of which has a neighborhood of the form $T^3 \times B^4/\{\pm\}$ where $B^4/\{\pm\}$ is the standard 4-ball around the origin in $\mathbf{R}^4 = \mathbf{C}^2$ divided by the equivalence relation $v \sim -v$.

Now, it has been known for a long time that $\mathbf{R}^4/\{\pm\}$ is metrically the scaling limit of the $SU(2)$ -holonomy metric on $T^*\mathbf{C}P^1$ constructed by Eguchi and Hanson as one scales the metric to contract the zero section to a point. Because $I_3 \times SU(2) \subset SO(7)$ is a subgroup of G_2 , it follows that one can regard the flat G_2 -structure on X in each singular locus neighborhood $T^3 \times B^4/\{\pm\}$ as the limit of a scaling of a G_2 -structure on $T^3 \times T^*\mathbf{C}P^1$. Joyce cuts out these singular neighborhoods and glues back in $T^3 \times N$ where N is a neighborhood of the zero section in $T^*\mathbf{C}P^1$, smoothly joining the flat 3-form with the Eguchi-Hanson-derived (closed) 3-form on the overlaps. This ‘surgery’ produces a smooth manifold \hat{X} , but does not disturb the fundamental group, which remains trivial.

By being very careful (here is where Joyce’s estimates are extremely delicate), he shows that he can do this in such a way that the resulting closed 3-form $\sigma \in \Omega_+^3(\hat{X})$ satisfies his estimate. I.e., it is close enough to being coclosed that it can be perturbed to a $\tilde{\sigma} \in \Omega_+^3(\hat{X})$ that is both closed and coclosed. Of course, since X is simply connected, it follows that the resulting torsion-free G_2 -structure has holonomy equal to G_2 .

By applying this idea to a number of different finite groups Γ , Joyce has been able to construct G_2 -metrics on a number of different 7-manifolds.

A similar set of ideas allows Joyce to construct compact 8-manifolds with holonomy $Spin(7)$, once one realizes that $Spin(7)$ can be defined in $GL(8, \mathbf{R})$ as the stabilizer of a certain 4-form in eight variables. The interested reader should consult [Jo1,2].

3. IRREDUCIBLE TORSION-FREE NON-METRIC CONNECTIONS

3.1. TWISTOR CONSTRUCTIONS. As previously mentioned, I found the first examples of ‘exotic’ holonomy groups by studying the geometry of the moduli space M of rational curves (i.e., complex curves of genus 0) in a complex surface S with normal bundle $\mathcal{O}(3)$. Following the examples of Hitchin’s study of rational curves in a surface S with normal bundle $\mathcal{O}(k)$ for $k = 1$ and 2, I knew that M would have dimension 4 and would have a natural G_3 -structure, where $G_3 \subset GL(4, \mathbf{C})$ is the image of $GL(2, \mathbf{C})$ acting by linear substitutions on the 4-dimensional space V_3 of cubic polynomials in two variables. I also knew that there would be a canonical G_3 -connection from general principles, but I was very surprised to find that this connection was

torsion-free.

In the examples Hitchin had analyzed, the geometry on the moduli space allowed one to reconstruct the surface S and so I fully expected to be able to do the same in this case. However, it turned out that the story was more subtle than that. In the standard double fibration picture:

$$\begin{array}{ccc} & I & \\ \lambda \swarrow & & \searrow \rho \\ M & & S \end{array}$$

where $I \subset M \times S$ is the set of pairs (C, p) where $p \in S$ is a point of the rational curve $C \in M$ and λ and ρ are just the projections onto the factors, each $p \in S$ would correspond to a hypersurface $H_p = \lambda(\rho^{-1}(p))$ in M (since it is only one condition for a curve in S to pass through a given point p). The members of this 2-parameter family of hypersurfaces in M would be expected to be the solutions of some differential geometric problem in M , but I was not able to find a geometrically defined 2-parameter family of hypersurfaces in the general 4-manifold carrying a torsion-free G_3 -structure.

However, a 3-parameter family Y of 2-dimensional surfaces did present itself. This can be described as follows: By the defining property of a G_3 -structure B on M^4 , each tangent space $T_x M$ can be thought of as the space of homogeneous cubic polynomials in two variables. This defines a distinguished \mathbb{P}^1 of lines, namely the perfect cubes, and a distinguished \mathbb{P}^1 of 2-planes, namely the multiples of a perfect square. I called such lines and 2-planes *null*. It was not difficult to show that, when B had a torsion-free connection, each null 2-plane was tangent to a unique totally geodesic 2-surface in M all of whose tangent planes were null. This family Y then fit into a double fibration⁷

$$\begin{array}{ccc} & N^5 & \\ \lambda \swarrow & & \searrow \rho \\ M^4 & & Y^3 \end{array}$$

where the fibers of λ are \mathbb{P}^1 's. Moreover, I was able to show that Y carried a natural structure as a contact manifold, that the family of \mathbb{P}^1 's given by $C_x = \rho(\lambda^{-1}(x))$ for $x \in M$ were all Legendrian curves for this contact structure, and that, moreover, this family was an open set in the space of Legendrian rational curves in Y . (In the surface case that I had started out with, Y turned out to be the projectivized tangent bundle of S .)

I then showed that the picture could be reversed: Starting with a holomorphic contact 3-manifold Y , one could look at the moduli space M of rational Legendrian curves C in Y to which the contact bundle $L \subset T^*Y$ restricts to be isomorphic to $\mathcal{O}(-3)$ and show that it was a smooth moduli space of dimension 4 on which there was a canonical torsion-free G_3 -structure.

⁷ Assuming Y to be Hausdorff in its natural topology, which can always be arranged by replacing M by a θ -convex open set in M .

Merkulov's generalization. In a remarkable series of papers, Merkulov [Me1,2,3] showed that this moduli space and double fibration construction obtains in a very general setting, starting from the data of an irreducibly acting (and therefore reductive) complex subgroup $H \subset \mathrm{GL}(n, \mathbb{C})$, a complex n -manifold M , and a holomorphic H -structure $B \subset F(M, \mathbb{C}^n)$ endowed with a holomorphic torsion-free connection θ .

The semi-simple part $H_s \subset H$ acts irreducibly on \mathbb{C}^n . With respect to a Cartan subalgebra of H_s and an ordering of its roots, there will be a unique line $E \subset (\mathbb{C}^n)^*$ spanned by a vector of highest weight. The H_s -orbit $F \subset \mathbb{P}((\mathbb{C}^n)^*)$ of E is a minimal H_s -orbit in $\mathbb{P}((\mathbb{C}^n)^*)$, a generalized flag variety of H_s of some dimension $k \leq n-1$, endowed with the hyperplane section bundle \mathcal{L} . (In the original case I treated, F is the projectivization of the set of perfect cubes and hence is a \mathbb{P}^1 . The bundle \mathcal{L} is $\mathcal{O}(-3)$.) The H -structure B provides identifications $T_x M \simeq \mathbb{C}^n$ unique up to an action of H , so there is a subbundle $N \subset \mathbb{P}(T^*M)$ whose fiber $N_x \subset \mathbb{P}(T_x^*M)$ over x corresponds to $F \subset \mathbb{P}((\mathbb{C}^n)^*)$ via any B -identification.

The projectivized cotangent bundle of any manifold is canonically a contact manifold, and the torsion-free condition on θ immediately implies that N is an *involutive* submanifold of $\mathbb{P}(T^*M)$, i.e., the restriction of the contact structure to N is degenerate, with Cauchy leaves of the largest possible dimension, namely $n-k-1$, the codimension of N in $\mathbb{P}(T^*M)$. When it is Hausdorff⁸, the leaf space Y of this Cauchy foliation is canonically a contact manifold, yielding the double fibration

$$\begin{array}{ccc} & N^{n+k} & \\ \lambda \swarrow & & \searrow \rho \\ M^n & & Y^{2k+1} \end{array}$$

where the manifolds $F_x = \rho(\lambda^{-1}(x))$ are Legendrian k -dimensional submanifolds of Y .

Merkulov then goes on to prove that, nearly always, one can recover M as the complete moduli space of the Legendrian immersions $F \subset Y$ that pull back the contact bundle $L \subset T^*Y$ to be \mathcal{L} . Moreover, when H_s acts as the full biholomorphism group of F (which, again, is nearly always) one can recover the original H -structure on M up to conformal scaling from the family of submanifolds $S_y = \lambda(\rho^{-1}(y))$ for $y \in Y$.

Finally, Merkulov gives representation theoretic criteria on an irreducibly acting subgroup $H \subset \mathrm{GL}(\mathfrak{m})$ with associated generalized flag variety $F \subset \mathbb{P}(\mathfrak{m}^*)$ of dimension k and hyperplane bundle \mathcal{L} , which guarantee that taking a $(2k+1)$ -dimensional contact manifold Y and considering the moduli space $M(Y)$ of Legendrian embeddings $F \subset Y$ that pull back the contact bundle L to be \mathcal{L} yields a smooth moduli space endowed with an H -structure and a torsion-free connection.

This last step is extremely important, for it provides a way to determine which irreducibly acting subgroups $H \subset \mathrm{GL}(\mathfrak{m})$ can occur as torsion-free holonomy in terms of representation theory, specifically, in terms of the vanishing of certain H -representations constructed functorially from \mathfrak{m} . This provides a different approach to solving the torsion-free holonomy problem, one that was carried out successfully by a combination of efforts of Chi, Merkulov, and Schwachhöfer. In particular, this

⁸ This can always be arranged by replacing M by a suitably θ -convex open set in M .

approach led to the discovery of the remaining groups in Table IV and, finally, the proof that Tables I, II, III, and IV exhaust the possibilities for irreducibly acting torsion-free holonomy.

3.2. POISSON CONSTRUCTIONS. The straightforward application of exterior differential systems to the holonomy problem outlined in §0.6 does not work for the entries of Table IV, at least at the level described so far. To see where the problem is, recall the structure equations derived so far for a torsion-free connection with holonomy $H \subset \mathrm{GL}(\mathfrak{m})$. They are

$$(3.1) \quad \begin{aligned} d\omega &= -\theta \wedge \omega \\ d\theta &= -\frac{1}{2}[\theta, \theta] + \frac{1}{2}R(\omega \wedge \omega) \\ dR &= -\theta.R + R'(\omega) \end{aligned}$$

where $R : B \rightarrow K(\mathfrak{h})$ and $R' : B \rightarrow K^1(\mathfrak{h}) \subset K(\mathfrak{h}) \otimes \mathfrak{m}^*$ are as already defined.

When one considers the first entry of Table IV, where $H = \mathrm{SL}(2, \mathbb{R})$ and $\mathfrak{m} \simeq V_3 = \mathcal{S}^3(V_1)$, it is not difficult to see that $K(\mathfrak{h}) \simeq V_2 = \mathcal{S}^2(V_1)$ has dimension 3 and that $K^1(\mathfrak{h}) \simeq V_3$ has dimension 4. Its Cartan characters are $(s_1, s_2, s_3, s_4) = (3, 1, 0, 0)$ but, as is easily computed, the prolongation $K^2(\mathfrak{h}) \subset K^1(\mathfrak{h}) \otimes \mathfrak{m}^*$ has dimension 1, so the tableau is not involutive and Cartan's existence theorem cannot be applied at this level.

However, there is a quadratic map $Q : K(\mathfrak{h}) \rightarrow K^1(\mathfrak{h}) \otimes \mathfrak{m}^*$ so that the exterior derivative of the third equation in (3.1) is $(dR' + \theta.R' - Q(R)(\omega))(\omega) = 0$, implying that there is a function $R'' : B \rightarrow K^2(\mathfrak{h})$ so that the equation

$$(3.2) \quad dR' = -\theta.R' + (R'' + Q(R))(\omega)$$

holds. Moreover, it is possible to choose the quadratic map Q in a unique way so that differentiating this last equation yields $dR''(\omega) = 0$. Now the second prolongation of $K^1(\mathfrak{h})$ vanishes, so this forces the structure equation

$$(3.3) \quad dR'' = 0.$$

Obviously, differentiating this equation will yield no new information.

At this point, Cartan's general existence theorem for coframings satisfying prescribed differential identities (a generalization of Lie's third fundamental theorem) can be applied to the entire ensemble

$$(3.4) \quad \begin{aligned} d\omega &= -\theta \wedge \omega, \\ d\theta &= -\frac{1}{2}[\theta, \theta] + \frac{1}{2}R(\omega \wedge \omega), \\ dR &= -\theta.R + R'(\omega), \\ dR' &= -\theta.R' + (R'' + Q(R))(\omega), \\ dR'' &= 0. \end{aligned}$$

His theorem implies that for every $(R_0, R'_0, R''_0) \in K(\mathfrak{h}) \times K^1(\mathfrak{h}) \times K^2(\mathfrak{h})$, there is a torsion-free H -structure $B \subset F(\mathfrak{m}, \mathfrak{m})$, unique up to local diffeomorphism, so that, at

a point $u_0 \in B$ one has $R(u_0) = R_0$, $R'(u_0) = R'_0$, and $R''(u_0) = R''_0$. Consequently, the space of diffeomorphism classes of germs of such torsion-free H -structures is finite dimensional.

What Chi, Merkulov, and Schwachhöfer show in [CMS] is that this exact same picture holds for each entry in Table IV. Namely, $K(\mathfrak{h}) \simeq \mathfrak{h}$, $K^1(\mathfrak{h}) \simeq \mathfrak{m}$, $K^2(\mathfrak{h}) \simeq \mathbb{F}$ ($= \mathbb{R}$ or \mathbb{C}), and there is a quadratic map $Q : \mathfrak{h} \rightarrow \mathfrak{m} \otimes \mathfrak{m}^*$ so that the above analysis of the structure equations goes through exactly the same as for the original two cases (with the obvious interpretations of the pairings).

Cartan's theorem then applies and yields not only the existence of torsion-free connections with these prescribed holonomies, but that the space of germs of such H -structures, modulo diffeomorphism, is finite dimensional.

In the original two cases that I treated, the finer understanding of the moduli space of solutions entailed understanding how the images $(R, R', R'')(B) \subset \mathfrak{h} \times \mathfrak{m} \times \mathbb{F}$ partition $\mathfrak{h} \times \mathfrak{m} \times \mathbb{F}$ into subsets. This analysis would have been hopeless were it not for several (at the time) amazing identities that I found by brute force calculation. They even allowed me to prove existence without using Cartan's existence theorem.

What is shown in [CMS], however, is that these mysterious identities can be explained in terms of a natural Poisson structure on the space $\mathfrak{h} \times \mathfrak{m} \times \mathbb{F}$ (actually, they regard the last factor as a parameter and consider Poisson structures on $\mathfrak{h} \times \mathfrak{m}$). The images $(R, R', R'')(B)$ turn out to be the symplectic leaves of this Poisson structure and this point of view simplifies the reconstruction of the H -structure from the leaf image (though it does not entirely remove some of the global difficulties having to do with the symplectic realizations necessary in their construction).

3.3. ALGEBRAIC CLASSIFICATION. Once the full list of the irreducible torsion-free holonomies was known, there arose the question of whether this list could be derived through Berger's original approach, i.e., representation theory. Schwachhöfer [Schw] has shown that this can indeed be done. (As is so often the case, knowing the answer in advance helps to organize the proof.) His fully algebraic classification of the irreducibly acting subgroups $H \subset GL(\mathfrak{m})$ that satisfy Berger's first criterion and the subset of those that also satisfy Berger's second criterion still involves quite a bit of case checking, but the general outline of the argument is clear.

What is particularly intriguing is Ziller's observation (reported in [Schw]) that this list can be constructed by a simple Ansatz starting from the list of Hermitian and quaternionic symmetric spaces. A direct proof of Ziller's Ansatz would be highly desirable.

3.4. TWO LEFTOVER CASES. As of this writing, each entry in the four Tables, save two, Entries 3 and 4 in Table III, has been treated in the literature and shown to occur as holonomy, either by twistor methods or exterior differential systems methods. Existence proofs by twistor methods have some difficulty when \mathfrak{m} is a complex vector space and the group $H \subset GL(\mathfrak{m})$ is of the form $H = G_c \cdot H_s$ where H_s is the semi-simple part and $G_c \subset \mathbb{C}^*$ is a one-dimensional subgroup of \mathbb{C}^* , acting as scalar multiplication on \mathfrak{m} . The method of exterior differential systems does not have this problem, but each case does require a separate treatment.

In my survey article [Br3], I left these two entries unsettled in the case where G_c had dimension 1 because, at the time, they did not seem that interesting. Now that they are the last unsettled cases, it seems to be a good idea to resolve them, so I will do that here, though, for lack of space, I will not provide details, just give the results of the Cartan-Kähler analysis.

For the first case, $H = G_c \cdot \mathrm{SL}(2, \mathbb{R}) \subset \mathrm{GL}(2, \mathbb{C})$, one must assume that $G_c \not\subseteq \mathbb{R}^*$, otherwise H will not act irreducibly on $\mathfrak{m} \simeq \mathbb{C}^2$. This leaves a one parameter family of possibilities $H_\lambda = \{e^{(i+\lambda)t} \mid t \in \mathbb{R}\} \subset \mathbb{C}^*$ where λ is any real number. By conjugation, one can assume that $\lambda \geq 0$, so I will do this. It turns out that there are two cases:

If $\lambda = 0$, so that $H_0 = S^1$, it is not difficult to compute that $K(\mathfrak{h}) \simeq V_4 \oplus V_2 \oplus V_0$, while $K^1(\mathfrak{h}) \simeq 2V_5 \oplus 2V_3 \oplus 2V_1$ and is involutive, with characters $(s_1, s_2, s_3, s_4) = (9, 9, 5, 1)$. Moreover, the torsion is absorbable. By Cartan's theorem, solutions exist and depend on one function of four variables.

However, if $\lambda > 0$, so that $H_\lambda \simeq \mathbb{R}^+$, one computes that $K(\mathfrak{h}) \simeq V_4 \oplus V_2$ while $K^1(\mathfrak{h}) \simeq 2V_5 \oplus 2V_3$ and is involutive, with characters $(s_1, s_2, s_3, s_4) = (8, 8, 4, 0)$. Moreover, the torsion is absorbable. By Cartan's theorem, solutions exist and depend on four functions of three variables.

For the second case, $H = G_c \cdot \mathrm{SU}(2) \subset \mathrm{GL}(2, \mathbb{C})$, one must assume that $G_c \not\subseteq S^1$, otherwise $H = \mathrm{U}(2)$ will preserve a metric on $\mathfrak{m} \simeq \mathbb{C}^2$. This leaves a one parameter family of possibilities $J_\lambda = \{e^{(1+i\lambda)t} \mid t \in \mathbb{R}\} \subset \mathbb{C}^*$ where λ is any real number. By conjugation, one can assume that $\lambda \geq 0$, though I won't need to do this. Here there is only one case: One computes that $K(\mathfrak{h}) \simeq V_4^{\mathbb{R}} \oplus V_2^{\mathbb{R}} \simeq \mathbb{R}^8$ while $K^1(\mathfrak{h}) \simeq V_5 \oplus V_3 \simeq \mathbb{C}^{10} \simeq \mathbb{R}^{20}$. The tableau is involutive, with characters $(s_1, s_2, s_3, s_4) = (8, 8, 4, 0)$. Moreover, the torsion is absorbable. By Cartan's theorem, solutions exist and depend on four functions of three variables.

REFERENCES

- [Al1] D. ALEKSEEVSKII - *Riemannian spaces with unusual holonomy groups*, *Funct. Anal. Appl.* **2** (1968), 97–105.
- [Al2] D. ALEKSEEVSKII - *Classification of quaternionic spaces with a transitive solvable group of motions*, *Math. USSR-Izv* **9** (1975), 297–339.
- [AS] A. AMBROSE and I. SINGER - *A theorem on holonomy*, *Trans. Amer. Math. Soc.* **75** (1953), 428–443.
- [BB] L. BÉRARD-BERGERY - *On the holonomy of Lorentzian manifolds*, *Proc. Symp. Pure Math.* **54** (1993), 27–40.
- [BI] L. BÉRARD-BERGERY and A. IKEMAKHEN - *Sur l'holonomie des variétés pseudo-riemanniennes de signature (n, n)* , *Bull. Soc. Math. France* **125** (1997), 93–114.
- [Ber1] M. BERGER, *Sur les groupes d'holonomie des variétés à connexion affine et des variétés riemanniennes*, *Bull. Soc. Math. France* **83** (1955), 279–330.
- [Ber2] M. BERGER - *Les espaces symétriques non compacts*, *Ann. Sci. École Norm. Sup.* **74** (1957), 85–177.
- [Bes] A. BESSE - *Einstein Manifolds*, *Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 10*, Springer-Verlag, Berlin, 1987.
- [BL] A. BOREL and A. LICHNEROWICZ - *Groupes d'holonomie des variétés riemanniennes*, *C.R. Acad. Sci. Paris* **234** (1952), 1835–1837.
- [BG] R. BROWN and A. GRAY - *Riemannian manifolds with holonomy group $\text{Spin}(9)$* , in *Differential Geometry (in honor of Kentaro Yano)*, Kinokuniya, Tokyo, 1972, 41–59.
- [Br1] R. BRYANT - *Metrics with exceptional holonomy*, *Ann. Math.* **126** (1987), 525–576.
- [Br2] R. BRYANT - *Two exotic holonomies in dimension four, path geometries, and twistor theory*, *Amer. Math. Soc. Proc. Symp. Pure Math.* **53** (1991), 33–88.
- [Br3] R. BRYANT - *Classical, exceptional, and exotic holonomies: a status report*, in *Actes de la Table Ronde de Géométrie Différentielle en l'Honneur de Marcel Berger*, Soc. Math. France, 1996, 93–166.
- [BCG] R. BRYANT, *et al* - *Exterior Differential Systems*, MSRI Series **18**, Springer-Verlag, 1991.
- [BS] R. BRYANT and S. SALAMON - *On the construction of some complete metrics with exceptional holonomy*, *Duke Math. J.* **58** (1989), 829–850.
- [Ca1] É. CARTAN - *La Géométrie des Espaces de Riemann*, *Mémoires des Sciences Mathématiques, Fasc. IX* (1925). (Especially, Chapitre VII, Section II.)
- [Ca2] É. CARTAN - *Les groupes d'holonomie des espaces généralisés*, *Acta. Math.* **48** (1926), 1–42.
- [CMS] Q.-S. CHI, S. MERKULOV, and L. SCHWACHHÖFER - *On the existence of infinite series of exotic holonomies*. *Inventiones Math.* **126** (1996), 391–411.
- [ChS] Q.-S. CHI and L. SCHWACHHÖFER - *Exotic holonomy on moduli spaces of rational curves*, *Differential Geom. App.* **8** (1998), 105–134.
- [Gr] M. GROMOV - *Carnot-Carathéodory spaces seen from within*, in *Sub-Riemannian geometry*, *Progr. Math.* **144**, Birkhäuser, Basel, 1996, 79–323.

- [Jo1] D. JOYCE - *Compact Riemannian 7-manifolds with holonomy G_2 : I & II*, J. Differential Geom. **43** (1996), 291–328, 329–375.
- [Jo2] D. JOYCE - *Compact 8-manifolds with holonomy $\text{Spin}(7)$* , Inventiones Math. **123** (1996), 507–552.
- [Jo3] D. JOYCE - *Compact manifolds with special holonomy*, Oxford University Press, Oxford, 2000 (projected).
- [KNa] S. KOBAYASHI and K. NAGANO - *On filtered Lie algebras and geometric structures*, II, J. Math. Mech. **14** (1965), 513–521.
- [KNo] S. KOBAYASHI and K. NOMIZU - *Foundations of Differential Geometry*, Vols. 1 & 2, Wiley-Interscience, New York, 1963.
- [LM] H. B. LAWSON, JR. and M. L. MICHELSON - *Spin geometry*, Princeton Math. Series **38**, Princeton Univ. Press, Princeton, NJ, 1989.
- [Me1] S. MERKULOV - *Moduli spaces of compact complex submanifolds of complex fibered manifolds*, Math. Proc. Camb. Phil. Soc. **118** (1995), 71–91.
- [Me2] S. MERKULOV - *Geometry of Kodaira moduli spaces*, Proc. Amer. Math. Soc. **124** (1996), 1499–1506.
- [Me3] S. MERKULOV - *Existence and geometry of Legendre moduli spaces*, Math. Z. **226** (1997), 211–265.
- [MS1] S. MERKULOV and L. SCHWACHHÖFER - *Classification of irreducible holonomies of torsion-free affine connections*, Ann. Math. (to appear).
- [MS2] S. MERKULOV and L. SCHWACHHÖFER - *Twistor solution of the holonomy problem*, 395–402, The Geometric Universe, Science, Geometry and the work of Roger Penrose, S.A. Hugget (ed.), Oxford Univ. Press, 1998.
- [PoS] Y. S. POON and S. SALAMON - *Quaternionic Kähler 8-manifolds with positive scalar curvature*, J. Differential Geom. **33** (1991), 363–378.
- [Sa] S. SALAMON - *Riemannian geometry and holonomy groups*, Pitman Research Notes in Math., no. 201, Longman Scientific & Technical, Essex, 1989.
- [Scho] J. SCHOUTEN - *On the number of degrees of freedom of the geodetically moving systems*, Proc. Kon. Acad. Wet. Amsterdam **21** (1918), 607–613.
- [Schw] L. SCHWACHHÖFER - *On the classification of holonomy representations*, Habilitationsschrift, Universität Leipzig, 1998.
- [Wi] B. WILKING - *On compact Riemannian manifolds with noncompact holonomy groups*, 1999 preprint.
URL: <http://wwwmath.uni-muenster.de/math/inst/sfb/about/publ/wilking.ps>

Robert BRYANT

Duke University
Department of Mathematics
P.O. Box 90320
DURHAM, NC 27708-0320
USA

E-mail : bryant@math.duke.edu

**CLASSICAL COMPUTING, QUANTUM COMPUTING,
AND SHOR'S FACTORING ALGORITHM**

by **Yuri I. MANIN**

0. WHY QUANTUM COMPUTING ?

Information processing (computing) is the dynamical evolution of a highly organized physical system produced by technology (computer) or nature (brain). The initial state of this system is (determined by) its input; its final state is the output. Physics describes nature in two complementary modes: classical and quantum. Up to the nineties, the basic mathematical models of computing, Turing machines, were classical objects, although the first suggestions for studying quantum models date back at least to 1980.

Roughly speaking, the motivation to study quantum computing comes from several sources: physics and technology, cognitive science, and mathematics. We will briefly discuss them in turn.

(i) Physically, the quantum mode of description is more fundamental than the classical one. In the seventies and eighties it was remarked that, because of the superposition principle, it is computationally unfeasible to simulate quantum processes on classical computers ([Po], [Fe1]). Roughly speaking, quantizing a classical system with N states we obtain a quantum system whose state space is an $(N - 1)$ -dimensional complex projective space whose volume grows *exponentially* with N . One can argue that the main preoccupation of quantum chemistry is the struggle with resulting difficulties. Reversing this argument, one might expect that quantum computers, if they can be built at all, will be considerably more powerful than classical ones ([Fe1], [Ma2]).

Progress in the microfabrication techniques of modern computers has already led us to the level where quantum noise becomes an essential hindrance to the error-free functioning of microchips. It is only logical to start exploiting the essential quantum mechanical behavior of small objects in devising computers, instead of neutralizing it.

(ii) As another motivation, one can invoke highly speculative, but intriguing, conjectures that our brain is in fact a quantum computer. For example, the recent progress in writing efficient chess playing software (Deep Blue) shows that to simulate the world championship level using only classical algorithms, one has to be able to analyze about 10^6 positions/sec and use about 10^{10} memory bytes. Since the characteristic time of neuronal processing is about 10^{-3} sec, it is very difficult to explain

how the classical brain could possibly do the job and play chess as successfully as Kasparov does. A less spectacular, but not less resource consuming task, is speech generation and perception, which is routinely done by billions of human brains, but still presents a formidable challenge for modern computers using classical algorithms.

Computational complexity of cognitive tasks has several sources: basic variables can be fields; a restricted amount of small blocks can combine into exponentially growing trees of alternatives; databases of incompressible information have to be stored and searched.

Two paradigms have been developed to cope with these difficulties: logic-like languages and combinatorial algorithms, and statistical matching of observed data to an unobserved model (see D. Mumford's paper [Mu] for a lucid discussion of the second paradigm.)

In many cases, the second strategy efficiently supports an acceptable performance, but usually cannot achieve excellency of the Deep Blue level. Both paradigms require huge computational resources, and it is not clear, how they can be organized, unless hardware allows massive parallel computing.

The idea of "quantum parallelism" (see sec. 2 below) is an appealing theoretical alternative. However, it is not at all clear that it can be made compatible with the available experimental evidence, which depicts the central nervous system as a distinctly classical device.

The following way out might be worth exploring. The implementation of efficient quantum algorithms which have been studied so far can be provided by one, or several, quantum chips (registers) controlled by a classical computer. A very considerable part of the overall computing job, besides controlling quantum chips, is also assigned to the classical computer. Analyzing a physical device of such architecture, we would have direct access to its classical component (electrical or neuronal network), whereas locating its quantum components might constitute a considerable challenge. For example, quantum chips in the brain might be represented by macromolecules of the type that were considered in some theoretical models for high temperature superconductivity.

The difficulties are seemingly increased by the fact that quantum measurements produce non-deterministic outcomes. Actually, one could try to use this to one's advantage, because there exist situations where we can distinguish the quantum randomness from the classical one by analyzing the probability distributions and using the Bell-type inequalities. With hindsight, one recognizes in Bell's setup the first example of the game-like situation where quantum players can behave demonstrably more efficiently than the classical ones (cf. the description of this setup in [Ts], pp. 52-54).

It would be extremely interesting to devise an experimental setting purporting to show that some fragments of the central nervous system relevant for information processing can in fact be in a quantum superposition of classical states.

(iii) Finally, we turn to mathematics. One can argue that nowadays one does not even need additional motivation, given the predominant mood prescribing the quantization of "everything that moves". Quantum groups, quantum cohomology, quantum invariants of knots etc come to mind. This actually seemed to be the primary

motivation before 1994, when P. Shor ([Sh]) devised the first quantum algorithm showing that prime factorization can be done on quantum computers in polynomial time, that is, considerably faster than by any known classical algorithm. (P. Shor's work was inspired by the earlier work [Si] of D. Simon). Shor's paper gave a new boost to the subject. Another beautiful result due to L. Grover ([Gro]) is that a quantum search among N objects can be done in $c\sqrt{N}$ steps. A. Kitaev [Ki1] devised new quantum algorithms for computing stabilizers of abelian group actions; his work was preceded by that of D. Boneh and R. Lipton [BoL], who treated the more general problem by a modification of Shor's method (cf. also [Gri]). At least as important as the results themselves, are the tools invented by Shor, Grover, and Kitaev.

Shor's work is the central subject of this lecture. It is explained in sec. 4. This explanation follows the discussion of the general principles of quantum computing and massive quantum parallelism in sec. 2, and of four quantum subroutines, including Grover's searching algorithm, in sec. 3. The second of these subroutines involving quantum computations of classical computable functions shows how to cope with the basic issue of quantum reversibility vs classical irreversibility. For more on this, see [Ben1] and [Ben2]. The opening sec. 1 contains a brief report on the classical theory of computability. I made some effort to express certain notions of computer science, including P/NP, in the language of mainstream mathematics. The last section 5 discusses Kolmogorov complexity in the context of classical and quantum computations.

Last, but not least, the hardware for quantum computing does not exist as yet: see 3.3 below for a brief discussion of the first attempts to engineer it. The quantum algorithms invented and studied up to now will stimulate the search of technological implementation which – if successful – will certainly correct our present understanding of quantum computing and quantum complexity.

Acknowledgements. I am grateful to Alesha Kitaev, David Mumford, and Dimitri Manin for their interest and remarks on the earlier version of this report. Many of their suggestions are incorporated in the text.

1. CLASSICAL THEORY OF COMPUTATION

1.1. Constructive universe. In this section I deal only with deterministic computations, which can be modelled by classical discrete time dynamical systems and subsequently quantized.

Alan Turing undertook the microscopic analysis of the intuitive idea of algorithmic computation. In a sense, he found its genetic code. The atom of information is one bit, the atomary operators can be chosen to act upon one/two bits and to produce the outputs of the same small size. Finally, the sequence of operations is strictly determined by the local environment of bounded size, again several bits.

For a change, I proceed in the reverse direction, and start this section with a presentation of the macrocosm of the classical theory of computation. Categorical language is appropriate to this end.

Let \mathcal{C} be a category whose objects are countable or finite sets U . Elements x of

these sets will generally be finite sets with additional structure. Without waiting for all the necessary axioms to be introduced, we will call $x \in U$ a *constructive object of type U* (an integer, a finite graph, a word in a given alphabet, a Boolean expression, an instance of a mass problem ...) The set U itself will be called *the constructive world* of objects of fixed type, and \mathcal{C} *the constructive universe*. The category \mathcal{C} , which will be made more concrete below, will contain all finite products and finite unions of its objects, and also finite sets U of all cardinalities.

Morphisms $U \rightarrow V$ in \mathcal{C} are certain partial maps of the underlying sets. More precisely, such a morphism is a pair $(D(f), f)$ where $D(f) \subset U$ and $f : D(f) \rightarrow V$ is a set-theoretic map. Composition is defined by

$$(D(g), g) \circ (D(f), f) = (g^{-1}D(f), g \circ f).$$

We will omit $D(f)$ when it does not lead to a confusion.

The morphisms f that we will be considering are *(semi)computable functions* $U \rightarrow V$. An intuitive meaning of this notion, which has a very strong heuristic potential, can be explained as follows: there should exist an algorithm φ such that if one takes as input the constructive object $u \in U$, one of the three alternatives holds:

- (i) $u \in D(f)$, φ produces in a finite number of steps the output $f(u) \in V$.
- (ii) $u \notin D(f)$, φ produces in a finite number of steps the standard output meaning NO.
- (iii) $u \notin D(f)$, φ works for an infinitely long time without producing any output.

The necessity of including the alternative (iii) in the definition of (semi-)computability was an important and non-trivial discovery of the classical theory. The set of all morphisms $U \rightarrow V$ is denoted $\mathcal{C}(U, V)$.

The sets of the form $D(f) \subset U$ are called *enumerable subsets of U* . If both $E \subset U$ and $U \setminus E$ are enumerable, E is called *decidable*.

The classical computation theory makes all of this more precise in the following way.

1.2. Definition. A category \mathcal{C} as above is called a *constructive universe* if it contains the constructive world \mathbf{N} of all integers ≥ 1 , finite sets $\emptyset, \{1\}, \dots, \{1, \dots, n\}, \dots$ and satisfies the following conditions (a)–(d).

(a) $\mathcal{C}(\mathbf{N}, \mathbf{N})$ is defined as the set of all partially recursive functions (see e.g. [Ma1], Chapter V, or [Sa]).

(b) Any infinite object of \mathcal{C} is isomorphic to \mathbf{N} .

(c) If U is finite, $\mathcal{C}(U, V)$ consists of all partial maps $U \rightarrow V$. If V is finite, $\mathcal{C}(U, V)$ consists of such f that inverse image of any element of V is enumerable.

Before stating the last condition (d), we make some comments.

Statement (b) is a part of the famous Church Thesis. Any isomorphism (computable bijection) $\mathbf{N} \rightarrow U$ in \mathcal{C} is called a *numbering*. Thus, two different numberings of the same constructive world differ by a recursive permutation of \mathbf{N} . We will call such numberings equivalent ones. Notice that because of (c) two finite constructive worlds are isomorphic iff they have the same cardinality, and the automorphism group of any finite U consists of all permutations of U .

As a matter of principle, we always consider \mathcal{C} as an open category, and at any moment allow ourselves to add to it new constructive worlds. If some infinite V is added to \mathcal{C} , it must come together with a class of equivalent numberings. Thus, any finite union of constructive worlds can be naturally turned into the constructive world, so that the embeddings become computable morphisms, and their images are decidable. As another example, the world \mathbf{N}^* of finite sequences of numbers from \mathbf{N} (“words in alphabet \mathbf{N} ”) is endowed with Gödel’s numbering

$$(1) \quad (n_1, n_2, \dots, n_k, \dots) \mapsto 2^q 3^{n_1-1} \dots p_{k+1}^{n_k-1} - 1$$

where p_k is the k -th prime number, $q = \max\{i \mid n_k = \dots = n_{k-i+1} = 1\}$. Hence we may assume that \mathcal{C} is closed with respect to the construction $U \mapsto U^*$. All natural functions, such as length of the word $U^* \rightarrow \mathbf{N}$, or the i -th letter of the word $U^* \rightarrow U$ are computable.

Similarly, \mathcal{C} can be made closed with respect to the finite direct products by using the (inverse) numbering of \mathbf{N}^2 :

$$(2) \quad (m, n) \mapsto m + \frac{1}{2}(m+n-1)(m+n-2).$$

Projections, diagonal maps, fiber maps $V \rightarrow U \times V, v \mapsto (u_0, v)$ are all computable.

Decidable subsets of constructive worlds are again constructive.

Church Thesis is often invoked as a substitute for an explicit construction of a numbering, and it says that *the category \mathcal{C} is defined uniquely up to equivalence*.

We now turn to the computability properties of the sets of morphisms $\mathcal{C}(U, V)$. Again, it is a matter of principle that $\mathcal{C}(U, V)$ itself *is not a constructive world if U is infinite*. To describe the situation axiomatically, consider first any diagram

$$(3) \quad \text{ev} : P \times U \rightarrow V$$

in \mathcal{C} . It defines a partial map $P \rightarrow \mathcal{C}(U, V)$, $p \mapsto \bar{p}$, where $\bar{p}(u) := \text{ev}(p, u)$. We will say that the constructive world $P = P(U, V)$ together with the evaluation map ev is *a programming method* (for computing some maps $U \rightarrow V$). It is called *universal*, if the following two conditions are satisfied. First, the map $P \rightarrow \mathcal{C}(U, V)$ must be surjective. Second, for any programming method $Q = Q(U, V)$ with the same source U and target V , $\mathcal{C}(Q, P)$ contains translation morphisms

$$(4) \quad \text{trans} : Q(U, V) \rightarrow P(U, V)$$

which are, by definition, everywhere defined, computable maps $Q \rightarrow P$ such that if $q \mapsto p$, then $\bar{q} = \bar{p}$.

We now complete the Definition 1.2 by adding the last axiom forming part of the Church Thesis:

(d) *For every two constructive worlds U, V , there exist universal programming methods.*

The standard examples of P for $U = V = \mathbf{N}$ are (formalized descriptions of) Turing machines, or recursive functions.

From (d) it follows that the composition of morphisms can be lifted to a computable function on the level of programming methods. To be more precise, if Q (resp. P) is a programming method for U, V (resp. V, W), and R is a universal programming method for U, W , there exist computable composition maps

$$(5) \quad \text{comp} : P(V, W) \times Q(U, V) \rightarrow R(U, W), (p, q) \mapsto r$$

such that $\bar{r} = \bar{p} \circ \bar{q}$.

Concrete $P(U, V)$ are furnished by the choice of what is called the “model of computations” in computer science. This last notion comes with a detailed description not only of programs but also of all steps of the computational process. At this stage the models of kinematics and dynamics of the process first emerge, and the discussion of quantization can start.

A formalized description of the first n steps will be called a *history of computation* or, for short, a *protocol* (of length n .) For a fixed model, protocols (of all lengths) form a constructive world as well. We will give two formalized versions of this notion, for functions with infinite and finite domains respectively. The first will be well suited for the discussion of polynomial time computability, the second is the base for quantum computing.

1.3. Models of computations I: normal models. Let U be an infinite constructive world. In this subsection we will be considering partial functions $U \rightarrow U$. The more general case $U \rightarrow V$ can be reduced to this one by working with $U \amalg V$.

A *normal model of computations* is the structure $(P, U, I, F, s,)$ consisting of four sets and a map:

$$(6) \quad I \subset U, F \subset P \times U, s : P \times U \rightarrow P \times U.$$

Here s is an everywhere defined function such that $s(p, u) = (p, s_{\mathbf{p}}(u))$ for any $(p, u) \in P \times U$. Intuitively, p is a program, u is a configuration of the deterministic discrete time computing device, and $s_{\mathbf{p}}(u)$ is the new configuration obtained from u after one unit of time (clock tick). Two additional subsets $I \subset U$ (initial configurations, or inputs) and $F \subset P \times U$ (final configurations) must be given, such that if $(p, u) \in F$, then $s(p, u) = (p, u)$ i.e., u is a fixed point of $s_{\mathbf{p}}$.

In this setting, we denote by $f_{\mathbf{p}}$ the partial function $f_{\mathbf{p}} : I \rightarrow U$ such that we have

$$(7) \quad u \in D(f_{\mathbf{p}}) \text{ and } f_{\mathbf{p}}(u) = v \text{ iff for some } n \geq 0, (p, s_{\mathbf{p}}^n(u)) \in F \text{ and } s_{\mathbf{p}}^n(u) = v.$$

The minimal such n will be called the *time* (number of clock ticks) needed to calculate $f_{\mathbf{p}}(u)$ using the program p .

Any finite sequence

$$(8) \quad (p, u, s_{\mathbf{p}}(u), \dots, s_{\mathbf{p}}^m(u)), u \in I,$$

will be called a *protocol of computation of length m* .

We now add the constructivity conditions.

We require P, U to be constructive worlds, s computable. In addition, we assume that I, F are decidable subsets of $U, P \times U$ respectively. Then f_p are computable, and protocols of given length (resp. of arbitrary length, resp. or those stopping at F), form constructive worlds. If we denote by Q the world of protocols stopping at F and by $\text{ev} : Q \times U \rightarrow U$ the map $(p, u) \mapsto s_p^{\text{max}}(u)$, we get a programming method.

Such a model is called *universal*, if the respective programming method is universal.

The notion of normal model of computations generalizes both normal algorithms and Turing machines. For their common treatment see e.g. [Sa], Chapter 4. In broad terms, $p \in P$ is the list of Markov substitutions, or the table defining the operation of a Turing machine. The remaining worlds U, I, F consist of various words over the working alphabet.

1.3.1. Claim. *For any U , universal normal models of computations exist, and can be effectively constructed.*

For $U = \mathbf{N}$, this follows from the existence of universal Turing machines, and generally, from the Church Thesis. It is well known that the universal machine for calculating functions of k arguments is obtained by taking an appropriate function of $k + 1$ arguments and making the first argument the variable part of the program. Hence P , in this case, consists of pairs (q, m) , where q is the fixed program of the $(k + 1)$ -variable universal function (hardware) and m is a word written on the tape (software).

1.4. Models of computations II: Boolean circuits. Boolean circuits are classical models of computation well suited for studying maps between the finite sets whose elements are encoded by sequences of 0's and 1's.

Consider the Boolean algebra \mathbf{B} generated over \mathbf{F}_2 by a countable sequence of independent variables, say x_1, x_2, x_3, \dots . This is the quotient algebra of $\mathbf{F}_2[x_1, x_2, \dots]$ with respect to the relations $x_i^2 = x_i$. Each Boolean polynomial determines a function on $\bigoplus_{i=1}^{\infty} \mathbf{F}_2$ with values in $\mathbf{F}_2 = \{0, 1\}$.

We start with the following simple fact.

1.4.1. Claim. *Any map $f : \mathbf{F}_2^m \rightarrow \mathbf{F}_2^n$ can be represented by a unique vector of Boolean polynomials.*

Proof. It suffices to consider the case $n = 1$. Then f is represented by

$$(9) \quad F(x_1, \dots, x_n) := \sum_{y=(y_i) \in \mathbf{F}_2^m} f(y) \prod_i (x_i + y_i + 1)$$

because the product in (9) is the delta function in x supported by y . Moreover, the spaces of maps and of Boolean polynomials have the common dimension 2^m over \mathbf{F}_2 .

Now we can calculate any vector of Boolean polynomials iterating operations from a small finite list, which is chosen and fixed, e.g. $\mathcal{B} := \{x, 1, x + y, xy, (x, x)\}$. Such operators are called *classical gates*. A sequence of such operators, together with

indication of their arguments from the previously computed bits, is called a *Boolean circuit*. The number of steps in such a circuit is considered as (a measure of) the time of computation.

When the relevant finite sets are not \mathbf{F}_2^m and perhaps have a wrong cardinality, we encode their elements by finite sequences of bits and consider the restriction of the Boolean polynomial to the relevant subset.

As above, a protocol of computation in this model can be represented as the finite table consisting of rows (generally of variable length) which accommodate sequences of 0's and 1's. The initial line of the table is the input. Each subsequent line must be obtainable from the previous one by the application of one the basic functions in \mathcal{B} to the sequence of neighboring bits (the remaining bits are copied unchanged). The last line is the output. The exact location of the bits which are changed in each row and the nature of change must be a part of the protocol.

Physically, one can implement the rows as the different registers of the memory, or else as the consecutive states of the same register (then we have to make a prescription for how to cope with the variable length, e.g. using blank symbols).

1.4.2. Turing machines vs Boolean circuits. Any protocol of the Turing computation of a function can be treated as such a protocol of an appropriate Boolean circuit, and in this case we have only one register (the initial part of the tape) whose states are consecutively changed by the head/processor. We will still use the term “gate” in this context.

A computable function f with infinite domain is the limit of a sequence of functions f_i between finite sets whose graphs extend each other. A Turing program for f furnishes a computable sequence of Boolean circuits, which compute all f_i in turn. Such a sequence is sometimes called *uniform*.

1.5. Size, complexity, and polynomial time computability. The quantitative theory of computational models deals simultaneously with the space and time dimensions of protocols. The preceding subsection focused on time, here we introduce space. For Boolean (and Turing machine) protocols this is easy: the length of each row of the protocol is the space required at that moment (plus several more bits for specifying the next gate). The maximum of these lengths is the total space required.

The case of normal models and infinite constructive worlds is more interesting.

Generally we will call a *size function* $U \rightarrow \mathbf{N} : u \rightarrow |u|$ any function such that for every $B \in \mathbf{N}$, there are only finitely many objects with $|u| \leq B$. Thus the number of bits $|n| = \lceil \log_2 n \rceil + 1$ and the identical function $\|n\| = n$ are both size functions. Using a numbering, we can transfer them to any constructive world. In these two examples, the number of constructive objects of size $\leq H$ grows as $\exp cH$, resp. cH . Such a count in more general cases allows one to make a distinction between *the bit size*, measuring the length of a description of the object, and *the volume* of the object.

In most cases we require computability of size functions. However, there are exceptions: for example, Kolmogorov complexity is a non-computable size function with very important properties: see below and sec. 5.

Given a size function (on all relevant worlds) and a normal model of computations \mathcal{S} , we can consider the following complexity problems.

(A) For a given morphism (computable map) $f : U \rightarrow V$, estimate the smallest size $K_{\mathcal{S}}(f)$ of the program p such that $f = f_p$.

Kolmogorov, Solomonoff and Chaitin proved that there exists an *optimal* universal model of computations \mathcal{U} such that, with $P = \mathbf{N}$ and the bit size function, for any other model \mathcal{S} there exists a constant c such that for any f

$$K_{\mathcal{U}}(f) \leq K_{\mathcal{S}}(f) + c.$$

When \mathcal{U} is chosen, $K_{\mathcal{U}}(f)$ is called Kolmogorov's complexity of f . With a different choice of \mathcal{U} we will get the same complexity function up to $O(1)$ -summand.

This complexity measure is highly non-trivial (and especially interesting) for an one-element world U and infinite V . It measures then the size of the most compressed description of a variable constructive object in V . This complexity is quite "objective" being almost independent of any arbitrary choices. Being uncomputable, it cannot be directly used in computer science. However, it furnishes some basic restrictions on various complexity measures, somewhat similar to those provided by the conservation laws in physics.

On \mathbf{N} we have $K_{\mathcal{U}}(n) \leq |n| + O(1) = \log_2 \|n\| + O(1)$. The first inequality "generically" can be replaced by equality, but infinitely often $K_{\mathcal{U}}(n)$ becomes much smaller than $|n|$.

(B) For a given morphism (recursive map) $f : U \rightarrow V$, estimate the time needed to calculate $f(u), u \in D(f)$ using the program p and compare the results for different p and different models of computations.

(C) The same for the function "maximal size of intermediate configurations in the protocol of the computation of $f(u)$ using the program p " (space, or memory).

In the last two problems, we have to compare functions rather than numbers: time and space depend on the size of input. Here a cruder polynomial scale appears naturally. Let us show how this happens.

Fix a computational model \mathcal{S} with the transition function s computing functions $U \rightarrow U$, and choose a bit size function on U satisfying the following crucial assumption:

(•) $|u| - c \leq |s_p(u)| \leq |u| + c$ where the constant c may depend on p but not on u .

In this case we have $|s_p^m(u)| \leq |u| + c_p m$: the required space grows no more than linearly with time.

Let now (\mathcal{S}', s') be another model such that $s_p = s'_q$ for some q . For example, such q always exists if \mathcal{S}' is universal. Assume that s' satisfies (•) as well, and additionally

(••) s can be computed in the model \mathcal{S}' in time bounded by a polynomial F in the size of input.

This requirement is certainly satisfied for Turing and Markov models, and is generally reasonable, because an elementary step of an algorithm deserves its name only if it is computationally tractable.

Then we can replace one application of s_p to $s_p^m(u)$ by $\leq F(|u| + cm)$ applications of s'_q . And if we needed $T(u)$ steps in order to calculate $f_p(u)$ using \mathcal{S} , we will need

no more than $\leq \sum_{m=1}^{T(u)} F(|u| + cm)$ steps to calculate the same function using \mathcal{S}' and q .

In a detailed model, there might be a small additional cost of merging two protocols. This is an example of the translation morphism (4) lifted to the worlds of protocols.

Thus, from (\bullet) and $(\bullet\bullet)$ it follows that functions computable in polynomial time by \mathcal{S} have the same property for all reasonable models. Notice also that for such functions, $|f(u)| \leq G(|u|)$ for some polynomial G and that the domain $D(f)$ of such a function is decidable: if after $T(|u|)$ $s_{\mathcal{P}}$ -steps we are not in a final state, then $u \notin D(f)$.

Thus we can define the class PF of functions, say, $\mathbf{N}^k \rightarrow \mathbf{N}$ computable in polynomial time by using a fixed universal Turing machine and arguing as above that this definition is model-independent.

If we want to extend it to a constructive universe \mathcal{C} however, we will have to postulate additionally that any constructive world U comes together with a natural class of numberings which, together with their inverses, are computable in polynomial time. This seems to be a part of the content of the “*polynomial Church thesis*” invoked by M. Freedman in [Fr1]. If we take this strengthening of the Church thesis for granted, then we can define also the bit size of an arbitrary constructive object as the bit size of its number with respect to one of these numberings. The quotient of two such size functions is bounded from above and from zero.

Below we will be considering only the universes \mathcal{C} and worlds U with these properties, and $|u|$ will always denote one of the bit size norms. Gödel’s numbering (2) for $\mathbf{N} \times \mathbf{N}$ shows that such \mathcal{C} is still closed with respect to finite products. (Notice however that the beautiful numbering (3) of \mathbf{N}^* using primes is not polynomial time computable; it may be replaced by another one which is in PF).

1.6. P/NP problem. By definition, a subset $E \subset U$ belongs to the class P iff its characteristic function χ_E (equal to 1 on E and 0 outside) belongs to the class PF . Furthermore, $E \in U$ belongs to the class NP iff there exists a subset $E' \subset U \times V$ belonging to P and a polynomial G such that

$$u \in E \iff \exists (u, v) \in E' \text{ with } |v| \leq G(|u|).$$

Here V is another world (which may coincide with U). We will say that E is obtained from E' by a *polynomially truncated projection*.

The discussion above establishes in what sense this definition is model independent.

Clearly, $P \subset NP$. The inverse inclusion is highly problematic. A naive algorithm calculating χ_E from $\chi_{E'}$ by searching for v with $|v| \leq G(|u|)$ and $\chi_{E'}(u, v) = 1$ will take exponential time e.g. when there is no such v (because $|u|$ is a bit size function). Of course, if one can treat all such v in parallel, the required time will be polynomial. Or else, if an oracle tells you that $u \in E$ and supplies an appropriate v , you can convince yourself that this is indeed so in polynomial time, by computing $\chi_{E'}(u, v) = 1$.

Notice that the enumerable sets can be alternatively described as projections of decidable ones, and that in this context projection does create undecidable sets. Nobody was able to translate the diagonalization argument used to establish this to

the P/NP domain. M. Freedman ([Fr2]) suggested an exciting new approach to the problem $P \neq NP(?)$, based upon a modification of Gromov's strategy for describing groups of polynomial growth.

It has long been known that this problem can be reduced to checking whether some very particular sets – NP -complete ones – belong to P . The set $E \subset U$ is called NP -complete if, for any other set $D \subset V, D \in NP$, there exists a function $f : V \rightarrow U, f \in PF$, such that $D = f^{-1}(E)$, that is, $\chi_D(v) = \chi_E(f(v))$. We will sketch the classical argument (due to S. Cooke, L. Levin, R. Karp) showing the existence of NP -complete sets. In fact, the reasoning is constructive: it furnishes a polynomially computable map producing f from the descriptions of $\chi_{E'}$ and of the truncating polynomial G .

In order to describe one NP -complete problem, we will define an infinite family of Boolean polynomials b_u indexed by the following data, constituting objects u of the constructive world U . One u is a collection

$$(10) \quad m \in \mathbf{N}; \quad (S_1, T_1), \dots, (S_N, T_N),$$

where $S_i, T_i \subset \{1, \dots, m\}$, and b_u is defined as

$$(11) \quad b_u(x_1, \dots, x_m) = \prod_{i=1}^N \left(1 + \prod_{k \in S_i} (1 + x_k) \prod_{j \in T_i} x_j \right).$$

The size of (10) is by definition $|u| = mN$.

Put

$$E = \{u \in U \mid \exists v \in \mathbf{F}_2^m, b_u(v) = 1\}.$$

Using the language of Boolean truth values, one says that v satisfies b_u if $b_u(v) = 1$, and E is called the *satisfiability problem*, or *SAT*.

1.6.1. Claim. $E \in NP$.

In fact, let

$$(12) \quad E' = \{(u, v) \mid b_u(v) = 1\} \subset U \times \left(\bigoplus_{i=1}^{\infty} \mathbf{F}_2 \right).$$

Clearly, E is the full projection of E' . A contemplation will convince the reader that $E' \in P$. In fact, we can calculate $b_u(v)$ performing $O(Nm)$ Boolean multiplications and additions. The projection to E can be replaced by a polynomially truncated projection, because we have to check only v of size $|v| \leq m$.

1.6.2. Claim. E is NP -complete.

In fact, let $D \in NP, D \subset A$ where A is some universe. Take a representation of D as a polynomially truncated projection of some set $D' \subset A \times B, D' \in P$. Choose a normal, say Turing, model of computation and consider the Turing protocols of computation of $\chi_{D'}(a, b)$ with fixed a and variable polynomially bounded b . As we have explained above, for a given a , any such protocol can be imagined as a table of a fixed polynomially bounded size whose rows are the consecutive states of the

computation. In the “microscopic” description, the positions in this table can be filled only by 0 or 1. In addition, each row is supplied by the specification of the position and the inner state of the head/processor. Some of the arrangements are valid protocols, others are not, but the local nature of the Turing computation allows one to produce a Boolean polynomial b_u in appropriate variables such that the valid protocols are recognized by the fact that this polynomial takes value 1. For detailed explanations see e.g. [GaJ], sec. 2.6. This defines the function f reducing D to E . The construction is so direct that the polynomial time computability of f is straightforward.

Many natural problems are known to be NP -complete, in particular 3 -SAT. It is defined as the subset of SAT consisting of those u for which $\text{card}(S_i \cup T_i) = 3$ for all i .

1.6.3. Remark. Most of Boolean functions are not computable in polynomial time. Several versions of this statement can be proved by simple counting.

First of all, fix a finite basis \mathcal{B} of Boolean operations as in 1.4.1, each acting upon $\leq a$ bits. Then sequences of these operations of length t generate $O((bn^a)^t)$ Boolean functions $\mathbf{F}_2^n \rightarrow \mathbf{F}_2^n$ where $b = \text{card } \mathcal{B}$. On the other hand, the number of all functions 2^{n2^n} grows as a double exponential of n and for large n cannot be obtained in time t polynomially bounded in n .

The same conclusion holds if we consider not all functions but only permutations: Stirling’s formula for $\text{card } S_{2^n} = 2^n!$ involves a double exponential.

Here is one more variation of this problem: define the time complexity of a conjugacy class in S_{2^n} as the minimal number of steps needed to calculate some permutation in this class. This notion arises if we are interested in calculating automorphisms of a finite universe of cardinality 2^n , which is not supplied with a specific encoding by binary words. Then it can happen that a judicious choice of encoding will drastically simplify the calculation of a given function. However, for most functions we still will not be able to achieve polynomial type computability, because the asymptotical formula for the number of conjugacy classes (partitions)

$$p(2^n) \sim \frac{\exp\left(\pi \sqrt{\frac{2}{3}\left(2^n - \frac{1}{24}\right)}\right)}{4\sqrt{3}\left(2^n - \frac{1}{24}\right)}$$

again displays the double exponential growth.

2. QUANTUM PARALLELISM

In this section we will discuss the basics: how to use the superposition principle in order to accelerate (certain) classical computations.

2.1. Description of the problem. Let N be a large number, $F : \{0, \dots, N-1\} \rightarrow \{0, \dots, N-1\}$ a function such that the computation of each particular value $F(x)$ is tractable, that is, can be done in time polynomial in $\log x$. We want to compute (to recognize) some property of the graph $(x, F(x))$, for example:

(i) Find the least period r of F , i.e., the least residue $r \bmod N$ such that $F(x+r \bmod N) = F(x)$ for all x (the key step in the Factorization Problem).

(ii) Find some x such that $F(x) = 1$ or establish that such x does not exist (Search Problem).

As we already mentioned, the direct attack on such a problem consists in compiling the complete list of pairs $(x, F(x))$ and then applying to it an algorithm recognizing the property in question. Such a strategy requires at least exponential time (as a function of the bit size of N) since already the length of the list is N . Barring a theoretical breakthrough in understanding such problems (for example a proof that $P = NP$), a practical response might be in exploiting the possibility of parallel computing, i.e., calculating simultaneously many – or even all – values of $F(x)$. This takes less time but uses (dis)proportionally more hardware.

A remarkable suggestion due to D. Deutsch (see [DeuJ], [Deu]) consists in using a quantum superposition of the classical states $|x\rangle$ as the replacement of the union of N classical registers, each in one of the initial states $|x\rangle$. To be more precise, here is a mathematical model formulated as the definition.

2.2. Quantum parallel processing: version I. *Keeping the notation above, assume moreover that $N = 2^n$ and that F is a bijective map (the set of all outputs is a permutation of the set of all inputs).*

(i) *The quantum space of inputs/outputs is the 2^n -dimensional complex Hilbert space H_n with the orthonormal basis $|x\rangle$, $0 \leq x \leq N - 1$. Vectors $|x\rangle$ are called classical states.*

(ii) *The quantum version of F is the unique unitary operator $U_F : H_n \rightarrow H_n$ such that $U_F|x\rangle = |F(x)\rangle$.*

Quantum parallel computing of F is (a physical realization of) a system with the state space H_n and the evolution operator U_F .

Naively speaking, if we apply U_F to the initial state which is a superposition of all classical states with, say, equal amplitudes, we will get simultaneously all classical values of F (i.e., their superposition):

$$(14) \quad U_F \left(\frac{1}{\sqrt{N}} \sum |x\rangle \right) = \frac{1}{\sqrt{N}} \sum |F(x)\rangle.$$

We will now discuss various issues related to this definition, before passing to its more realistic modification.

(A) We put $N = 2^n$ above because we are imagining the respective classical system as an n -bit register: cf. the discussion of Boolean circuits. Every number $0 \leq x \leq N - 1$ is written in the binary notation $x = \sum_i \epsilon_i 2^i$ and is identified with the pure (classical) state $|\epsilon_{n-1}, \dots, \epsilon_0\rangle$ where $\epsilon_i = 0$ or 1 is the state of the i -th register. The quantum system H_1 is called *qubit*. We have $H_n = H_1^{\otimes n}$, $|\epsilon_{n-1}, \dots, \epsilon_0\rangle = |\epsilon_{n-1}\rangle \otimes \dots \otimes |\epsilon_0\rangle$.

This conforms to the general principles of quantum mechanics. The Hilbert space of the union of systems can be identified with the tensor product of the Hilbert spaces of the subsystems. Accordingly, decomposable vectors correspond to the states of the compound for which one can say that the individual subsystems are in definite states.

(B) Pure quantum states, strictly speaking, are points of the projective space $P(H_n)$ that is, complex lines in H_n . Traditionally, one considers instead vectors of norm one. This leaves undetermined an overall phase factor $\exp i\varphi$. If we have two state vectors, individual phase factors have no objective meaning, but their quotient, that is the difference of their phases, does have one. This difference can be measured by observing effects of interference. This possibility is used for implementing efficient quantum algorithms.

(C) If a quantum system S is isolated, its dynamical evolution is described by the unitary operator $U(t) = \exp iHt$ where H is the Hamiltonian, t is time. Therefore one option for implementing U_F physically is to design a device for which U_F would be a fixed time evolution operator. However, this seemingly contradicts many deeply rooted notions of the algorithm theory. For example, calculating $F(x)$ for different inputs x takes different times, and it would be highly artificial to try to equalize them already in the design.

Instead, one can try to implement U_F as the result of a sequence of brief interactions, carefully controlled by a classical computer, of S with environment (say, laser pulses). Mathematically speaking, U_F is represented as a product of some standard unitary operators $U_m \dots U_1$ each of which acts only on a small subset (two, three) of classical bits. These operators are called *quantum gates*.

The complexity of the respective quantum computation is determined by its length (the number m of the gates) and by the complexity of each of them. The latter point is a subtle one: continuous parameters, e.g. phase shifts, on which U_i may depend, makes the information content of each U_i potentially infinite and leads to a suspicion that a quantum computer will in fact perform an analog computation, only implemented in a fancy way. A very interesting discussion in [Ts], Lecture 9, convincingly refutes this viewpoint, by displaying those features of quantum computation which distinguish it from both analog and digital classical information processing. This discussion is based on the technique of fault tolerant computing using quantum codes for producing continuous variables highly protected from external noise.

(D) From the classical viewpoint, the requirement that F must be a permutation looks highly restrictive (for instance, in the search problem F takes only two values). Physically, the reason for this requirement is that only such F extend to unitary operators (“quantum reversibility”). The standard way out consists of introducing *two* n -bit registers instead of one, for keeping the value of the argument as well as that of the function. More precisely, if $F(|x\rangle)$ is an arbitrary function, we can replace it by the permutation $\tilde{F}(|x, y\rangle) := |x, F(x) \oplus y\rangle$, where \oplus is the Boolean (bitwise) sum. This involves no more than a polynomial increase of the classical complexity, and the restriction of \tilde{F} to $y = 0$ produces the graph of F which we need anyway for the type of problems we are interested in.

In fact, in order to process a classical algorithm (sequence of Boolean gates) for computing F into the quantum one, we replace each classical gate by the respective reversible quantum gate, i.e., by the unitary operator corresponding to it tensored by the identical operator. Besides two registers for keeping $|x\rangle$ and $F(|x\rangle)$ this trick introduces as well extra qubits in which we are not particularly interested. The corre-

sponding space and its content are sometimes referred to as “scratchpad”, “garbage”, etc. Besides ensuring reversibility, additional space and garbage can be introduced as well for considering functions $F : \{0, \dots, N - 1\} \rightarrow \{0, \dots, M - 1\}$ where N, M are not powers of two (then we extend them to the closest power of two). For more details, see the next section.

Notice that the choice of gate array (Boolean circuit) as the classical model of computation is essential in the following sense: a quantum routine cannot use conditional instructions. Indeed, to implement such an instruction we must observe the memory in the midst of calculation, but the observation generally will change its current quantum state.

In the same vein, we must avoid copying instructions, because the classical copying operator $|x\rangle \rightarrow |x\rangle \otimes |x\rangle$ is not linear. In particular, each output qubit from a quantum gate can be used only in one gate at the next step (if several gates are used parallelly): cloning is not allowed.

These examples show that the basics of quantum code writing will have a very distinct flavor.

We now pass to the problems posed by the input/output routines.

Input, or initialization, in principle can be implemented in the same way as a computation: we produce an input state starting e.g. from the classical state $|0\rangle$ and applying a sequence of basic unitary operators: see the next section. Output, however, involves an additional quantum mechanical notion: that of *observation*.

(E) The simplest model of observation of a quantum system with the Hilbert space H involves the choice of an orthonormal basis of H . Only elements of this basis $|\chi_i\rangle$ can appear as the results of observation. If our system is in some state $|\psi\rangle$ at the moment of observation, it will be observed in the state $|\chi_i\rangle$ with probability $|\langle \chi_i | \psi \rangle|^2$.

This means first of all that every quantum computation is inherently probabilistic. Observing (a part of) the quantum memory is not exactly the same as “printing the output”. We must plan a series of runs of the same quantum program and the subsequent classical processing of the observed results, and we can hope only to get the desired answer with probability close to one.

Furthermore, this means that by implementing quantum parallelism simply and as in (14), and then observing the memory as if it were the classical n -bit register, we will simply get some value $F(x)$ with probability $1/N$. This does not use the potential of the quantum parallelism. Therefore we formulate a corrected version of this notion, leaving more flexibility and stressing the additional tasks of the designer, each of which eventually contributes to the complexity estimate.

2.3. Quantum parallel processing: version II. *To solve efficiently a problem involving properties of the graph of a function F , we must design:*

(i) *An auxiliary unitary operator U carrying the relevant information about the graph of F .*

(ii) *A computationally feasible realization of U with the help of standard quantum gates.*

(iii) *A computationally feasible realization of the input subroutine.*

(iv) *A computationally feasible classical algorithm processing the results of many runs of quantum computation.*

All of this must be supplemented by quantum error-correcting encoding, which we will not address here. In the next section we will discuss some standard quantum subroutines.

3. SELECTED QUANTUM SUBROUTINES

3.1. Initialization. Using the same conventions as in (14) and the subsequent comments, in particular, the identification $H_n = H_1^{\otimes n}$, we have

$$(15) \quad \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle = \frac{1}{\sqrt{N}} \sum_{\epsilon_i=0,1} |\epsilon_{n-1} \dots \epsilon_0\rangle = \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \right)^{\otimes n}.$$

In other words,

$$(16) \quad \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle = U_1^{(n-1)} \dots U_1^{(0)} |0 \dots 0\rangle$$

where $U_1 : H_1 \rightarrow H_1$ is the unitary operator

$$|0\rangle \mapsto \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |1\rangle \mapsto \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle),$$

and $U_1^{(i)} = \text{id} \otimes \dots \otimes U_1 \otimes \dots \otimes \text{id}$ acts only on the i -th qubit.

Thus making the quantum gate U_1 act on each memory bit, one can in n steps initialize our register in the state which is the superposition of all 2^n classical states with equal weights.

3.2. Quantum computations of classical functions. Let \mathcal{B} be a finite basis of classical gates containing one-bit identity and generating all Boolean circuits, and $F : \mathbf{F}_2^m \rightarrow \mathbf{F}_2^n$ a function. We will describe how to turn a Boolean circuit of length L calculating F into another Boolean circuit of comparable length consisting only of reversible gates, and calculating a modified function, which however contains all information about the graph of F . Reversibility means that each step is a bijection (actually, an involution) and hence can be extended to a unitary operator, that is, a quantum gate. For a gate f , define $\tilde{f}(|x, y\rangle) = |x, f(x) + y\rangle$ as in 2.2(D) above.

3.2.1. Claim. *A Boolean circuit \mathcal{S} of length L in the basis \mathcal{B} can be processed into the reversible Boolean circuit $\tilde{\mathcal{S}}$ of length $O((L + m + n)^2)$ calculating a permutation $H : \mathbf{F}_2^{m+n+L} \rightarrow \mathbf{F}_2^{m+n+L}$ with the following property:*

$$H(x, y, 0) = (x, F(x) + y, 0) = (\tilde{F}(x, y), 0).$$

Here x, y, z have sizes m, n, L respectively.

Proof. We will understand L here as the sum of sizes of the outputs of all gates involved in the description of \mathcal{S} . We first replace in \mathcal{S} each gate f by its reversible counterpart \tilde{f} . This involves inserting extra bits which we put side by side into a new register of total length L . The resulting subcircuit will calculate a permutation $K : \mathbf{F}_2^{m+L} \rightarrow \mathbf{F}_2^{m+L}$ such that $K(x, 0) = (F(x), G(x))$ for some function G (garbage).

Now add to the memory one more register of size n keeping the variable y . Extend K to the permutation $\overline{K} : \mathbf{F}_2^{m+L+n} \rightarrow \mathbf{F}_2^{m+L+n}$ keeping y intact: $\overline{K} : (x, 0, y) \mapsto (F(x), G(x), y)$. Clearly, \overline{K} is calculated by the same Boolean circuit as K , but with extended register.

Extend this circuit by the one adding the contents of the first and the third register: $(F(x), G(x), y) \mapsto (F(x), G(x), F(x) + y)$. Finally, build the last extension which calculates \overline{K}^{-1} and consists of reversed gates calculating \overline{K} in reverse order. This clears the middle register (scratchpad) and produces $(x, 0, F(x) + y)$. The whole circuit requires $O(L+m+n)$ gates if we allow the application of them to not necessarily neighboring bits. Otherwise we must insert gates for local permutations which will replace this estimate by $O((L+m+n)^2)$.

3.3. Fast Fourier transform. Finding the least period of a function of one real variable can be done by calculating its Fourier transforms and looking at its maxima. The same strategy is applied by Shor in his solution of the factorization problem. We will show now that the discrete Fourier transform Φ_n is computationally easy (quantum polynomial time). We define $\Phi_n : H_n \rightarrow H_n$ by

$$(17) \quad \Phi_n(|x\rangle) = \frac{1}{\sqrt{N}} \sum_{c=0}^{N-1} |c\rangle \exp(2\pi icx/N)$$

In fact, it is slightly easier to implement directly the operator

$$(18) \quad \Phi_n^t(|x\rangle) = \frac{1}{\sqrt{N}} \sum_{c=0}^{N-1} |c^t\rangle \exp(2\pi icx/N).$$

where c^t is c read from the right to the left. The effects of the bit reversal can be then compensated at a later stage without difficulty.

Let $U_2^{(kj)} : H_n \rightarrow H_n$, $k < j$, be the quantum gate which acts on the pair of the k -th and j -th qubits in the following way: it multiplies $|11\rangle$ by $\exp(i\pi/2^{j-k})$ and leaves the remaining classical states $|00\rangle, |01\rangle, |10\rangle$ intact.

3.3.1. Lemma *We have*

$$(19) \quad \Phi_n^t = \prod_{k=0}^{n-1} \left(U_1^{(k)} \prod_{j=k+1}^{n-1} U_2^{(kj)} \right).$$

By our rules of the game, (19) has polynomial length in the sense that it involves only $O(n^2)$ gates. However, implementation of $U_2^{(kj)}$ requires controlling variable

phase factors which tend to 1 as $k - j$ grows. Moreover, arbitrary pairs of qubits must allow quantum mechanical coupling so that for large n the interaction between qubits must be non-local. The contribution of these complications to the notion of complexity cannot be estimated without going into the details of physical arrangement. Therefore I will add a few words to this effect.

The implementation of quantum register suggested in [CZ] consists of a collection of ions (charged atoms) in a linear harmonic trap (optical cavity). Two of the electronic states of each ion are denoted $|0\rangle$ and $|1\rangle$ and represent a qubit. Laser pulses transmitted to the cavity through the optical fibers and controlled by the classical computer are used to implement gates and read out. The Coulomb repulsion keeps ions apart (spatial selectivity) which allows the preparation of each ion separately in any superposition of $|0\rangle$ and $|1\rangle$ by timing the laser pulse properly and preparing its phase carefully. The same Coulomb repulsion allows for collective excitations of the whole cluster whose quanta are called phonons. Such excitations are produced by laser pulses as well under appropriate resonance conditions. The resulting resonance selectivity combined with the spatial selectivity implements a controlled entanglement of the ions that can be used in order to simulate two and three bit gates. For a detailed and lucid mathematical explanation, see [Ts], Lecture 8.

Another recent suggestion ([GeC]) is to use a single molecule as a quantum register, representing qubits by nuclear spins of individual atoms, and using interactions through chemical bonds in order to perform multiple bit logic. The classical technique of nuclear magnetic resonance developed since the 1940's, which allows one to work with many molecules simultaneously, provides the start up technology for this project.

3.4. Quantum search. All the subroutines described up to now boiled down to some identities in the unitary groups involving products of not too many operators acting on subspaces of small dimension. They did not involve output subroutines and therefore did not “compute” anything in the traditional sense of the word. We will now describe the beautiful quantum search algorithm due to L. Grover which produces a new identity of this type, but also demonstrates the effect of observation and the way one can use quantum entanglement in order to exploit the potential of quantum parallelism.

We will treat only the simplest version. Let $F : \mathbf{F}_2^n \rightarrow \{0, 1\}$ be a function taking the value 1 at exactly one point x_0 . We want to compute x_0 . We assume that F is computable in polynomial time, or else that its values are given by an oracle. Classical search for x_0 requires on the average about $N/2$ evaluations of F where $N = 2^n$.

In the quantum version, we will assume that we have a quantum Boolean circuit (or quantum oracle) calculating the unitary operator $H_n \rightarrow H_n$

$$I_F : |x\rangle \mapsto e^{\pi i F(x)} |x\rangle.$$

In other words, I_F is the reflection inverting the sign of $|x_0\rangle$ and leaving the remaining classical states intact.

Moreover, we put $J = -I_\delta$, where $\delta : \mathbf{F}_2^n \rightarrow \{0, 1\}$ takes the value 1 only at 0, and $V = U_1^{(n-1)} \dots U_1^{(0)}$, as in (16).

3.4.1. Claim. (i) *The real plane in H_n spanned by the uniform superposition ξ of all classical states (15) and by $|x_0\rangle$ is invariant with respect to $T := VJVIF$.*

(ii) *T restricted to this plane is the rotation (from ξ to $|x_0\rangle$) by the angle φ_N where*

$$\cos \varphi_N = 1 - \frac{2}{N}, \quad \sin \varphi_N = 2 \frac{\sqrt{N-1}}{N}.$$

The check is straightforward.

Now, φ_N is close to $\frac{2}{\sqrt{N}}$, and for the initial angle φ between ξ and $|x_0\rangle$ we have

$$\cos \varphi = -\frac{1}{\sqrt{N}}.$$

Hence in $[\varphi/\varphi_N] \approx \frac{\pi\sqrt{N}}{4}$ applications of T to ξ we will get the state very close to $|x_0\rangle$. Stopping the iteration of T after as many steps and measuring the outcome in the basis of classical states, we will obtain $|x_0\rangle$ with probability very close to one.

One application of T replaces in the quantum search one evaluation of F . Thus, thanks to quantum parallelism, we achieve a polynomial speed-up in comparison with the classical search. The case when F takes value 1 at several points and we only want to find one of them, can be treated by an extension of this method. If there are n such points, the algorithm requires about $\sqrt{N/n}$ steps, and n needs not be known a priori: see [BoyBHT].

4. SHOR'S FACTORING ALGORITHM

4.1. Notation. Let M be a number to be factored. We will assume that it is odd and is not a power of a prime number.

Denote by N the size of the basic memory register we will be using (not counting scratchpad). Its bit size n will be about twice that of M . More precisely, choose $M^2 < N = 2^n < 2M^2$. Finally, let $1 < t < M$ be a random parameter with $\gcd(t, M) = 1$. This condition can be checked classically in time polynomial in n .

Below we will describe one run of Shor's algorithm, in which t (and of course, M, N) is fixed. Generally, polynomially many runs will be required, in which the value of t can remain the same or be chosen anew. This is needed in order to gather statistics. Shor's algorithm is a probabilistic one, with two sources of randomness that must be clearly distinguished. One is built into the classical probabilistic reduction of factoring to the finding of the period of a function. Another stems from the necessity of observing quantum memory, which, too, produces random results.

More precise estimates than those given here show that a quantum computer which can store about $3n$ qubits can find a factor of M in time of order n^3 with probability close to 1 : see [BCDP]. On the other hand, it is widely believed that no recursive function of the type $M \mapsto a$ *proper factor of M* belongs to PF . This is why the most popular public key encryption schemes rely upon the difficulty of the factoring problem.

4.2. Classical algorithm. Put

$$r := \min \{ \rho \mid t^\rho \equiv 1 \pmod{M} \}$$

which is the least period of $F : a \mapsto t^a \pmod{M}$.

4.2.1. Claim. *If one can efficiently calculate r as a function of t , one can find a proper divisor of M in polynomial in $\log_2 M$ time with probability $\geq 1 - M^{-m}$ for any fixed m .*

Assume that for a given t the period r satisfies

$$r \equiv 0 \pmod{2}, \quad t^{r/2} \not\equiv -1 \pmod{M}.$$

Then $\gcd(t^{r/2} + 1, M)$ is a proper divisor of M . Notice that \gcd is computable in polynomial time.

The probability that this condition holds is $\geq 1 - \frac{1}{2^{k-1}}$ where k is the number of different odd prime divisors of M , hence $\geq \frac{1}{2}$ in our case. Therefore we will find a good t with probability $\geq 1 - M^{-m}$ in $O(\log M)$ tries. The longest calculation in one try is that of $t^{r/2}$. The usual squaring method takes polynomial time as well.

4.3. Quantum algorithm calculating r . Here we describe one run of the quantum algorithm which purports to compute r , given M, N, t . We will use the working register that can keep a pair consisting of a variable $0 \leq a \leq N - 1$ and the respective value of the function $t^a \pmod{M}$. One more register will serve as the scratchpad needed to compute $|a, t^a \pmod{M}\rangle$ reversibly. When this calculation is completed, the content of the scratchpad will be reversibly erased: cf. 3.2.1. In the remaining part of the computation the scratchpad will not be used anymore, we can decouple it, and forget about it.

The quantum computation consists of four steps, three of which were described in sec. 3:

- (i) Partial initialization produces from $|0, 0\rangle$ the superposition

$$\frac{1}{\sqrt{N}} \sum_{a=0}^{N-1} |a, 0\rangle.$$

- (ii) Reversible calculation of F processes this state into

$$\frac{1}{\sqrt{N}} \sum_{a=0}^{N-1} |a, t^a \pmod{M}\rangle.$$

- (iii) Partial Fourier transform then furnishes

$$\frac{1}{N} \sum_{a=0}^{N-1} \sum_{c=0}^{N-1} \exp(2\pi i ac/N) |c, t^a \pmod{M}\rangle.$$

(iv) The last step is the observation of this state with respect to the system of classical states $|c, m \bmod M\rangle$. This step produces some concrete output

$$(20) \quad |c, t^k \bmod M\rangle$$

with probability

$$(21) \quad \left| \frac{1}{N} \sum_{a: t^a \equiv t^k \bmod M} \exp(2\pi i ac/N) \right|^2.$$

The remaining part of the run is assigned to the classical computer and consists of the following steps.

(A) Find the best approximation (in lowest terms) to $\frac{c}{N}$ with denominator $r' < M < \sqrt{N}$:

$$(22) \quad \left| \frac{c}{N} - \frac{d'}{r'} \right| < \frac{1}{2N}.$$

As we will see below, we may hope that r' will coincide with r in at least one run among at most polynomially many. Hence we try r' in the role of r right away:

(B) If $r' \equiv 0 \pmod{2}$, calculate $\gcd(t^{r'/2} \pm 1, M)$.

If r' is odd, or if r' is even, but we did not get a proper divisor of M , repeat the run $O(\log \log M)$ times with the same t . In case of failure, change t and start a new run.

4.3.1. Justification. We will now show that, given t , from the observed values of $|c, t^k \bmod M\rangle$ in $O(\log \log M)$ runs we can find the correct value of r with probability close to 1.

Let us call the observed value of c *good*, if

$$\exists l \in \left[-\frac{r}{2}, \frac{r}{2} \right], \quad rc \equiv l \pmod{N}.$$

In this case there exists such d that

$$-\frac{r}{2} \leq rc - dN = l \leq \frac{r}{2}$$

so that

$$\left| \frac{c}{N} - \frac{d}{r} \right| < \frac{1}{2N}.$$

Hence if c is good, then r' found from (22) in fact divides r .

Now call c *very good* if $r' = r$.

Estimating the exponential sum (21), we can easily check that the probability of observing a good c is $\geq \frac{1}{3r^2}$. On the other hand, there are $r\varphi(r)$ states $|c, t^k \bmod M\rangle$

with very good c . Thus to find a very good c with high probability, $O(r^2 \log r)$ runs will suffice.

5. KOLMOGOROV COMPLEXITY AND GROWTH OF RECURSIVE FUNCTIONS

Consider general functions $f : \mathbf{N} \rightarrow \mathbf{N}$. Computability theory uses several growth scales for such functions, of which two are most useful: f may be majorized by some recursive function (e.g. when it is itself recursive), or by a polynomial (e.g. when it is computable in polynomial time). Linear growth does not seem particularly relevant in this context. However, this impression is quite misleading, at least if one allows re-ordering \mathbf{N} . In fact, we have:

5.1. Claim. *There exists a permutation $\mathbf{K} : \mathbf{N} \rightarrow \mathbf{N}$ such that for any partially recursive function $f : \mathbf{N} \rightarrow \mathbf{N}$ there exists a constant c with the property*

$$(23) \quad \mathbf{K} \circ f \circ \mathbf{K}^{-1}(n) \leq cn \text{ for all } n \in \mathbf{K}(D(f)).$$

Moreover, \mathbf{K} is bounded by a linear function, but \mathbf{K}^{-1} is not bounded by any recursive function.

Proof. We will use the Kolmogorov complexity measure. For a recursive function $u : \mathbf{N} \rightarrow \mathbf{N}$, $x \in \mathbf{N}$, put $C_u(x) := \min \{k \mid f(k) = x\}$, or ∞ if such k does not exist. Call such a function u *optimal* if, for any other recursive function v , there exists a constant $c_{u,v}$ such that $C_u(x) \leq c_{u,v} C_v(x)$ for all x . Optimal functions do exist (see e.g. [Ma1], Theorem VI.9.2); in particular, they take all positive integer values (however they certainly are not everywhere defined). Fix one such u and call $C_u(x)$ the (exponential) complexity of x . By definition, $\mathbf{K} = \mathbf{K}_u$ rearranges \mathbf{N} in the order of increasing complexity. In other words,

$$(24) \quad \mathbf{K}(x) := 1 + \text{card} \{y \mid C_u(y) < C_u(x)\}.$$

We first show that

$$(25) \quad \mathbf{K}(x) = \exp(O(1)) C_u(x).$$

Since C_u takes each value at most once, it follows from (24) that $\mathbf{K}(n) \leq C_u(n)$. In order to show that $C_u(x) \leq c \mathbf{K}(x)$ for some c it suffices to check that

$$\text{card} \{k \leq N \mid \exists x, C_u(x) = k\} \geq bN$$

with some $b > 0$. In fact, at least half of the numbers $x \leq N$ have the complexity which is no less than $x/2$.

Now, VI.9.7(b) in [Ma1] implies that, for any recursive function f and all $x \in D(f)$, we have $C_u(f(x)) \leq \text{const} C_u(x)$. Since $C_u(x)$ and $\mathbf{K}(x)$ have the same order of growth up to a bounded factor, our claim follows.

5.2. Corollary. *Denote by S_∞^{rec} the group of recursive permutations of \mathbf{N} . Then $\mathbf{K} S_\infty^{\text{rec}} \mathbf{K}^{-1}$ is a subgroup of permutations of no more than linear growth.*

Actually, appealing to the Proposition VI.9.6 of [Ma1], one can considerably strengthen this result. For example, let σ be a recursive permutation, $\sigma^{\mathbf{K}} = \mathbf{K}\sigma\mathbf{K}^{-1}$. Then $\sigma^{\mathbf{K}}(x) \leq cx$ so that $(\sigma^{\mathbf{K}})^n(x) \leq c^n x$ for $n > 0$. But actually the last inequality can be replaced by

$$(\sigma^{\mathbf{K}})^n(x) \leq c'n$$

for a fixed x and variable n . With both x and n variable one gets the estimate $O(xn \log(xn))$.

In the same way as finite permutations appear in the quantum versions of Boolean circuits, infinite (computable) permutations are natural for treating quantum Turing machines ([Deu]) and our normal computation models. In fact, if one assumes that the transition function s is a permutation, and then extends it to the unitary operator U_s in the infinite-dimensional Hilbert space, one might be interested in studying the spectral properties of such operators. But the latter depend only on the conjugacy class. Perhaps the universal conjugation $U_{\mathbf{K}}$ might be a useful theoretical tool in this context. In the purely classical situation, (23) may play a role in studying the limiting behavior of polynomial time algorithms, as suggested in [Fr1] and [Fr2].

Finally, I would like to comment upon the hidden role of Kolmogorov complexity in the real life of classical computing. The point is that in a sense (which is difficult to formalize), we are interested only in the calculation of sufficiently nice functions, because a random Boolean function will have (super)exponential complexity anyway. A nice function, at the very least, has a short description and, therefore, a small Kolmogorov complexity. Thus, dealing with practical problems, we actually work not with small numbers, graphs, circuits, . . . , but rather with an initial segment of the respective constructive world reordered with the help of \mathbf{K} . We systematically replace a large object by its short description, and then try to overcome the computational difficulties generated by this replacement.

APPENDIX

The following text is a contribution to the prehistory of quantum computing. It is the translation from Russian of the last three paragraphs of the Introduction to [Ma2] (1980). For this reference I am grateful to A. Kitaev [Ki].

“Perhaps, for better understanding of this phenomenon [DNA replication], we need a mathematical theory of quantum automata. Such a theory would provide us with mathematical models of deterministic processes with quite unusual properties. One reason for this is that the quantum state space has far greater capacity than the classical one: for a classical system with N states, its quantum version allowing superposition accommodates c^N states. When we join two classical systems, their number of states N_1 and N_2 are multiplied, and in the quantum case we get exponential growth $c^{N_1 N_2}$.

These crude estimates show that the quantum behavior of the system might be much more complex than its classical simulation. In particular, since there is no

unique decomposition of a quantum system into its constituent parts, a state of the quantum automaton can be considered in many ways as a state of various virtual classical automata. Cf. the following instructive comment at the end of the article [Po]: ‘The quantum–mechanical computation of one molecule of methane requires 10^{42} grid points. Assuming that at each point we have to perform only 10 elementary operations, and that the computation is performed at the extremely low temperature $T = 3 \cdot 10^{-3} K$, we would still have to use all the energy produced on Earth during the last century.’

The first difficulty we must overcome is the choice of the correct balance between the mathematical and the physical principles. The quantum automaton has to be an abstract one: its mathematical model must appeal only to the general principles of quantum physics, without prescribing a physical implementation. Then the model of evolution is the unitary rotation in a finite dimensional Hilbert space, and the decomposition of the system into its virtual parts corresponds to the tensor product decomposition of the state space. Somewhere in this picture we must accommodate interaction, which is described by density matrices and probabilities.”

BIBLIOGRAPHY

- [BCDP] D. Beckman, A. N. Chari, Sr. Devabhaktuni, J. Preskill - *Efficient networks for quantum computing*. Phys. Rev. A, 54:2 (1996), 1034–1063.
- [Ben1] P. Benioff - *The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines*, J. Stat. Phys., 22 (1980), 563–591.
- [Ben2] P. Benioff - *Quantum mechanical Hamiltonian models of Turing machines that dissipate no energy*, Phys. Rev. Lett., 48 (1980), 1581–1585.
- [BoL] D. Boneh, R. Lipton - *Quantum cryptanalysis of hidden linear functions*, Proc. of Advances in Cryptology — CRYPTO ’95, Springer LN in Computer Science, vol. 963 (1995), 424–437.
- [BoyBHT] M. Boyer, G. Brassard, P. Høyer, A. Tapp - *Tight bounds on quantum searching*, Preprint, 1996.
- [CZ] J. Cirac, P. Zoller - *Quantum computation with cold trapped ions*, Phys. Rev. Lett., 74:20 (1995), 4091–4094.
- [Deu] D. Deutsch - *Quantum theory, the Church–Turing principle and the universal quantum computer*, Proc. R. Soc. Lond. A 400 (1985), 97–117.
- [DeuJ] D. Deutsch, R. Jozsa - *Rapid solutions of problems by quantum computation*, Proc. Roy. Soc. London, Ser. A, 449 (1992), 553–558.
- [Fe1] R. Feynman - *Simulating physics with computers*, Int. J. of Theor. Phys., 21 (1982), 467–488.
- [Fe2] R. Feynman. *Quantum mechanical computers*, Found. Phys., 16 (1986), 507–531.
- [Fr1] M. Freedman - *Topological views on computational complexity*, In Proc. ICM Berlin 1998, vol. II, 453–464.
- [Fr2] M. Freedman - *Limit, logic, and computation*, Proc. Nat. Ac. Sci. USA, 95 (1998), 95–97.

- [Fr3] M. Freedman - *P/NP, and the quantum field computer*, Proc. Nat. Ac. Sci. USA, 95 (1998), 98–101.
- [GaJ] M. Garey, D. Johnson - *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San-Francisco, 1979.
- [GeC] N. Gershenfield, I. Chuang - *Bulk spin-resonance quantum computation*, Science 275 (1997), 350–355.
- [Gri] D. Grigoriev - *Testing the shift-equivalence of polynomials using quantum mechanics*, In: Manin's Festschrift, Journ. of Math. Sci., 82:1 (1996), 3184–3193.
- [Gro] L. K. Grover - *Quantum mechanics helps in searching for a needle in a haystack*, Phys. Rev. Lett. 79 (1997), 325–328.
- [Ki1] A. Kitaev - *Quantum computations: algorithms and error correction*, Russian Math. Surveys, 52:6 (1997), 53–112.
- [Ki2] A. Kitaev - *Classical and quantum computations*, Lecture notes, Independent University, Moscow, 1998.
- [Ma1] Yu. Manin - *A Course in Mathematical Logic*, Springer Verlag, 1977, pp. xiii+286.
- [Ma2] Yu. Manin - *Computable and uncomputable (in Russian)*, Moscow, Sovetskoye Radio, 1980.
- [Mu] D. Mumford - *The statistical description of visual signals*, Preprint.
- [Po] R. P. Poplavskii - *Thermodynamical models of information processing (in Russian)*, Uspekhi Fizicheskikh Nauk, 115:3 (1975), 465–501.
- [Sa] A. Salomaa - *Computation and Automata*, Cambridge UP, 1985.
- [Sh] P. W. Shor - *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput., 26:5 (1997), 1484–1509.
- [Si] D. Simon - *On the power of quantum computation*, Proc. of the 35th Ann. Symp. on Foundations of Comp. Sci. (1994), 116–123.
- [Ts] B. Tsirelson - *Quantum information processing*, Lecture notes, Tel-Aviv University, 1997.

Yu. I. MANIN

Max-Planck-Institut für Mathematik
 Vivatgasse 7
 D-53111 BONN (Germany)
 E-mail : manin@mpim-bonn.mpg.de

DENSITÉ MAXIMALE
DES EMPILEMENTS DE SPHÈRES EN DIMENSION 3
[d'après Thomas C. HALES et Samuel P. FERGUSON]

par Joseph OESTERLÉ (*)

Au nouvel an 1610, Johannes Kepler offre à son protecteur et ami Johannes Matthaüs Wackher von Wackenfels un opuscule d'une vingtaine de pages ([14]), intitulé *Strena seu de nive sexangula* (l'étrenne ou la neige sexangulaire). Il y médite sur l'origine de la forme des flocons de neige en étoile à six branches, ce qui l'amène à s'intéresser entre autres à la forme des alvéoles des abeilles, au dodécaèdre rhomboédrique, à la forme des grains de grenades et à l'empilement de sphères que nous appelons aujourd'hui *cubique à faces centrées*, et qui est par exemple celui suivant lequel les artilleurs empilent les boulets de canon. Kepler affirme qu'il n'est pas possible de trouver un empilement de sphères (toutes de même rayon) plus serré, assertion qui est devenue la fameuse *conjecture de Kepler*.

En 1997, T. Hales publie une stratégie pour démontrer la conjecture de Kepler, en la ramenant à un problème d'optimisation non linéaire en dimension finie. Il mène ce programme à bien, en collaboration avec son étudiant S. Ferguson, et l'achève en juillet 1998. Tant le choix du problème d'optimisation considéré que sa résolution s'appuient sur un usage intensif de l'ordinateur. Mais près de 4 siècles après avoir été formulée, la conjecture de Kepler est enfin démontrée :

THÉORÈME 1 (Hales – Ferguson). — *Tout empilement de sphères dans l'espace euclidien de dimension 3 a une densité au plus égale à la densité $\frac{\pi}{3\sqrt{2}}$ de l'empilement cubique à faces centrées.*

Précisons le vocabulaire utilisé : Le mot *sphère* signifie ici sphère pleine, *i.e.* est synonyme de boule fermée. Un *empilement de sphères* est une famille de boules fermées, toutes de même rayon, d'intérieurs mutuellement disjoints. Soit B la réunion de ces boules ; la *densité de l'empilement* est la limite supérieure de $\text{vol}(B \cap \Omega) / \text{vol}(\Omega)$, pour Ω un cube dont le côté tend vers l'infini.

On peut donner du th. 1 la formulation équivalente suivante, qui ne fait intervenir que des empilements finis :

THÉORÈME 1'. — *Une famille de boules disjointes, toutes de même rayon et contenues dans un parallélépipède Ω de l'espace euclidien de dimension 3, occupe une fraction*

(*) L'auteur remercie J. C. Lagarias de lui avoir communiqué ses notes personnelles sur le sujet.

du volume de Ω au plus égale à $\frac{\pi}{3\sqrt{2}}$.

Le fait que le th. 1' implique le th. 1 résulte de la définition de la densité d'un empilement de sphères ; l'implication réciproque se déduit de ce que l'on peut paver l'espace par des translatés d'un parallélépipède.

L'empilement cubique à faces centrées est ainsi nommé car on l'obtient en plaçant les centres des sphères aux sommets et aux centres des faces des mailles d'un réseau cubique (le rayon des sphères étant choisi de façon à ce qu'elles soient tangentes à leurs plus proches voisines). En prenant pour verticale la direction soit d'un côté, soit d'une diagonale de la maille cubique, on peut voir cet empilement comme une superposition de couches horizontales de sphères réparties soit en réseau carré (chaque sphère étant tangente à 4 autres de la même couche), soit en réseau hexagonal (chaque sphère étant tangente à 6 autres de la même couche) ; chaque couche s'insère dans les trous les plus profonds de la précédente.

Remarque 1. — Lorsqu'on superpose des couches de sphères réparties comme ci-dessus en réseau hexagonal, chaque couche peut être posée de deux façons distinctes dans les trous les plus profonds de la précédente. Cela conduit à une infinité (non dénombrable) d'empilements de sphères différents, tous de même densité $\frac{\pi}{3\sqrt{2}}$ que l'empilement cubique à faces centrées.

1. Quelques mots sur l'histoire de la conjecture de Kepler

En 1831, dans son étude des formes quadratiques ternaires, C. F. Gauss ([6]) démontre le cas particulier suivant de la conjecture de Kepler : si les centres des sphères d'un empilement forment un réseau, la densité de l'empilement est $\leq \frac{\pi}{3\sqrt{2}}$, avec égalité si et seulement si l'empilement est cubique à faces centrées. Ce résultat sera largement généralisé en dimension supérieure (cf. [1], [15]).

La conjecture de Kepler est reprise et généralisée par Hilbert dans son 18^{ème} problème.

En 1910, A. Thue ([18]) démontre l'analogie en dimension 2 de la conjecture de Kepler : *tout empilement de cercles dans le plan euclidien a une densité au plus égale à la densité $\frac{\pi}{2\sqrt{3}}$ de l'empilement hexagonal.*

En 1953, L. Fejes Tóth ([3]) propose un programme pour démontrer la conjecture de Kepler en la ramenant à un problème d'optimisation non linéaire en dimension finie. Ce problème, toujours ouvert, semble trop complexe pour être résolu avec les moyens de calcul actuels. Mais Fejes Tóth est le premier à deviner l'importance que le développement rapide des ordinateurs aura pour ce type de stratégie.

En 1958, C. A. Rogers ([16]) démontre que la densité supérieure d'un empilement de sphères dans un espace euclidien de dimension n est au plus égale à σ_n , portion du volume d'un simplexe régulier de côté 2 recouverte par les sphères unité centrées en ses sommets. Pour $n = 2$, cela fournit une nouvelle démonstration du théorème de Thue, car on peut paver le plan euclidien par des triangles équilatéraux.

L'espace euclidien de dimension 3 ne peut être pavé par des tétraèdres réguliers. La borne de Rogers, $\sigma_3 = 0,7796\dots$, est de ce fait moins bonne que celle prédite par la conjecture de Kepler, $\frac{\pi}{3\sqrt{2}} = 0,7404\dots$. La phrase célèbre de Rogers : *Many*

mathematicians believe, and all physicists know, that the density cannot exceed $\frac{\pi}{3\sqrt{2}}$, exprime mieux que toute autre la redoutable subtilité de la conjecture de Kepler, malgré son apparente simplicité.

La borne de Rogers pour la densité des empilements de sphères en dimension 3 sera améliorée par divers auteurs ; avant les travaux de Hales et Ferguson, la meilleure borne connue était 0,7731 (obtenue par D. J. Muder en 1993). Hales mentionne que les travaux de Muder ont influencé les siens.

Dans un article publié en 1993, Wu-yi Hsiang ([13]) affirme démontrer la conjecture de Kepler. Les spécialistes s'accordent à considérer que les preuves de nombreuses assertions n'y sont pas présentées avec suffisamment de détails pour être qualifiées de complètes. De ce fait, l'article de Hsiang est au mieux vu comme un plan d'attaque de la conjecture de Kepler. Je partage cette opinion. Les textes ultérieurs de Hsiang qui m'ont été communiqués, en particulier le manuscrit d'un livre en préparation, ne me semblent pas pour l'instant de nature à modifier ce point de vue.

2. La conjecture du dodécaèdre et le problème des 13 sphères

Considérons un empilement de sphères de rayon 1 dans un espace euclidien E . Soit S l'ensemble des centres des sphères. La *cellule de Voronoi* de s , notée $V(s)$ se compose des points de E qui sont plus proches de s que des autres points de S . C'est un ensemble convexe fermé, localement polyédrique. Les cellules de Voronoi des points de S pavent l'espace E : autrement dit, elles recouvrent E et leurs intérieurs sont mutuellement disjoints. Chacune de ces cellules contient une unique sphère de l'empilement ; on en déduit facilement que si leur volume est minoré par v , la densité de l'empilement est majorée par b/v , où b est le volume de la boule unité.

Cette observation conduit naturellement au problème suivant : *quel est le volume minimal des cellules de Voronoi des empilements de sphères de rayon 1 ?*

En dimension 2, le problème a été résolu par Fejes Tóth en 1943 ([2]) : l'aire d'une cellule de Voronoi d'un empilement de cercles de rayon 1 est $\geq 2\sqrt{3}$, avec égalité si et seulement si la cellule est un hexagone régulier circonscrit à un cercle de rayon 1. Cet énoncé implique le théorème de Thue relatif à la densité des empilements de cercles.

En dimension 3, le problème a été résolu par Hales et Mc Laughlin en novembre 1998, par des techniques analogues à celles utilisées pour prouver la conjecture de Kepler. Ils démontrent ainsi la *conjecture du dodécaèdre*, formulée par Fejes Tóth en 1943 (et démontrée par lui dans le cas particulier où les cellules de Voronoi n'ont que 12 faces au plus) :

THÉORÈME 2 (Hales – Mc Laughlin). — *Le volume d'une cellule de Voronoi d'un empilement de sphères de rayon 1 dans l'espace euclidien de dimension 3 est $\geq 10\sqrt{130} - 58\sqrt{5}$, avec égalité si et seulement si la cellule est un dodécaèdre régulier circonscrit à une sphère de rayon 1.*

Notons que ce théorème implique que la densité d'un empilement de sphères en dimension 3 est majorée par 0,7547, une borne nettement meilleure que celles obtenues avant que ne soit démontrée la conjecture de Kepler.

Remarque 2. — Pour démontrer le théorème 2, on peut supposer que l’empilement considéré est saturé (*i.e.* maximal). Dans ce cas, pour tout $s \in S$, la cellule de Voronoi $V(s)$ est contenue dans la boule de centre s et de rayon 2. Sa forme, et donc son volume, ne dépendent que des points de S dont la distance à s est ≤ 4 , et le nombre de ces points est majoré par une constante absolue. La conjecture du dodécaèdre est donc un problème d’optimisation non linéaire : il s’agit de trouver le maximum de la fonction volume sur un espace de configurations qui est une partie compacte d’un espace vectoriel de dimension finie. Une des difficultés vient du fait que cette dimension est de l’ordre de 150 : en effet, on sait seulement que le nombre de faces de $V(s)$ est ≤ 49 et on connaît des exemples où il est 44.

En 1694, une controverse oppose Isaac Newton et l’abbé Gregory. Newton pense qu’on ne peut placer qu’au plus 12 sphères identiques (ne s’interpénétrant pas) au contact d’une sphère centrale de même rayon, Gregory pense qu’il est possible d’en placer 13. Cette alternative, le *problème des 13 sphères*, ne sera tranchée qu’en 1953, par Schütte et Van der Waerden ([17]) : ils donnent raison à Newton. Ils exhibent par ailleurs un arrangement de 13 sphères de rayon 1 tangentes à une sphère centrale de rayon $r = 1,0911\dots$. On pense que ce rayon est le plus petit possible, mais cela n’a pour l’instant été démontré. Peut-être les techniques de Hales et Ferguson permettront-elles de le faire.

3. Principe d’attaque de la conjecture de Kepler

Pour majorer la densité des empilements de sphères dans un espace euclidien E de dimension 3, il suffit de considérer les empilements de sphères de rayon 1 qui sont saturés (*i.e.* maximaux). Choisissons donc un tel empilement. Notons B la réunion des boules qui le composent, et S l’ensemble de leurs centres : les distances mutuelles des points de S sont ≥ 2 , et tout point de E est à une distance < 2 de S .

Si A est une partie de E , notons χ_A sa fonction caractéristique. Supposons que, pour des nombres réels a, b, c tels que $a > \frac{3c}{4\pi}$, on puisse écrire

$$(1) \quad a\chi_B - b\chi_E = \sum_{s \in S} \phi_s \quad (\text{presque partout}),$$

où ϕ_s est une fonction intégrable dont la norme L^1 est majorée par une constante indépendante de s , dont le support est contenu dans une boule de centre s et de rayon indépendant de s , et dont l’intégrale satisfait la relation

$$(2) \quad \int \phi_s \leq c.$$

Démontrons que la densité δ de l’empilement satisfait alors l’inégalité

$$(3) \quad \left(a - \frac{3c}{4\pi}\right)\delta \leq b$$

En intégrant (1) sur un cube Ω de côté r , on obtient lorsque r tend vers l’infini

$$a \operatorname{vol}(B \cap \Omega) - b \operatorname{vol}(\Omega) \leq c \operatorname{Card}(S \cap \Omega) + O\left(\frac{1}{r}\right) \leq \frac{3c}{4\pi} \operatorname{vol}(B \cap \Omega) + O(r^2).$$

d'où $(a - \frac{3c}{4\pi}) \frac{\text{vol}(B \cap \Omega)}{\text{vol}(\Omega)} \leq b + O(r^{-1})$; l'inégalité (3) s'en déduit par passage à la limite supérieure.

Exemples. — 1) La relation (1) est satisfaite lorsque l'on prend $a = 0$, $b = -1$ et $\phi_s = -\chi_{V(s)}$, où $V(s)$ est la cellule de Voronoi de s ; cette cellule est contenue dans la boule de centre s et de rayon 2, puisque l'empilement est supposé saturé. La relation (2) est satisfaite en prenant $c = -v$, où v est un minorant du volume des cellules de Voronoi. La relation (3) donne l'inégalité $\delta \leq \frac{4\pi}{3v}$ que nous avons rencontrée au n° 2.

2) Soit r un nombre réel > 2 . Disons que deux points distincts de S sont voisins si leur distance est $\leq r$. Fejes Tóth propose pour démontrer la conjecture de Kepler de prendre le plus grand r qui garantisse que tout point de S a au plus 12 voisins (probablement 2,0534...), et de poser $\phi_s = -\frac{12-m}{12}\chi_{V(s)} - \frac{1}{12} \sum \chi_{V(t)}$, où t parcourt les voisins de s et m est leur nombre. La relation (1) est dans ce cas satisfaite avec $a = 0$ et $b = -1$, mais la détermination de la constante c optimale dans (2) apparaît totalement hors d'atteinte.

3) On appelle *décomposition de Delaunay* associée à l'empilement une décomposition de l'espace E en tétraèdres, dont les sommets appartiennent à S , et dont les boules circonscrites n'ont pas de point intérieur dans S . Il existe au moins une telle décomposition ; celle-ci est unique lorsqu'aucun point de E n'appartient à plus de 4 cellules de Voronoi (ce qui est le cas générique), et est dans ce cas duale de celle de Voronoi. Fixons une décomposition de Delaunay associée à l'empilement et appelons *étoile* de s la réunion $D(s)$ des simplexes de la décomposition contenant s . Posons $\phi_s = \chi_{B \cap D(s)} - \delta_{oct}\chi_{D(s)}$, où δ_{oct} est la portion du volume d'un octaèdre régulier de côté 2 recouverte par les boules unités centrées en ses sommets. La première tentative de Hales pour démontrer en 1992 la conjecture de Kepler revenait à utiliser ces fonctions ϕ_s , avec $a = 4$ et $b = 4\delta_{oct}$. Mais il s'aperçut que pour certaines étoiles (les prismes pentagonaux), l'intégrale de ϕ_s était trop grande pour conclure.

4. Articulation de la démonstration de Hales et Ferguson

La démonstration de Hales et Ferguson de la conjecture de Kepler se compose de 8 articles, totalisant environ 300 pages, dont l'ordre logique est le suivant :

An overview of the Kepler conjecture, de T.C. Hales ([7]).

A formulation of the Kepler conjecture, de S. Ferguson et T.C. Hales ([5]).

Sphere packings I, de T.C. Hales ([8]).

Sphere packings II, de T.C. Hales ([9]).

Sphere packings III, de T.C. Hales ([10]).

Sphere packings IV, de T.C. Hales ([11]).

Sphere packings V, de S. Ferguson ([4]).

The Kepler conjecture (Sphere packings VI), de T.C. Hales ([12]).

Les premiers articles rédigés furent *Sphere packings I* et *Sphere packings II*. Ils ont été publiés en 1997. Les autres articles sont pour l'instant des prépublications électroniques, disponibles sur la page personnelle de T. Hales.

Dans *Sphere packings I*, Hales décrit sa stratégie pour démontrer la conjecture de Kepler. Le problème d'optimisation qu'il considère est de même nature que celui

considéré dans l'exemple 3 du n° 3, mais s'appuie sur une décomposition de l'espace hybride entre la décomposition de Voronoi et celle de Delaunay. À l'étoile $D(s)$ d'un point s pour cette décomposition, Hales associe un nombre réel qu'il appelle *le score* de $D(s)$ et qui est l'intégrale de la fonction ϕ dans l'approche du n° 3.

L'ensemble \mathcal{D} des étoiles possibles, pour $s = 0$, a une structure naturelle d'espace compact et le score est une fonction continue sur \mathcal{D} . Les étoiles qui proviennent de l'empilement cubique faces centrée et de ses variantes décrites dans la remarque 1 forment deux orbites \mathcal{D}_0 et \mathcal{D}'_0 sous l'action du groupe orthogonal. Pour démontrer la conjecture de Kepler, il suffit de démontrer que le maximum du score est atteint sur $\mathcal{D}_0 \cup \mathcal{D}'_0$. Pour des raisons de commodité, Hales note $8pt$ la valeur prise par le score sur cet ensemble.

Chaque étoile possède une structure combinatoire, décrite par une carte sphérique. Dans *Sphere packings II*, Hales démontre une propriété forte de maximum local : les étoiles ayant même structure combinatoire que celles de $\mathcal{D} \cup \mathcal{D}'_0$ ont un score $\leq 8pt$, avec égalité si et seulement si elles appartiennent à $\mathcal{D} \cup \mathcal{D}'_0$.

La stratégie consiste alors à démontrer que pour les autres structures combinatoires, le score est $< 8pt$. Les difficultés techniques rencontrées, en particulier dans l'étude des prismes pentagonaux, conduisent Ferguson et Hales à modifier le problème d'optimisation initial, en proposant une nouvelle décomposition de l'espace et une nouvelle fonction score, plus compliquées apparemment mais mieux adaptées à l'étude numérique envisagée. C'est ce qu'ils font dans *A formulation of the Kepler conjecture*, en prenant soin de vérifier que les résultats de *Sphere packings I* et *Sphere packings II* restent valables dans ce nouveau cadre.

Dans *Sphere packings III*, Hales démontre que le score (nouvelle version) de D est $< 8pt$ si la carte associée n'a que des faces triangulaires ou quadrilatérales, et ne provient pas d'un prisme pentagonal (ni de \mathcal{D}_0 et \mathcal{D}'_0) ; le cas des prismes pentagonaux est traité par Ferguson dans *Sphere packings V*, celui des cartes possédant des faces à 5 côtés ou plus par Hales dans *Sphere packings IV* et *Sphere packings VI*.

Au total, ce sont environ 5000 cartes, recensées par ordinateur, qui interviennent dans le problème. Des méthodes générales de programmation linéaire permettent de se ramener à n'examiner qu'une centaine de cas ; ceux-ci sont traités un à un. Beaucoup des problèmes d'optimisation non linéaires rencontrés sont remplacés par des problèmes d'optimisation linéaires qui les dominent, que l'on sait résoudre par les méthodes de programmation linéaire. Au total, ce sont près de 100000 problèmes linéaires, en 100 à 200 variables, avec 1000 à 2000 contraintes, qu'il a fallu traiter pour achever la démonstration. Les calculs ont été effectués en utilisant une *arithmétique des intervalles* : chaque nombre réel r est représenté par un couple (a, b) de nombres rationnels qui l'encadrent, de manière à garantir que les erreurs d'arrondis cumulées n'entachent pas la validité des résultats obtenus.

5. La conjecture de Kepler est-elle démontrée ?

La situation n'est pas sans rappeler celle du théorème des 4 couleurs, dont la démonstration par K. Appel et W. Haken en 1977 avait nécessité l'examen de quelques milliers de configurations par ordinateur.

Je ne prétends pas avoir vérifié l'intégralité de la démonstration de Hales et Ferguson, et encore moins les programmes informatiques utilisés. Je ne pense pas que quiconque à part les auteurs l'ait fait.

Ma conviction personnelle que les travaux de Hales et Ferguson fournissent effectivement une preuve de la conjecture de Kepler repose sur les arguments suivants :

— Ni erreurs substantielles, ni points obscurs n'ont été relevés jusqu'à présent par les spécialistes ayant examiné ces travaux.

— Dans toutes les parties du travail que j'ai pris la peine de lire en détail, j'ai trouvé des démonstrations claires et complètes.

— Le recours à l'intuition géométrique, dangereux dans ce type de problème, est réduit au minimum et remplacé par des inégalités précises.

— Les algorithmes employés sont décrits avec soin, et les codes rédigés sont disponibles pour vérification. L'arithmétique des intervalles tient compte des erreurs commises dans l'approximation des nombres réels.

— Si d'aventure, on s'apercevait ultérieurement qu'un cas à examiner a échappé à la sagacité des auteurs, il est vraisemblable que leurs techniques permettraient de le traiter sans devoir introduire d'idées nouvelles.

BIBLIOGRAPHIE

- [1] J. H. CONWAY, N. J. A. SLOANE – *Sphere packings, lattices and groups*, Grundlehren der mathematischen Wissenschaften 290, Springer Verlag, 1988.
- [2] L. FEJES TÓTH – *Über die dichteste Kugellagerung*, Math. Z. 48 (1943), 676-684.
- [3] L. FEJES TÓTH – *Lagerungen in der Ebene, auf der Kugel und im Raum*, Springer, Berlin, 1953.
- [4] S. P. FERGUSON – *Sphere packings V*, thèse, Université du Michigan, 1997.
- [5] S. P. FERGUSON, T. C. HALES – *A formulation of the Kepler conjecture*, prépublication.
- [6] C. F. GAUSS – *Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen von Ludwig August Seber*, Göttingische gelehrte Anzeigen, 1831, Juli 9 (= Werke, vol 2, Königliche Gesellschaft der Wissenschaften, Göttingen, 1876, 188-196).
- [7] T. C. HALES – *An overview of the Kepler conjecture*, prépublication.
- [8] T. C. HALES – *Sphere packings I*, Disc. Comp. Geom. 17 (1997), 1-51.
- [9] T. C. HALES – *Sphere packings II*, Disc. Comp. Geom. 18 (1997), 135-149.
- [10] T. C. HALES – *Sphere packings III*, prépublication.
- [11] T. C. HALES – *Sphere packings IV*, prépublication.
- [12] T. C. HALES – *The Kepler conjecture (Sphere packings VI)*, prépublication.
- [13] W-Y. HSIANG – *On the sphere packing problem and the proof of Kepler's conjecture*, Intern. J. Math. 93 (1993), 739-831.
- [14] J. KEPLER – *Strena seu de nive sexangula*, Frankfurt, Tampach, 1611, 24 pages. (Traduction allemande de F. Rossmann, *Neujahrsgabe*, Berlin, 1943 ; traduction anglaise de C. Hardie, *The six-cornered snowflake*, Oxford, 1966 ; traduction

française de R. Halleux, *L'étrenne ou la neige sexangulaire*, librairie J. Vrin et éditions du CNRS, Paris, 1975.)

- [15] J. OESTERLÉ – *Empilements de sphères*, Séminaire Bourbaki 1989-90, exposé n° 727, Astérisque 189-190 (1990), 375-397.
- [16] C. A. ROGERS – *The packing of equal spheres*, Proc. London Math. Soc. 8 (1958), 609-620.
- [17] K. SCHÜTTE, B. L. VAN DER WAERDEN – *Das Problem der dreizehn Kugeln*, Proc. London Math. Ann. 125 (1953), 325-334.
- [18] A. THUE – *Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene*, Norske Vid. Selsk. Skr.1 (1910), 1-9.

Joseph OESTERLÉ

Institut Henri Poincaré

11, rue Pierre et Marie Curie

F-75005 PARIS

Adresse électronique : oesterle@ihp.jussieu.fr

SOUS-GROUPES FINIS DES GROUPES DE LIE

par Jean-Pierre SERRE

INTRODUCTION

Les sous-groupes finis du groupe des rotations $SO_3(\mathbf{R})$ sont bien connus. Ce sont :

- les groupes cycliques C_n d'ordre $n = 1, 2, \dots$;
- les groupes diédraux D_n d'ordre $2n$, $n = 2, 3, \dots$;
- le groupe alterné Alt_4 d'ordre 12 ;
- le groupe symétrique Sym_4 d'ordre 24 ;
- le groupe alterné Alt_5 d'ordre 60.

On aimerait avoir une liste analogue pour d'autres groupes de Lie compacts, ou d'autres groupes algébriques (en caractéristique zéro, et même en caractéristique > 0). Ce serait utile pour beaucoup de questions (représentations ℓ -adiques, par exemple). Bien sûr, c'est trop demander, vu que tout groupe fini se plonge dans un groupe unitaire convenable ! On va voir que l'on peut tout de même dire pas mal de choses si l'on se borne à des groupes finis qui sont, soit abéliens, soit simples.

HYPOTHÈSES ET NOTATIONS

Plutôt que de travailler dans la catégorie des groupes de Lie compacts, on préfère se placer dans celle des groupes réductifs complexes. Cela ne change rien : on sait que, si K est un groupe de Lie compact, il possède un complexifié G qui est un groupe réductif sur \mathbf{C} ; le groupe K est un sous-groupe compact maximal de $G(\mathbf{C})$. Tout sous-groupe fini de $G(\mathbf{C})$ est conjugué à un sous-groupe de K ; de plus, K "contrôle la fusion de K dans $G(\mathbf{C})$ " au sens suivant : si A, B sont deux sous-groupes de K , et si $g \in G(\mathbf{C})$ est tel que $gAg^{-1} = B$, il existe un élément g_0 de K tel que $g_0ag_0^{-1} = gag^{-1}$ pour tout $a \in A$ (cela se déduit de la décomposition de Cartan de $G(\mathbf{C})$).

(Dans le cas particulier $K = SO_3(\mathbf{R})$, on a $G = PGL_2$, de sorte que les groupes C_n, D_n, \dots, Alt_5 s'interprètent comme des sous-groupes finis de $PGL_2(\mathbf{C})$, c'est-à-dire comme des groupes finis d'automorphismes de la droite projective.)

Dans ce qui suit, on adoptera le point de vue des groupes algébriques (qui a, entre autres avantages, celui de permettre des réductions modulo p). On fixe un corps k algébriquement clos de caractéristique zéro, ainsi qu'un groupe réductif connexe G défini sur k ; on se permet d'identifier G à $G(k)$. Le cas le plus intéressant est celui où G est "presque simple", i.e. semi-simple à système de racines irréductible ; le groupe adjoint G^{ad} est alors un groupe simple, au sens usuel du terme.

1. LE CAS (PRESQUE) AB ÉLIEN

Lorsque $G = \mathrm{PGL}_2$ les sous-groupes abéliens finis de G sont les groupes cycliques C_n et le groupe diédral D_2 qui est abélien élémentaire de type $(2, 2)$. Les C_n sont contenus dans un tore maximal, alors que D_2 ne l'est pas ; le nombre premier $p = 2$ joue donc un rôle particulier pour PGL_2 . Nous allons trouver une situation analogue dans le cas général.

1.1. Sous-groupes toraux

Un sous-groupe fini A de G est dit *toral* s'il est contenu dans un tore maximal T de G . La structure d'un tel sous-groupe est évidente : si $r = \dim T$ est le rang de G , A peut être engendré par r éléments ; inversement, tout groupe abélien ayant cette propriété est isomorphe à un sous-groupe toral de G .

Soit $N = N_G(T)$ le normalisateur de T dans G . Le quotient $W = N/T$ est le *groupe de Weyl* de G (plus correctement : du couple (G, T)). Ce groupe opère sur T par conjugaison, et il contrôle la fusion de T dans G :

1.1.1. *Si A et B sont des sous-groupes de T , et si $g \in G$ est tel que $gAg^{-1} = B$, il existe $w \in W$ tel que $w(a) = gag^{-1}$ pour tout $a \in A$.*

Cet énoncé est l'exact analogue d'un théorème de Burnside sur les sous-groupes du centre d'un p -groupe de Sylow. Il se démontre de la même manière : on remarque que T et $g^{-1}Tg$ sont des tores maximaux du centralisateur $Z_G(A)$ de A , donc sont conjugués par $Z_G(A)$. Cela permet de remplacer g par un élément de N ; d'où le résultat cherché.

Les groupes abéliens ayant très peu de générateurs sont toraux :

1.1.2. *Soit A un sous-groupe abélien fini de G . Alors A est toral dans chacun des deux cas suivants :*

- a) *A est cyclique ;*
- b) *G est simplement connexe, et A est engendré par deux éléments.*

Le cas a) est immédiat : tout élément d'ordre fini est semi-simple, donc contenu dans un tore maximal. Dans le cas b), supposons A engendré par x, y . Du fait que G est simplement connexe, le centralisateur $Z_G(x)$ est connexe. Le même argument que dans a) montre qu'il existe un tore maximal T de $Z_G(x)$ qui contient y . Ce tore est un tore maximal de G et il contient x , donc A .

1.2. Plongements dans N

À défaut de pouvoir plonger un groupe abélien fini dans un tore maximal, on peut essayer de le plonger dans le normalisateur d'un tel tore. C'est toujours possible. Plus généralement (cf. Borel-Serre [6], Borel-Mostow [5] et Springer-Steinberg [33], II.5.6) :

1.2.1. *Soit A un sous-groupe fini hyper-résoluble de G . Il existe un tore maximal T de G dont le normalisateur N contient A .*

Rappelons qu'un groupe A est dit hyper-résoluble ("supersolvable") s'il admet une suite de composition :

$$1 = A_0 \subset A_1 \subset \dots \subset A_n = A,$$

où les A_i sont normaux dans A , et A_i/A_{i-1} est cyclique pour tout $i \geq 1$. On a les implications :

$$\text{abélien} \implies \text{nilpotent} \implies \text{hyper-résoluble} \implies \text{résoluble}.$$

Voici une application simple de 1.2.1 :

1.2.2. *Soit p un nombre premier ne divisant pas l'ordre du groupe de Weyl W . Si A est un p -groupe contenu dans G , A est abélien et toral.*

En effet, on peut supposer, d'après 1.2.1, que A est contenu dans N . Vu l'hypothèse faite sur p , son image dans $W = N/T$ est triviale. Il est donc contenu dans T .

Remarque : Le groupe N est une extension, en général non triviale, de W par T . On trouvera dans Tits ([35],[36]) une description de cette extension, en termes d'un certain groupe fini N_Z défini explicitement par générateurs et relations ; voir aussi Bourbaki, LIE IX, p. 115, exerc. 12.

1.3. Nombres premiers de torsion

(Références : Borel [4], Steinberg [34] et Bourbaki, LIE IX, p. 120-121, exerc. 7 à 12.)

Un nombre premier p est dit *de torsion* (pour G) s'il vérifie les conditions équivalentes suivantes :

- a) *Il existe un p -sous-groupe abélien de G qui n'est pas toral.*
- a') *Il existe un p -sous-groupe abélien élémentaire de G , de rang ≤ 3 , qui n'est pas toral.*

On note $\text{Tors}(G)$ l'ensemble de ces nombres premiers ; d'après 1.2.2, c'est un sous-ensemble de l'ensemble des diviseurs premiers de l'ordre de W . Dans le cas particulier où $G = \text{PGL}_2$, on a $\text{Tors}(G) = \{2\}$.

Le terme de "torsion" provient du résultat suivant, dans lequel je suppose que $k = \mathbf{C}$ (sinon il faut faire intervenir la cohomologie étale) :

1.3.1. (cf. [4],[34]) *Pour que p appartienne à $\text{Tors}(G)$, il faut et il suffit que l'un des groupes d'homologie $H_i(G, \mathbf{Z})$ contienne un élément d'ordre p .*

(Noter qu'il revient au même de considérer l'homologie de $G = G(\mathbf{C})$ ou celle d'un compact maximal K , car $G(\mathbf{C})$ et K ont même type d'homotopie.)

On trouvera dans [4] et [34] une longue liste de propriétés caractérisant les éléments de $\text{Tors}(G)$. En voici quelques-unes :

1.3.2. *On a $\text{Tors}(G) = \text{Tors}(G')$, où G' est le groupe dérivé de G .*

Comme G' est semi-simple, cela ramène l'étude de $\text{Tors}(G)$ au cas où G est semi-simple. Dans ce cas, notons \overline{G} le revêtement universel de G , notons $\pi_1(G)$ le noyau de $\overline{G} \rightarrow G$ et soit $\text{Tors}(\pi_1(G))$ l'ensemble des nombres premiers qui divisent l'ordre du groupe fini $\pi_1(G)$. Alors :

1.3.3. *On a $\text{Tors}(G) = \cup_H \text{Tors}(\pi_1(H'))$, où H parcourt les sous-groupes réductifs connexes de G ayant même rang que G .*

1.3.4. *On a $\text{Tors}(G) = \text{Tors}(\overline{G}) \cup \text{Tors}(\pi_1(G))$.*

Cet énoncé ramène la détermination de $\text{Tors}(G)$ au cas où G est simplement connexe. En utilisant 1.3.3, on en déduit (cf. [4],[34]) :

1.3.5. *Supposons G simplement connexe et presque simple. Soit (α_i) une base de son système de racines, soit β la plus grande racine, et écrivons la racine duale β^\vee de β sous la forme :*

$$\beta^\vee = \sum n_i \alpha_i^\vee,$$

où les n_i sont des entiers > 0 . Alors, pour que p soit de torsion pour G , il faut et il suffit qu'on ait $p \leq \sup(n_i)$.

D'où :

1.3.6. *Supposons G simplement connexe et presque simple. Alors :*

$\text{Tors}(G) = \emptyset$ si G est de type A_n ou C_n ;

$\text{Tors}(G) = \{2\}$ si G est de type B_n ($n \geq 3$), D_n ($n \geq 4$) ou G_2 ;

$\text{Tors}(G) = \{2, 3\}$ si G est de type F_4 , E_6 ou E_7 ;

$\text{Tors}(G) = \{2, 3, 5\}$ si G est de type E_8 .

1.4. Exemples de groupes abéliens élémentaires non toraux

(Références : Adams [1], Borel [4], Borel–Serre [6], Cohen-Seitz [10], Steinberg [34] et (surtout) Griess [17].)

Je me borne à deux exemples, l’un relatif à $p = 2$ et l’autre à $p = 5$.

1.4.1. Supposons que -1 appartienne au groupe de Weyl W ; c’est le cas pour les groupes de type A_1, B_n, C_n, D_n (n pair), G_2, F_4, E_7, E_8 . Soit $g \in N$ un représentant de l’élément -1 de W . On peut montrer que g^2 est d’ordre 1 ou 2, et appartient au centre de G . Supposons que $g^2 = 1$ (c’est le cas si G est de type adjoint). Soit A le groupe engendré par g et par les éléments d’ordre 2 de T ; c’est un groupe abélien élémentaire d’ordre 2^{n+1} , où n est le rang de G , i.e. la dimension de T . *Ce groupe n’est pas toral* ; on peut même montrer que son centralisateur $Z_G(A)$ est fini.

Lorsque G est PGL_2 , le groupe A est le groupe diédral D_2 . Lorsque G est de type G_2, F_4 ou E_8 , A est d’ordre $2^3, 2^5, 2^9$; de tels sous-groupes jouent un grand rôle dans la cohomologie (usuelle – ou galoisienne) du groupe G . Noter que, dans ces trois cas, A est un sous-groupe élémentaire *maximal* de G : de façon générale, si G est simplement connexe, les p -sous-groupes abéliens de G sont de rang $\leq n + 1$ si $p = 2$, et de rang $\leq n$ si $p > 2$, cf. Borel [4] et Cohen-Seitz [10].

1.4.2. Le groupe $G = E_8$ contient un élément z d’ordre 5 dont le centralisateur $Z_G(z)$ est de la forme $G_1 \cdot G_2$, où G_1 et G_2 sont isomorphes à SL_5 , commutent, et ont pour intersection $\langle z \rangle$ (cela se déduit du diagramme de Dynkin complété de E_8 en remarquant que, si l’on en retranche la racine simple qui a le coefficient 5 dans la plus grande racine, on trouve deux diagrammes de type A_4). Dans $G_1 = \mathrm{SL}_5$, il est facile de trouver des éléments x_1, y_1 d’ordre 5 tels que $x_1 y_1 x_1^{-1} y_1^{-1} = z$; de même, il existe dans G_2 des éléments x_2, y_2 d’ordre 5 tels que $x_2 y_2 x_2^{-1} y_2^{-1} = z^{-1}$. Si l’on pose $x = x_1 x_2$ et $y = y_1 y_2$, on constate que le groupe $A = \langle x, y, z \rangle$ est abélien élémentaire d’ordre 5^3 . *Ce groupe n’est pas toral* ; on peut même montrer que $Z_G(A)$ est égal à A .

1.5. Relations entre cohomologie galoisienne et sous-groupes non toraux

Qu’il existe de telles relations est connu depuis longtemps. Voici deux exemples :

1.5.1 (Grothendieck [23]). *Les deux propriétés suivantes sont équivalentes :*

a) $\mathrm{Tors}(G) = \emptyset$ (cela équivaut à dire que tout sous-groupe abélien de G est toral).

b) $H^1(K, G) = 0$ pour toute extension K de k .

(Pour la définition de $H^1(K, G)$, voir par exemple [30].)

Lorsque G est semi-simple, ces propriétés sont satisfaites si et seulement si G est un produit de groupes simplement connexes de type A ou C , cf. § 1.3. Un tel groupe est parfois dit “spécial”.

1.5.2. *Supposons que G soit égal à PGL_2 , ou soit de type G_2 . Soit A le 2-sous-groupe élémentaire non toral de G défini dans 1.4.1. Alors, pour toute extension K de k , l’application $H^1(K, A) \rightarrow H^1(K, G)$ est surjective.*

Noter que $H^1(K, A)$ n’est autre que $\mathrm{Hom}(\mathrm{Gal}(\overline{K}/K), A)$. Dans le cas de PGL_2 , les éléments de $H^1(K, A)$ peuvent donc s’interpréter comme des couples (λ, μ) d’éléments

de K^*/K^{*2} et l'élément correspondant de $H^1(K, \mathrm{PGL}_2)$ est l'algèbre de quaternions définie par deux générateurs i et j soumis aux relations

$$i^2 = \lambda, \quad j^2 = \mu, \quad ij = -ji.$$

Même chose pour G_2 , les quaternions étant remplacés par les octonions.

L'énoncé 1.5.2, pour agréable qu'il soit, ne donne pas de moyen de prouver la non trivialité des éléments de $H^1(K, G)$ ainsi obtenus ; il faut le compléter par la construction d'*invariants cohomologiques*, cf. [30], §§ 6,7 et n et Youssin [29] ont obtenu un résultat bien plus satisfaisant. Pour le formuler, il faut d'abord définir la *dimension essentielle* $\mathrm{ed}(x)$ d'un élément x de $H^1(K, G)$: c'est la borne inférieure des degrés de transcendance sur k des sous-extensions K' de K telles que x appartienne à l'image de $H^1(K', G) \rightarrow H^1(K, G)$. (En termes plus géométriques – et plus vagues – c'est le nombre minimum de paramètres dont on a besoin pour écrire le G -torseur x .) La borne supérieure des $\mathrm{ed}(x)$, quand K et x varient, est la *dimension essentielle* de G ; elle est notée $\mathrm{ed}(G)$. Nous pouvons maintenant énoncer le théorème principal de [29] :

1.5.3. *Si G contient un p -sous-groupe abélien élémentaire A dont le centralisateur est fini, on a $\mathrm{ed}(G) \geq \mathrm{rang}(A)$.*

En combinant cet énoncé avec 1.4.1, on obtient :

1.5.4. *On a $\mathrm{ed}(E_7^{\mathrm{ad}}) \geq 8$ et $\mathrm{ed}(E_8) \geq 9$.*

Ainsi, il existe des E_8 -torseurs dont la construction exige au moins 9 paramètres !

Remarques

1) L'hypothèse faite sur A dans 1.5.3 est équivalente à dire que A n'appartient à aucun sous-groupe parabolique propre de G . Elle entraîne que A n'est pas toral.

2) L'énoncé démontré dans [29] est plus précis que 1.5.3 ; c'est :

$$\mathrm{ed}(G; p) \geq \mathrm{rang}(A),$$

où $\mathrm{ed}(G; p)$ est la dimension essentielle de G "en p " (i.e. en considérant comme négligeables les extensions de corps de degré premier à p).

3) Les démonstrations de [29] utilisent la *résolution des singularités* (sous forme équivariante). Elles ne s'étendent pas, pour l'instant, aux corps de caractéristique $\neq 0$.

2. LE CAS (PRESQUE) SIMPLE

On va maintenant s'intéresser aux plongements d'un groupe simple (fini, non abélien) dans G .

Il est commode de considérer, plus généralement, les plongements des groupes \overline{S} qui sont des extensions centrales de S (on peut imposer à \overline{S} d'être égal à son groupe dérivé, cela ne change rien). Exemple typique :

$$S = L_2(q) = \mathrm{PSL}_2(\mathbf{F}_q) \quad \text{et} \quad \overline{S} = 2 \cdot L_2(q) = \mathrm{SL}_2(\mathbf{F}_q), \quad q \text{ impair.}$$

(Les notations $L_2(q)$ et $2 \cdot L_2(q)$ sont celles de l'ATLAS [15].)

Un plongement d'un tel groupe \bar{S} dans G est appelé un *plongement projectif* de S . Le principal avantage de cette notion est la propriété d'invariance suivante : si $G' \rightarrow G$ est une isogénie, S a un plongement projectif dans G si et seulement si il a un plongement projectif dans G' .

Lorsque S et G sont donnés, et que G est un groupe classique, l'examen de la table des caractères de S (et de ses extensions centrales) permet de décider si S a un plongement (ou un plongement projectif) dans G ; c'est clair lorsque G est de type A_n , et c'est facile pour les types B_n, C_n, D_n . Une méthode analogue s'applique à G_2 (en utilisant sa représentation irréductible de degré 7 et la forme trilinéaire alternée correspondante), cf. Aschbacher [2] et Cohen–Wales [11]. Les types F_4, E_6, E_7, E_8 sont plus difficiles ; ce n'est que récemment (Griess–Ryba [22]) que la liste des S possibles a été complétée. Avant de donner cette liste (que l'on trouvera au § 2.4), je vais parler du cas $S = L_2(q)$, qui est le seul où l'on ait des énoncés généraux, i.e. valables aussi bien pour les groupes classiques que pour les groupes exceptionnels.

2.1. Plongements projectifs de $L_2(q)$ dans G ; énoncé du résultat

On suppose $q > 2$. Le groupe $S = L_2(q) = \text{PSL}_2(\mathbf{F}_q)$ est alors un groupe simple (sauf si $q = 3$, où c'est le groupe Alt_4), et toute extension centrale \bar{S} de S , égale à son groupe dérivé, est isomorphe, soit à $2 \cdot S = \text{SL}_2(\mathbf{F}_q)$, soit à S (sauf si $q = 4$ ou 9). On va donc s'intéresser aux homomorphismes

$$f : \text{SL}_2(\mathbf{F}_q) \longrightarrow G$$

de noyau égal à 1 ou à (± 1) . Un tel homomorphisme sera dit *non dégénéré*.

Écrivons q sous la forme p^e , avec p premier, $e \geq 1$. Le p -groupe de Sylow $U = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$ de $\text{SL}_2(\mathbf{F}_q)$ est isomorphe à \mathbf{F}_q . Le groupe $A = f(U)$ est un p -groupe abélien élémentaire de G de rang e . Nous dirons que f est *de type toral* si A est toral au sens du § 1.1. C'est le cas si $e = 1$, ou si $e = 2$ et G est simplement connexe (1.1.2), ou si p n'est pas un nombre premier de torsion pour G .

Nous allons donner un *critère pour l'existence d'un f non dégénéré de type toral*. Supposons G presque simple, de rang r ; soient k_i ($i = 1, \dots, r$) les exposants de son groupe de Weyl et soient $d_i = k_i + 1$ les degrés correspondants (Bourbaki, LIE V, § 6, prop. 3). L'énoncé suivant résume une série de résultats dus à divers auteurs ([2], [8], [9], [11], [12], [14], [19], [20], [24], [28], [31]) :

2.1.1. *Pour qu'il existe un homomorphisme non dégénéré de type toral de $\text{SL}_2(\mathbf{F}_q)$ dans G , il faut et il suffit que $q - 1$ divise l'un des entiers $2d_1, \dots, 2d_r$.*

Remarque.— Lorsque $p = 2$ ou 3 , il existe quelques plongements de $L_2(p^e)$ qui ne sont pas de type toral, par exemple :

$$\begin{aligned} L_2(4) &\longrightarrow \text{PGL}_2, & L_2(8) &\longrightarrow G_2, & L_2(16) &\longrightarrow D_8, & L_2(32) &\longrightarrow E_8; \\ L_2(9) &\longrightarrow \text{PGL}_3, & L_2(27) &\longrightarrow F_4. \end{aligned}$$

Je ne sais pas en donner de description systématique.

Exemples

1) Si $G = \mathrm{SL}_2$, on a $r = 1$ et $d_1 = 2$; la condition dit alors que $q - 1$ divise 4, d'où $q = 3$ et $q = 5$, ce qui donne des plongements de $\mathrm{SL}_2(\mathbf{F}_3)$ et $\mathrm{SL}_2(\mathbf{F}_5)$ dans SL_2 ; d'où des plongements de $\mathrm{PSL}_2(\mathbf{F}_3) = \mathrm{Alt}_4$ et de $\mathrm{PSL}_2(\mathbf{F}_5) = \mathrm{Alt}_5$ dans PGL_2 . On retrouve ainsi les groupes du tétraèdre et de l'icosaèdre (quant au groupe du cube, Sym_4 , il s'interprète aussi comme $\mathrm{PGL}_2(\mathbf{F}_3)$, et c'est le normalisateur du groupe Alt_4).

2) Si $G = G_2$, on a $r = 2$, $d_1 = 2$, $d_2 = 6$; la condition dit que $q - 1$ divise 12, ce qui donne des plongements projectifs pour $q = 3, 5, 7, 13$. En fait, si $q = 7$ ou 13, ces plongements projectifs sont de vrais plongements de $L_2(q)$, car sinon leurs images seraient contenues dans le centralisateur d'un élément d'ordre 2, qui est de type $A_1 \cdot A_1$, et cela contredirait l'exemple 1. (Ce genre d'argument s'applique à beaucoup d'autres cas : les plongements projectifs intéressants sont de vrais plongements.)

3) Si $G = E_8$, on a $(d_1, \dots, d_8) = (2, 8, 12, 14, 18, 20, 24, 30)$ et l'on en déduit notamment des plongements de $L_2(q)$ pour $q = 16, 31, 41, 49, 61$.

4) Le plus grand des entiers d_i est le *nombre de Coxeter* h , égal à $(\dim G)/r - 1$. L'énoncé 2.1.1 contient donc comme cas particulier la *conjecture de Kostant* : *si $q = 2h + 1$ est une puissance d'un nombre premier, le groupe G^{ad} contient un sous-groupe isomorphe à $L_2(q)$.*

5) On a un énoncé analogue à celui de Kostant lorsque $h + 1$ est une puissance d'un nombre premier, d'où par exemple $L_2(19) \rightarrow E_7^{\mathrm{ad}}$ et $L_2(31) \rightarrow E_8$. Lorsque $h + 1$ est égal à un nombre premier p , on a un résultat plus précis (cf. [31]) : le groupe $\mathrm{PGL}_2(\mathbf{F}_p)$ (qui est "deux fois plus grand" que $\mathrm{PSL}_2(\mathbf{F}_p)$) est, lui aussi, plongeable dans G^{ad} . Lorsque $G = \mathrm{PGL}_2$, on retrouve le groupe du cube $\mathrm{PGL}_2(\mathbf{F}_3)$, cf. exemple 1. De ce point de vue, on peut dire que "les analogues" pour E_8 des groupes Alt_4 , Sym_4 , et Alt_5 du début de l'exposé sont respectivement $\mathrm{PSL}_2(\mathbf{F}_{31})$, $\mathrm{PGL}_2(\mathbf{F}_{31})$ et $\mathrm{PSL}_2(\mathbf{F}_{61})$.

2.2. Le critère 2.1.1 : démonstration de la nécessité

Il s'agit de prouver que, si $f : \mathrm{SL}_2(\mathbf{F}_q) \rightarrow G$ est un homomorphisme non dégénéré de type toral, alors $q - 1$ divise l'un des entiers $2d_i$.

On utilise :

2.2.1. *Soit A un p -sous-groupe élémentaire toral de G , et soit $g \in N_G(A)$. Soit $I_g \in \mathrm{GL}(A)$ l'automorphisme de A (vu comme espace vectoriel sur \mathbf{F}_p) défini par la conjugaison par g . Soit λ une valeur propre de I_g dans $\overline{\mathbf{F}}_p$ et soit m l'ordre de λ (dans $\overline{\mathbf{F}}_p^*$). Alors m divise l'un des d_i .*

(On peut supposer que A est contenu dans T ; d'après 1.1.1, il existe $w \in W$ qui induit I_g sur A . L'une des valeurs propres de w en caractéristique 0 a pour réduction λ en caractéristique p . Son ordre est donc de la forme mp^a , avec $a \geq 0$. D'après un théorème de Springer ([32], th. 3.4 (i)), mp^a divise l'un des d_i . Il en est donc de même de m .)

Revenons à f , et au p -Sylow $U = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$. Soit $A = f(U)$, et soit $g = f(h)$, où $h = \begin{pmatrix} c & 0 \\ 0 & c^{-1} \end{pmatrix}$ est un générateur du sous-groupe diagonal de $\mathrm{SL}_2(\mathbf{F}_q)$. Si l'on identifie U à \mathbf{F}_q , l'action de h sur ce groupe est l'homothétie de rapport c^2 ; ses valeurs propres (dans $\overline{\mathbf{F}_p}$) sont les conjugués de c^2 , qui sont d'ordre $m = (q-1)/2$ si $p > 2$ et $m = q-1$ si $p = 2$. En appliquant 2.2.1 à A et g , on voit que m divise l'un des entiers d_i ; donc $q-1$ divise l'un des entiers $2d_i$.

Remarque.— Le même argument montre que, si f est un homomorphisme de $\mathrm{GL}_2(\mathbf{F}_q)$ dans G , qui est non dégénéré et de type toral (en un sens évident), alors $q-1$ divise l'un des d_i .

2.3. Le critère 2.1.1 : vérification de la suffisance

On doit montrer que, si $q-1$ divise l'un des entiers $2d_i$, il existe $f : \mathrm{SL}_2(\mathbf{F}_q) \rightarrow G$ qui est non dégénéré de type toral.

On ne connaît pas de démonstration générale de cet énoncé. On procède cas par cas :

1) Le cas où G est de type classique se traite facilement, grâce à la connaissance de la table des caractères de $\mathrm{SL}_2(\mathbf{F}_q)$; les caractères irréductibles de degré $(q \pm 1)/2$ sont particulièrement utiles. La condition de toralité est trivialement satisfaite si $p \neq 2$; dans le cas où $p = 2$, et où G est un groupe orthogonal, il faut faire un peu attention. La même méthode s'applique à G_2 (voir aussi [2], [11], [28]).

2) Pour les groupes exceptionnels, les inclusions des groupes classiques dans ceux-ci, et les plongements

$$G_2 \longrightarrow F_4 \longrightarrow E_6 \longrightarrow E_7 \longrightarrow E_8,$$

montrent qu'il suffit de traiter les cas suivants :

$$F_4 (q = 25) \quad ; \quad E_6 (q = 19) \quad ; \quad E_7 (q = 29, 37) \quad ; \quad E_8 (q = 16, 31, 41, 49, 61).$$

Le cas $(E_8 ; 16)$ se traite en remarquant que $L_2(16)$ se plonge dans un groupe de type D_8 , donc dans un groupe de type E_8 ; ce plongement n'est pas de type toral dans D_8 , mais il le devient dans E_8 comme on le voit en appliquant [8], prop. 3.8.

Le cas $(F_4 ; 25)$ se déduit de ce que $L_2(25)$ se plonge dans le groupe de Tits ${}^2F_4(2)'$, qui lui-même se plonge dans E_6 , cf. 2.4.2, b) ci-après. On vérifie par un calcul de caractères que le sous-groupe de E_6 ainsi obtenu est contenu dans un conjugué de F_4 , cf. Cohen–Wales [14].

Les cas $(E_6 ; 19)$, $(E_7 ; 37)$, $(E_8 ; 31)$, $(E_8 ; 41)$, $(E_8 ; 49)$, $(E_8 ; 61)$ ont été vérifiés par des calculs sur ordinateur, cf. Cohen–Wales [14], Kleidman–Ryba [24], Griess–Ryba [19] et [20], Cohen–Griess–Lisser [9].

Les cas $(E_7 ; 37)$, $(E_8 ; 31)$ et $(E_8 ; 61)$ sont traités dans [31] par une méthode p -adique qui consiste à relever un plongement (bien choisi) de la caractéristique p à la caractéristique 0. Une variante non encore publiée de cette méthode permet de traiter

aussi $(E_6; 19)$, $(E_7; 29)$ et $(E_8; 41)$. Ainsi, tous les cas où q est premier peuvent être obtenus sans ordinateur.

Remarque.— Les calculs sur ordinateur ont un inconvénient évident : ils ne sont pas vérifiables pas à pas, comme une démonstration doit l'être. Ils ont toutefois un avantage : dans certains cas, ils montrent l'unicité du plongement considéré (à conjugaison près), cf. [19], [20]. C'est là un résultat que la méthode p -adique ne donne pas, au moins pour le moment.

2.4. Plongements projectifs des groupes finis simples dans les groupes de type exceptionnel

La table suivante est extraite de Griess-Ryba [22] (avec une petite correction relative à F_4). Elle donne la liste des groupes simples ayant un plongement projectif dans G_2, \dots, E_8 . Je renvoie à [22] et [27] pour divers renseignements supplémentaires sur ces plongements, ainsi que pour des références.

Table

G_2 – Alt_n , $n = 5, 6$; $L_2(q)$, $q = 7, 8, 13$; $\text{SU}_3(3) = G_2(2)'$.

F_4 – ceux de G_2 et : Alt_n , $n = 7, 8, 9, 10$; $L_2(q)$, $q = 17, 25, 27$; $L_3(3)$; $\text{SU}_4(2)$; $\text{Sp}_6(2) = O_7(2)$; $O_8^+(2)$; ${}^3D_4(2)$.

E_6 – ceux de F_4 et : Alt_{11} ; $L_2(q)$, $q = 11, 19$; $L_3(4)$; $\text{PSU}_4(3)$; ${}^2F_4(2)'$; M_{11} ; $HJ = J_2$.

E_7 – ceux de E_6 et : Alt_n , $n = 12, 13$; $L_2(q)$, $q = 29, 37$; $\text{PSU}_3(8)$; M_{12} .

E_8 – ceux de E_7 et : Alt_n , $n = 14, 15, 16, 17$; $L_2(q)$, $q = 16, 31, 32, 41, 49, 61$; $L_3(5)$; $\text{PSp}_4(5)$; $G_2(3)$; ${}^2B_2(8) = Sz(8)$.

(Les notations sont celles de l'ATLAS [15]. En particulier $L_n(q)$ désigne le groupe $\text{PSL}_n(\mathbf{F}_q)$. Vu que $\text{Alt}_5 = L_2(4) = L_2(5)$ et $\text{Alt}_6 = L_2(9)$, la liste pour G_2 pourrait aussi être écrite :

$$G_2 - L_2(q), q = 4, 5, 7, 8, 9, 13 ; \text{SU}_3(3) = G_2(2)'$$

De même, pour F_4 , on peut remplacer Alt_8 par $L_4(2)$.)

La vérification de l'exactitude de cette table comporte deux parties. Tout d'abord :

2.4.1. Un groupe simple qui ne figure pas dans la table n'a pas de plongement projectif dans G .

Comme on peut s'y attendre, le point de départ est la *classification des groupes simples finis*, qui est admise (le lecteur curieux de savoir quelle partie de cette classification reste à démontrer pourra consulter Aschbacher [3]). Cela permet de passer en revue les différents cas possibles : groupes alternés, groupes de type algébrique, groupes sporadiques. Pour éliminer un groupe S , on utilise des arguments variés, par exemple 1.2.2 ou 2.2.1 (qui suffisent si le groupe est très gros), ou (dans les cas difficiles) la table des caractères du groupe. C'est un travail délicat. La moindre erreur peut conduire à éliminer à tort le groupe en question. C'est ce qui s'était passé

dans une liste précédente [8] pour les groupes $L_2(41)$, $L_2(49)$ et $Sz(8)$ qui avaient été déclarés non plongeables dans E_8 .

2.4.2. *Tout groupe figurant dans la table a au moins un plongement projectif dans G .*

On utilise différentes méthodes. Par exemple :

a) Le cas le plus facile est celui où l'on connaît un sous-groupe de G dans lequel S a un plongement projectif. Ainsi, pour traiter le cas de Alt_{10} et F_4 , il suffit de remarquer que Alt_{10} a une représentation orthogonale évidente de degré 9, autrement dit se plonge dans un groupe de type B_4 , et l'on utilise le plongement de B_4 dans F_4 .

b) Certains cas peuvent se traiter à partir de la table des caractères de S (et de ses extensions centrales). Outre $G = G_2$, déjà signalé, il faut mentionner le cas où $G = E_6$ et où S est le groupe de Tits ${}^2F_4(2)'$ (Cohen–Wales [14]). On part du fait que le groupe $S \cdot 2 = {}^2F_4(2)$ a une représentation irréductible V de dimension 78 (cf. [15], p. 75). Un calcul de caractères montre que $\wedge^2 V$ contient V ; il existe donc un homomorphisme non nul $\wedge^2 V \rightarrow V$ compatible avec l'action de $S \cdot 2$, et un autre calcul de caractères montre que l'identité de Jacobi est satisfaite. D'où une structure d'algèbre de Lie sur V . Il est clair que cette algèbre de Lie est simple ; puisqu'elle est de dimension 78, elle est de type B_6 , C_6 ou E_6 . On élimine les types B_6 et C_6 qui conduiraient à des représentations de $S \cdot 2$ de degré trop petit. L'algèbre de Lie V est donc de type E_6 , ce qui fournit un plongement de $S \cdot 2$ dans E_6^{ad} , donc *a fortiori* un plongement de S . (On aimerait avoir davantage d'exemples de ce genre !)

c) La plupart des autres plongements ont été construits au moyen de calculs sur ordinateur. Je renvoie à [22] pour une description des méthodes employées. Je signale seulement que les calculs ne se font pas sur le corps k , mais sur un corps fini \mathbf{F}_ℓ , où ℓ est un nombre premier ne divisant pas l'ordre de S et tel que \mathbf{F}_ℓ contienne les racines de l'unité intervenant dans la construction : ainsi, pour plonger $L_2(61)$ dans E_8 , Cohen–Griess–Lisser [9] choisissent $\ell = 1831$. Le relèvement de \mathbf{F}_ℓ à \mathbf{Z}_ℓ (donc à la caractéristique 0) ne présente aucune difficulté vu que ℓ ne divise pas $|S|$. Il semble que, dans chaque cas, le calcul comporte suffisamment de vérifications internes pour qu'on puisse lui faire confiance.

2.5. Compléments

2.5.1. Classification en caractéristique > 0

L'analogie de 2.4 en caractéristique p a été fait par Liebeck–Seitz [27]. Tout groupe S intervenant en caractéristique 0 intervient aussi en caractéristique p (quel que soit p) ; c'est là une conséquence simple de la théorie de Bruhat–Tits, cf. [31], § 5. Outre ces groupes, et ceux qui sont “de caractéristique p ”, Liebeck–Seitz donnent la liste suivante :

$\mathbf{G}_2 - p = 2 : J_2 ; p = 5 : \text{Alt}_7 ; p = 11 : J_1.$

$\mathbf{F}_4 - \text{ceux de } G_2 \text{ et } p = 2 : L_4(3) ; p = 3 : L_3(4) ; p = 5 : Sz(8) ; p = 11 : M_{11}, \text{Alt}_{11}.$

$\mathbf{E}_6 - \text{ceux de } F_4 \text{ et } p = 2 : M_{12}, \text{Alt}_{12}, G_2(3), O_7(3), M_{22}, J_3, Fi_{22} ; p = 3 : M_{12}, \text{Alt}_{12} ; p = 5 : M_{12} ; p = 7 : M_{22}.$

$\mathbf{E}_7 - \text{ceux de } E_6 \text{ et } p = 5 : M_{22}, Ru, HS ; p = 7 : \text{Alt}_{14}.$

$\mathbf{E}_8 - \text{ceux de } E_7 \text{ et } p = 2 : L_4(5) ; p = 3 : \text{Alt}_{18}, Th ; p = 5 : Sz(32).$

Noter en particulier le groupe de Janko J_1 dans $G_2(\mathbf{F}_{11})$ et le groupe de Thompson Th dans $E_8(\mathbf{F}_3)$.

2.5.2. Classes de conjugaison de plongements

On aimerait pouvoir compléter la table 2.4 en décrivant les plongements à conjugaison près. Cela a été fait dans certains cas, mais pas dans tous, cf. [22]. Le cas des plongements de Alt_5 dans E_8 est particulièrement intéressant (cf. Frey [16]) ; on peut déterminer les triplets (x, y, z) de classes de conjugaison de E_8 d'ordres $(2, 3, 5)$ qui sont représentables dans un même sous-groupe Alt_5 . Pour tous ces triplets, sauf un (celui appelé “844” dans [16]), Frey détermine le nombre de classes de conjugaison correspondantes (une ou deux). Par contre, pour le cas “844” (qui est le seul où le centralisateur du sous-groupe Alt_5 soit fini), on ne sait pas combien il y a de classes de conjugaison ; on dispose de plusieurs tels sous-groupes (par exemple un sous-groupe du groupe de Borovik [7], ou un sous-groupe de $L_2(41)$, ou de $L_2(61), \dots$), mais il n'est pas facile de voir s'ils sont ou non conjugués. Comme Alt_5 admet la présentation :

$$(x, y, z \mid x^2 = y^3 = z^5 = 1, xyz = 1),$$

c'est là un problème analogue à celui de la “rigidité” intervenant pour la classification des revêtements galoisiens de la droite projective ramifiés en 3 points.

2.5.3. Rationalité

On sait que G provient par extension des scalaires d'un groupe *déployé* G_{dep} défini sur \mathbf{Q} . Si S (ou \bar{S}) est plongeable dans $G(k)$, on peut se demander quels sont les sous-corps k' de k tels que S soit plongeable dans $G_{\text{dep}}(k')$. Cette question est étroitement liée à la précédente (celle des classes de conjugaison) : voir là-dessus [19], App. 2. Voici un exemple typique :

D'après Aschbacher [2], le groupe $S \cdot 2 = G_2(2)$ admet un plongement dans $G_2(k)$, et un seul, à conjugaison près. Or, à la fois $S \cdot 2$ et G_2 ont un centre trivial, et pas d'automorphisme externe. De plus, le centralisateur de $2 \cdot S$ est trivial. Soit P l'ensemble de ces plongements ; c'est un G_2 -torseur qui est défini de façon naturelle sur \mathbf{Q} . Il définit donc une \mathbf{Q} -forme G_2^0 de G_2 , et l'on peut plonger $S \cdot 2$ dans $G_2^0(\mathbf{Q})$ par définition même de G_2^0 . Or, il n'y a que deux formes de G_2 sur \mathbf{Q} , que l'on distingue par leurs points réels ; la forme déployée ne peut pas contenir $2 \cdot S$: son compact maximal est trop petit. Ainsi, G_2^0 est la forme non déployée de G_2 , celle qui correspond aux octonions usuels. On conclut de là que le plongement cherché de $S \cdot 2$ dans $G_{\text{dep}}(k')$ existe si et seulement si G_{dep} et G_2^0 sont k' -isomorphes, i.e. *si et seulement si -1 est somme de 4 carrés dans k'* . Un argument analogue montre que $L_2(13)$ est plongeable dans $G_{\text{dep}}(k')$ si et seulement si k' contient $\sqrt{13}$ et -1 est somme de 4 carrés dans k' ; même chose pour $L_2(8)$, avec $\sqrt{13}$ remplacé par $z_9 + \bar{z}_9$, où z_9 est une racine primitive 9-ème de l'unité. (Noter l'analogie de ces énoncés avec le suivant, connu depuis longtemps : Alt_4 , Sym_4 et Alt_5 sont plongeables dans $\text{PGL}_2(k')$ si et seulement si -1 est somme de 2 carrés dans k' et (pour Alt_5) k' contient $\sqrt{5}$.)

BIBLIOGRAPHIE

- [1] J.F. ADAMS - *2-tori in E_8* , Math. Ann. **287** (1987), 29-39 (= *Selected Works*, vol. II, 264-274).
- [2] M. ASCHBACHER - *Chevalley groups of type G_2 as the group of a trilinear form*, J. Alg. **109** (1987), 193-259.
- [3] M. ASCHBACHER - *Quasithin groups*, in *Algebraic Groups and their Representations* (R. Carter and J. Saxl edit.), NATO AS series, vol. **517**, 321-340, Kluwer, 1998.
- [4] A. BOREL - *Sous-groupes commutatifs et torsion des groupes de Lie compacts connexes*, Tôhoku Math. J. **13** (1961), 216-240 (= *Oe. II*, n° 53 et *Commentaires*, 775-777).
- [5] A. BOREL and G.D. MOSTOW - *On semi-simple automorphisms of Lie algebras*, Ann. Math. **61** (1955), 389-405 (= A. Borel, *Oe. I*, n° 36).
- [6] A. BOREL et J.-P. SERRE - *Sur certains sous-groupes des groupes de Lie compacts*, Comm. Math. Helv. **27** (1953), 128-139 (= A. Borel, *Oe. I*, n° 24).
- [7] A.V. BOROVIK - *A maximal subgroup in the simple finite group $E_8(q)$* , Contemp. Math. A.M.S. **131** (1992), vol. I, 67-79.
- [8] A.M. COHEN and R.L. GRIESS, Jr - *On finite simple subgroups of the complex Lie groups of type E_8* , AMS Proc. Symp. Pure Math. **47** (1987), vol. II, 367-405.
- [9] A.M. COHEN, R.L. GRIESS, Jr and B. LISSER - *The group $L(2, 61)$ embeds in the Lie group of type E_8* , Comm. Alg. **21** (1993), 1889-1907.

- [10] A.M. COHEN and G.M. SEITZ - *The r -rank of the groups of exceptional Lie type*, Indag. Math. **49** (1987), 251-259.
- [11] A.M. COHEN and D.B. WALES - *Finite subgroups of $G_2(\mathbf{C})$* , Comm. Alg. **11** (1983), 441-459.
- [12] A.M. COHEN and D.B. WALES - *Embeddings of the group $L(2, 13)$ in groups of Lie type E_6* , Israel J. Math. **82** (1993), 45-86.
- [13] A.M. COHEN and D.B. WALES - *Finite simple subgroups of semisimple complex Lie groups - a survey*, in *Groups of Lie type and their geometries*, LMS Lect. Notes **207** (1995), 77-96.
- [14] A.M. COHEN and D.B. WALES - *Finite subgroups of $F_4(\mathbf{C})$ and $E_6(\mathbf{C})$* , Proc. London Math. Soc. **74** (1997), 105-150.
- [15] J.H. CONWAY, R.T. CURTIS, S.P. NORTON, R.A. PARKER and R.A. WILSON - *Atlas of Finite Groups*, Clarendon Press, Oxford, 1985.
- [16] D. FREY - *Conjugacy of alternating groups of degree 5 and $SL(2, 5)$ subgroups of the complex Lie groups of type E_8* , Memoirs AMS **634** (1998).
- [17] R.L. GRIESS, Jr - *Elementary abelian subgroups of algebraic groups*, Geom. Dedicata **39** (1991), 253-305.
- [18] R.L. GRIESS, Jr and A.J.E. RYBA - *Embeddings of $U(3, 8)$, $Sz(8)$ and the Rudvalis group in algebraic groups of type E_7* , Invent. math. **116** (1994), 215-241.
- [19] R.L. GRIESS, Jr and A.J.E. RYBA - *Embeddings of $PGL(2, 31)$ and $SL(2, 32)$ in $E_8(\mathbf{C})$* , Duke Math. J. **94** (1998), 181-211.
- [20] R.L. GRIESS, Jr and A.J.E. RYBA - *Embeddings of $PSL(2, 41)$ and $PSL(2, 49)$ in $E_8(\mathbf{C})$* , J. Symb. Comp. **11** (1999), 1-17.
- [21] R.L. GRIESS, Jr and A.J.E. RYBA - *Embeddings of $Sz(8)$ into exceptional Lie groups*, soumis au J. Crelle.
- [22] R.L. GRIESS, Jr and A.J.E. RYBA - *Finite simple groups which projectively embed in an exceptional Lie group are classified !*, Bull. AMS **36** (1999), 75-93.
- [23] A. GROTHENDIECK - *Torsion homologique et sections rationnelles*, Sém. Chevalley (1958), *Anneaux de Chow et Applications*, exposé 5.
- [24] P.B. KLEIDMAN and A.J.E. RYBA - *Kostant's conjecture holds for E_7 : $L_2(37) < E_7(\mathbf{C})$* , J. Alg. **161** (1993), 535-540.
- [25] M.-A. KNUS, A. MERKURJEV, M. ROST and J.-P. TIGNOL - *The Book of Involutions*, AMS Colloquium Publ. **44**, 1998.
- [26] M.W. LIEBECK - *Subgroups of exceptional groups*, in *Algebraic Groups and their Representations* (R.W. Carter and J. Saxl edit.), NATO AS series, vol. **517**, 275-290, Kluwer, 1998.
- [27] M.W. LIEBECK and G.M. SEITZ - *On finite subgroups of exceptional algebraic groups*, 1999, à paraître au J. Crelle.
- [28] A. MEURMAN - *An embedding of $PSL(2, 13)$ in $G_2(\mathbf{C})$* , Lect. Notes in Math. **933** (1982), 157-162.
- [29] Z. REICHSTEIN and B. YOUSSEIN - *Essential dimensions of algebraic groups and a resolution theorem for G -varieties*, 1999, à paraître.
- [30] J.-P. SERRE - *Cohomologie galoisienne : progrès et problèmes*, Sém. Bourbaki 1993-94, exposé n° 783 (SMF, Astérisque **227** (1995), 229-257).

- [31] J.-P. SERRE - *Exemples de plongements des groupes $\mathbf{PSL}_2(\mathbf{F}_p)$ dans des groupes de Lie simples*, Invent. math. **124** (1996), 525-562.
- [32] T.A. SPRINGER - *Regular elements of finite reflection groups*, Invent. math. **25** (1974), 159-198.
- [33] T.A. SPRINGER and R. STEINBERG - *Conjugacy classes*, Lect. Notes in Math. **131** (1970), 281-312 (= R. Steinberg, *C.P.*, 293-323).
- [34] R. STEINBERG - *Torsion in reductive groups*, Adv. in Math. **15** (1975), 63-92 (= *C.P.*, 415-444).
- [35] J. TITS - *Normalisateurs de tores, I. Groupes de Coxeter étendus*, J. Alg. **4** (1966), 96-116.
- [36] J. TITS - *Sur les constantes de structure et le théorème d'existence des algèbres de Lie semi-simples*, Publ. Math. IHES **31** (1966), 21-58.

Jean-Pierre SERRE
Collège de France
3, rue d'Ulm
F-75005 PARIS
E-mail : serre@dmi.ens.fr

L'ORDRE DE DEHORNOY SUR LES TRESSSES

par Christian KASSEL

Une des découvertes récentes les plus intéressantes et inattendues concernant le groupe des tresses B_n d'Artin est due à Dehornoy qui a démontré fin 1991 que B_n a un ordre total invariant par multiplication à gauche. Il en résulte notamment que l'anneau du groupe B_n n'a pas de diviseurs de zéro. Dehornoy déduit cet ordre de l'étude générale des systèmes autodistributifs, définis comme des ensembles munis d'une loi de composition vérifiant l'identité $x(yz) = (xy)(xz)$. Des travaux sur un axiome indémontrable de théorie des ensembles avaient mis en évidence un système autodistributif remarquable et suggéré que l'étude de tels systèmes devait donner des résultats non triviaux.

Fenn, Greene, Rolfsen, Rourke et Wiest ont donné en 1997 une construction géométrique de l'ordre de Dehornoy. Leurs techniques ont permis également de mettre des structures d'ordre sur des groupes de difféotopies de surfaces. Plus récemment, Thurston a indiqué comment construire un ordre total sur B_n en utilisant les travaux de Nielsen sur les surfaces.

Le but de cet exposé est de présenter ces travaux ainsi que le lien inattendu avec la théorie des ensembles. Au n°1 nous énonçons le résultat principal sur l'ordre des tresses et quelques propriétés de cet ordre. Les deux numéros suivants sont consacrés à une présentation des techniques de Dehornoy et à sa démonstration du résultat principal ; le lien entre tresses et systèmes autodistributifs est exposé au n° 2 tandis qu'au n° 3 nous étudions deux systèmes autodistributifs particulièrement importants. Au n° 4 nous indiquons comment l'étude des grands cardinaux en théorie des ensembles a été à l'origine de ces travaux. La construction géométrique de Fenn *et al.* ainsi que l'approche à la Nielsen seront présentées au n° 5. Pour plus de détails, le lecteur pourra consulter les références citées, et notamment la monographie [D00] en cours d'achèvement.

Thomas Delzant m'a aidé à comprendre l'idée de Thurston et à mettre le n° 5.2 en forme. Je le remercie très chaleureusement ainsi que Patrick Dehornoy, Dale Rolfsen et Bertold Wiest qui m'ont fait profiter de leurs commentaires.

1. LES RÉSULTATS PRINCIPAUX

Dans ce numéro nous rappelons la définition des groupes de tresses et celle des groupes ordonnables avant d'énoncer le théorème principal de Dehornoy impliquant l'existence d'un ordre sur les tresses.

1.1. Le groupe des tresses d'Artin. Soit n un entier ≥ 2 . Définissons B_n comme le groupe engendré par $n - 1$ générateurs $\sigma_1, \dots, \sigma_{n-1}$ soumis aux relations

$$(1.1) \quad \sigma_i \sigma_j = \sigma_j \sigma_i$$

si $|i - j| > 1$ et

$$(1.2) \quad \sigma_i \sigma_j \sigma_i = \sigma_j \sigma_i \sigma_j$$

si $|i - j| = 1$. On utilisera souvent le mot *tresse* pour désigner un élément de B_n .

E. Artin a introduit le groupe B_n dans les années 1920 et a résolu le problème des mots¹ pour ce groupe (voir [A26], [A47]). En 1967 Garside [Ga] résout à la fois ce problème et le problème de conjugaison¹ (voir aussi [Mk]). D'autres solutions ont été proposées, fondées notamment sur une forme normale due à Adyan [Ad], et redécouverte un peu plus tard par ElRifai et Morton [EM] et par Thurston, ce dernier élaborant à cette occasion la théorie des groupes automatiques et établissant l'automaticité de B_n (les résultats de Thurston sont exposés au chapitre 9 de [E+]). Rappelons aussi que les groupes de tresses jouent un rôle important dans des branches des mathématiques aussi diverses que la théorie des nœuds, la théorie de l'homotopie, la géométrie algébrique, la théorie des groupes, celle des représentations, *etc.* (on trouvera un reflet de cette diversité dans [B+] et dans [Ca]).

Un *mot de tresse* w est un mot formé à partir des générateurs $\sigma_1, \dots, \sigma_{n-1}$ et de leurs inverses. Il est commode de représenter w par une configuration plane formée par n intervalles, appelés les brins de w , plongés de manière lisse dans une bande verticale bornée du plan. Une telle configuration s'obtient comme suit.

¹ Résoudre le problème des mots (resp. le problème de conjugaison) dans un groupe, c'est trouver un algorithme permettant de reconnaître ou de décider si deux mots donnés représentent ou non le même élément dans le groupe (resp. la même classe de conjugaison).

(a) On représente le générateur σ_i et son inverse σ_i^{-1} par les configurations de la figure 1. Le croisement de la configuration de σ_i (resp. de σ_i^{-1}) est dit *positif* (resp. *négatif*).

(b) Si D est la configuration associée au mot w et D' celle du mot w' , alors la configuration associée à ww' est obtenue en plaçant D au-dessus de D' et en joignant les extrémités inférieures des brins de D aux extrémités supérieures de D' .

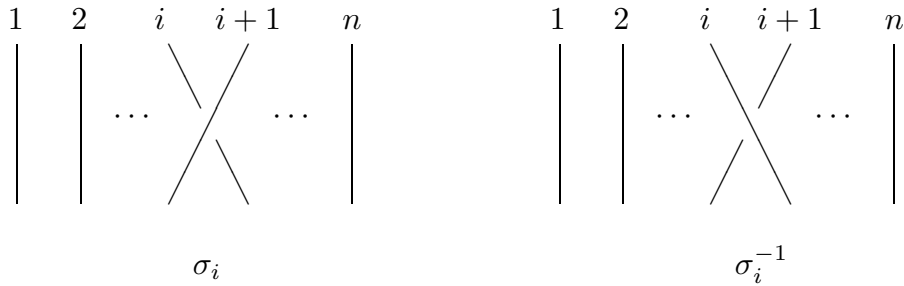


Figure 1

1.2. Groupes ordonnables. Un groupe G est dit *ordonnable* s'il possède un ordre total $<$ invariant par multiplication à gauche, c'est-à-dire vérifiant $a < b \Rightarrow ca < cb$ pour tout $a, b, c \in G$. Si, de plus, l'ordre est invariant par multiplication à droite (c'est-à-dire $a < b \Rightarrow ac < bc$), on dit que le groupe est *biordonnable*.

Pour tout groupe ordonnable G , notons P l'ensemble $\{x \in G \mid x > 1\}$ de ses éléments positifs. Le caractère total de l'ordre se traduit par la partition $G = P \amalg \{1\} \amalg P^{-1}$ où $P^{-1} = \{x \in G \mid x^{-1} \in P\}$. L'invariance par multiplication à gauche se traduit par $P^2 \subset P$, où P^2 est formé des produits de deux éléments de P dans G . Réciproquement, s'il existe un sous-ensemble P de G vérifiant

$$G = P \amalg \{1\} \amalg P^{-1} \quad \text{et} \quad P^2 \subset P,$$

alors G est ordonnable pour l'ordre défini par $x < y$ si $x^{-1}y \in P$. L'ordre en question fait de G un groupe biordonnable si et seulement si $xPx^{-1} \subset P$ pour tout $x \in G$.

L'ordonnabilité d'un groupe a de nombreuses conséquences (voir [MR], chap. 7 ou [Pa], chap. 13). D'abord, tout groupe ordonnable G est sans torsion. Si, de plus, R est un anneau sans diviseurs de zéro, alors l'algèbre du groupe $R[G]$ n'a pas de diviseurs de zéro (elle vérifie donc une fameuse conjecture de Kaplansky). Il en résulte que les seuls idempotents de $R[G]$ sont triviaux, à savoir 0 ou 1. En outre, les seuls éléments inversibles de $R[G]$ sont triviaux, c'est-à-dire de la forme rx où r est un élément inversible de R et $x \in G$.

Les groupes libres de type fini, les groupes fondamentaux de surfaces fermées, les groupes abéliens libres de type fini sont biordonnables. Pour ce qui concerne les

tresses, les groupes B_n ne sont pas biordonnables alors que les groupes de tresses pures le sont (*cf.* [KR] et [RZ]).

1.3. L'ordre de Dehornoy sur B_n . Pour $i \in \{1, \dots, n-1\}$, on dit qu'un mot de la forme $w_0 \sigma_i w_1 \sigma_i \dots \sigma_i w_r$ est σ_i -positif si les sous-mots w_0, w_1, \dots, w_r sont des mots dans les lettres $\sigma_j^{\pm 1}$ avec $j > i$. En d'autres termes, dans un mot σ_i -positif, le générateur σ_i d'indice minimal n'apparaît qu'avec des puissances positives. S'il n'apparaît qu'avec des puissances négatives, on dit que le mot est σ_i -négatif (donc un mot est σ_i -négatif si son inverse est σ_i -positif). Une tresse est dite σ_i -positive (resp. σ_i -négative) si elle peut être représentée par un mot σ_i -positif (resp. σ_i -négatif). Par exemple, le mot $\sigma_1 \sigma_2 \sigma_1^{-N}$ n'est pas σ_1 -positif si N est un entier ≥ 1 , mais la tresse qu'il représente l'est (le montrer !).

Nous dirons qu'une tresse est σ -positive (resp. σ -négative) s'il existe un entier i tel qu'elle est σ_i -positive (resp. σ_i -négative). On peut alors énoncer le théorème de Dehornoy.

Théorème 1.— *Toute tresse différente de 1 est soit σ -positive, soit σ -négative.*

La démonstration de ce théorème occupera les numéros 2 et 3. Celui-ci a pour conséquence les résultats nouveaux suivants non connus antérieurement.

Corollaire.— (a) *Pour tout n , le groupe B_n est ordonnable.*

(b) *Pour tout n et tout anneau R sans diviseurs de zéro, l'algèbre $R[B_n]$ du groupe B_n n'a pas de diviseurs de zéro, ni d'unités ou d'idempotents non triviaux.*

DÉMONSTRATION.— (a) Soit P l'ensemble des tresses σ -positives de B_n . Il est clair que P^2 est inclus dans P . Le théorème 1 implique $B_n = P \amalg \{1\} \amalg P^{-1}$. Il en résulte que B_n est ordonnable.

(b) C'est une conséquence de (a) et des remarques faites au n° 1.2. /

1.4. Propriétés. L'ordre de Dehornoy ainsi défini est l'unique ordre total de B_n invariant par multiplication à gauche tel que $\beta \sigma_1 \beta' > 1$ pour tous les éléments β, β' du sous-groupe de B_n engendré par $\sigma_2, \dots, \sigma_{n-1}$. Dans [D97b] Dehornoy a construit un algorithme de résolution du problème des mots fondé sur l'ordre total des tresses. En pratique, cet algorithme est beaucoup plus efficace que les algorithmes antérieurs (mais le problème reste ouvert de montrer qu'il est quadratique).

Soit B_n^+ le sous-monoïde de B_n constitué des tresses représentées par des produits des générateurs $\sigma_1, \dots, \sigma_{n-1}$, mais non de leurs inverses. Laver [L96] a montré que l'ordre de Dehornoy prolonge l'ordre partiel considéré par ElRifai et Morton dans [EM] ; il en déduit le résultat remarquable suivant, qui distingue probablement cet ordre parmi les ordres de B_n mentionnés au n° 5.2.

Théorème 2.— *L'ordre de Dehornoy restreint à B_n^+ est un bon ordre.*

Rappelons qu'un ordre total sur un ensemble E est un *bon ordre* si toute partie de E a un élément minimal. Dans un bon ordre, chaque élément a un rang ; par exemple, le générateur σ_{n-1} de B_n^+ est le plus petit élément > 1 . Comme tout élément de B_n possède une écriture canonique comme fraction de deux éléments de B_n^+ (voir [E+]), le théorème de Laver permet de spécifier une tresse à l'aide d'un couple d'ordinaux. Burckel [Bu] a donné une autre preuve du théorème 2 et montré que le type de l'ordre de B_n^+ est $\omega^{\omega^{n-2}}$, où ω désigne le plus petit ordinal infini. Voir aussi l'article de Wiest [W] basé sur la définition géométrique de l'ordre de B_n présentée au n° 5.1.

2. COLORIAGE DES TRESSES ET SYSTÈMES AUTODISTRIBUTIFS

Le but du n° 2 est d'indiquer comment Dehornoy démontre le théorème 1.

2.1. Comment colorier les tresses ? L'idée n'est pas nouvelle ; elle a notamment été utilisée par Joyce [Joy], Matveev [Mt] et Brieskorn [Br] dans les années 1980. Partons d'un mot de tresse w représentant un élément de B_n . Colorions de gauche à droite les extrémités supérieures des brins de la configuration plane associée à w avec les éléments a_1, a_2, \dots, a_n d'un ensemble S . Si nous propageons ces couleurs le long des brins, nous lirons au bas de la configuration une suite de couleurs qui est une permutation de la suite a_1, a_2, \dots, a_n de départ : c'est évidemment la permutation sous-jacente à la tresse.

Les choses deviennent plus intéressantes si nous autorisons les couleurs à interagir à chaque croisement des brins. Supposons que les croisements soient exclusivement positifs, c'est-à-dire que le mot w soit un produit des générateurs σ_i et non de leurs inverses. On dira d'un tel mot qu'il est *positif*. Décidons qu'à chaque croisement la couleur a du brin inférieur reste inchangée tandis que la couleur b du brin supérieur s'amalgame avec la couleur a de l'autre brin pour former une nouvelle couleur $a*b$ qui dépend de a et de b (voir la figure 2). Ceci revient à munir S d'une loi de composition $(a, b) \mapsto a*b$. Procédant de cette manière pour tous les croisements, nous construisons ainsi une action (à droite) du monoïde des mots de tresse positifs à n brins sur la puissance n -ième S^n de l'ensemble S . Si on note ε le mot vide, alors cette action est définie par récurrence sur la longueur des mots par $(a_1, \dots, a_n) \varepsilon = (a_1, \dots, a_n)$ et

$$(2.1) \quad (a_1, \dots, a_n) \sigma_i w = (a_1, \dots, a_{i-1}, a_i * a_{i+1}, a_i, a_{i+2}, \dots, a_n) w$$

où $(a_1, a_2, \dots, a_n) \in S^n$, w est un mot de tresse et $1 \leq i < n$.

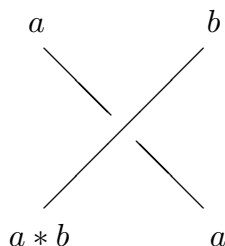


Figure 2

Au n° 1.4 nous avons défini B_n^+ comme le sous-monoïde du groupe B_n formé des tresses représentées par des mots positifs. On sait que le monoïde B_n^+ a une présentation donnée par les relations (1.1) et (1.2).

Lemme 1.— *La formule (2.1) définit une action du monoïde B_n^+ sur S^n si et seulement si, pour tout $a, b, c \in S$, on a*

$$(AD) \quad a * (b * c) = (a * b) * (a * b).$$

DÉMONSTRATION.— On se ramène aux calculs suivants où a, b, c, d désignent des éléments de l'ensemble S . Le premier calcul reflète la relation (1.1) :

$$(a, b, c, d) \sigma_1 \sigma_3 = (a * b, a, c * d, c) = (a, b, c, d) \sigma_3 \sigma_1.$$

La relation non triviale (1.2) montre la nécessité de l'identité (AD) :

$$(a, b, c) \sigma_1 \sigma_2 \sigma_1 = ((a * b) * (a * c), a * b, a),$$

$$(a, b, c) \sigma_2 \sigma_1 \sigma_2 = (a * (b * c), a * b, a). \quad \checkmark$$

Définition.— *Un système autodistributif (ou encore un système AD) est un ensemble muni d'une loi de composition vérifiant la relation (AD) du lemme 1.*

Par la suite, on utilisera les notions évidentes de morphisme de systèmes AD, de sous-système AD, de partie génératrice d'un système AD.

2.2. Ensembles automorphes. Dans ce numéro, nous donnons des exemples de systèmes autodistributifs. Il s'agit ici de ce que Brieskorn [Br] a appelé des *ensembles automorphes*, à savoir des systèmes autodistributifs pour lesquels les multiplications à gauche $b \mapsto a * b$ sont bijectives (ces systèmes particuliers ont également été utilisés sous d'autres noms, comme “wrack” chez Conway et Wraith, “rack” chez Fenn et

Rourke [FR], “crystal” chez Kauffman [Kau] ; on trouvera dans [FR] une introduction historique aux ensembles automorphes).

Si S est un ensemble automorphe, alors l'action de B_n^+ sur S^n , qui est à valeurs dans un ensemble de bijections, s'étend en une action du groupe B_n tout entier. On obtient ainsi une méthode systématique de construction de représentations de B_n , méthode dont les exemples (a) et (b) ci-dessous montrent l'intérêt.

(a) (*Conjugaison*) Tout groupe G est muni d'une structure de système AD dont la loi de composition $*$ est définie par $a * b = aba^{-1}$ où $a, b \in G$.

Si G est le groupe libre F_n sur n générateurs x_1, \dots, x_n , posons $(y_1, \dots, y_n) = (x_1, \dots, x_n) \beta$ pour tout $\beta \in B_n$. La correspondance $(x_1, \dots, x_n) \mapsto (y_1, \dots, y_n)$ définit un automorphisme $\hat{\beta}$ de F_n . L'application $\beta \mapsto \hat{\beta}^{-1}$ est l'homomorphisme injectif de groupes $B_n \rightarrow \text{Aut}(F_n)$ bien connu (voir par exemple [BZ], chap. 10).

(b) (*Barycentre*) L'anneau $\mathbf{Z}[t, t^{-1}]$ des polynômes de Laurent à coefficients entiers a une structure de système AD avec $a * b = (1 - t)a + tb$ où $a, b \in \mathbf{Z}[t, t^{-1}]$. L'action linéaire de B_n^+ sur $\mathbf{Z}[t, t^{-1}]^n$ s'étend en un homomorphisme de groupes $B_n \rightarrow \text{GL}_n(\mathbf{Z}[t, t^{-1}])$: c'est la *représentation de Burau*.

Les ensembles automorphes pour lesquels $a * a = a$ pour tout a , comme dans les exemples (a) et (b), ont été appelés “quandles” par Joyce [Joy] et “distributive groupoids” par Matveev [Mt] (tous deux ont utilisé ce type de systèmes AD pour classifier les nœuds dans \mathbf{R}^3). Voici deux autres exemples d'ensembles automorphes.

(c) (*Systèmes de racines*) Les vecteurs non nuls d'un espace vectoriel euclidien forment un ensemble automorphe : si a et b sont non nuls, on définit $a * b$ comme l'image de b par la symétrie orthogonale par rapport à l'hyperplan orthogonal à a . En particulier, tout système de racines est un ensemble automorphe.

(d) Si l'on note F_X le groupe libre engendré par les éléments d'un ensemble X , on peut munir l'ensemble-produit $B_X = F_X \times X$ d'une loi autodistributive par

$$(w_1, x_1) * (w_2, x_2) = (w_1 x_1 w_1^{-1} w_2, x_2)$$

où $w_1, w_2 \in F_X$ et $x_1, x_2 \in X$. Le système B_X est un ensemble automorphe libre ; tout ensemble automorphe S est quotient de B_X où X est une partie génératrice de S .

2.3. Cycles et acyclicité. Si a et b sont des éléments d'un système autodistributif S , nous dirons que a divise b (à gauche) s'il existe $c \in S$ tel que $b = a * c$. Un *cycle* est une suite (a_1, \dots, a_k) d'éléments de S tels que a_i divise a_{i+1} pour tout $i < k$ et a_k divise a_1 . On dit qu'un système AD est *acyclique* s'il ne possède pas de cycle. Un ensemble automorphe n'est pas acyclique puisque tout élément se divise lui-même.

L'observation suivante montre l'intérêt des systèmes acycliques. Sur un système autodistributif S on peut définir une relation binaire \prec comme suit : $a \prec b$ s'il existe un entier $k \geq 1$ et des éléments a_1, \dots, a_k de S tels que $a_1 = a$, $a_k = b$ et a_i divise a_{i+1} pour tout $i < k$. La relation \prec est transitive, et c'est une relation d'ordre partiel si S est acyclique. Avant les travaux de théorie des ensembles exposés au n° 4, on ne connaissait pas d'exemple de système acyclique.

Proposition 1.— *Il existe des systèmes autodistributifs acycliques.*

DÉMONSTRATION.— Il suffit d'exhiber un exemple. Voici celui de Larue [La] qui utilise l'image dans les automorphismes d'un groupe libre de l'opération sur les tresses que nous décrirons au n° 3.3. Soit F_∞ le groupe libre sur une infinité dénombrable de lettres x_1, x_2, \dots et $\text{Aut}(F_\infty)$ le groupe de ses automorphismes de groupe. Nous allons munir $\text{Aut}(F_\infty)$ d'une loi autodistributive $*$. Soit α l'automorphisme de F_∞ déterminé par

$$\alpha(x_i) = \begin{cases} x_1 x_2 x_1^{-1} & \text{si } i = 1, \\ x_1 & \text{si } i = 2, \\ x_i & \text{si } i > 2, \end{cases}$$

et T_0 l'endomorphisme de F_∞ donné par $T_0(x_i) = x_{i+1}$ pour tout $i \geq 0$. Ce dernier permet de définir l'endomorphisme T de $\text{Aut}(F_\infty)$ par les formules $T(\varphi)(x_1) = x_1$ et $T(\varphi)(x_i) = T_0(\varphi(x_{i-1}))$ si $i > 1$ et $\varphi \in \text{Aut}(F_\infty)$. Pour φ et $\psi \in \text{Aut}(F_\infty)$, posons

$$(2.2) \quad \varphi * \psi = \varphi \circ T(\psi) \circ \alpha \circ T(\varphi^{-1}).$$

On vérifie par un calcul direct que la loi de composition $*$ est autodistributive.

Nous affirmons que le système $(\text{Aut}(F_\infty), *)$ est acyclique. En effet, soit E l'ensemble des éléments de F_∞ dont l'expression réduite en les générateurs x_1, x_2, \dots et leurs inverses se termine par x_1^{-1} . On observe d'abord que α , ainsi que tout élément $\varphi \in \text{Aut}(F_\infty)$ qui est dans l'image du décalage T , préserve le sous-ensemble E . Supposons maintenant que le système $\text{Aut}(F_\infty)$ possède un cycle. Cette hypothèse se traduit par une égalité de la forme $\varphi = \varphi \circ T(\varphi_0) \circ \alpha \circ T(\varphi_1) \circ \dots \circ T(\varphi_{k-1}) \circ \alpha \circ T(\varphi_k)$, où $k \geq 1$ et $\varphi, \varphi_0, \dots, \varphi_{k-1}, \varphi_k$ sont des automorphismes de F_∞ . De manière équivalente, on a

$$(2.3) \quad \text{id} = T(\varphi_0) \circ \alpha \circ T(\varphi_1) \circ \dots \circ T(\varphi_{k-1}) \circ \alpha \circ T(\varphi_k)$$

où id représente l'automorphisme identité de F_∞ . Appliquons les deux membres de (2.3) à l'élément x_1 de F_∞ . A gauche, on obtient x_1 qui n'appartient pas au sous-ensemble E . Par contre, comme $(\alpha \circ T(\varphi_k))(x_1) = \alpha(x_1) = x_1 x_2 x_1^{-1}$ appartient à E , il résulte de l'observation ci-dessus que l'élément de F_∞ obtenu en appliquant le membre de droite de (2.3) à x_1 appartient à E . D'où une contradiction. \surd

2.4. Retournement des mots. Nous avons vu au n° 2.1 comment définir une action du monoïde B_n^+ sur la puissance n -ième d'un système AD S et au n° 2.2 que cette action s'étend au groupe B_n tout entier si (et seulement si) S est un ensemble automorphe. Or, à la vue des remarques faites au n° 2.3 et des travaux dont nous parlerons au n° 4, la question se pose d'étendre l'action du monoïde B_n^+ à B_n lorsque S est acyclique. En développant une méthode spécifique, dite de retournement des mots, Dehornoy a réussi à définir une action partielle de B_n sur S^n lorsque S est un système AD *simplifiable* (à gauche), c'est-à-dire un système pour lequel les multiplications à gauche sont injectives.

On dit qu'un mot de tresse w se transforme par retournement (à gauche) en le mot w' si w' s'obtient en remplaçant dans w des facteurs du type $\sigma_i \sigma_j^{-1}$ par le mot vide si $i = j$, par $\sigma_j^{-1} \sigma_i$ si $|i - j| > 1$, et par $\sigma_j^{-1} \sigma_i^{-1} \sigma_j \sigma_i$ si $|i - j| = 1$. Par définition du retournement, w et w' représentent le même élément dans B_n .

Garside [Ga] avait démontré que toute tresse est le quotient de deux tresses représentées par des mots de tresse positifs. Dehornoy reprend les ingrédients de l'approche de Garside, mais travaille sur les mots de tresse plutôt que sur les tresses, ce qui lui permet d'établir dans [D97a] le résultat plus fin suivant.

Lemme 2.— *Tout mot de tresse se transforme par retournement en un mot de la forme $u^{-1}v$ où u et v sont des mots de tresse positifs.*

Remarquons qu'il n'est pas évident *a priori* que l'algorithme de retournement converge en un nombre fini d'étapes car la longueur des mots peut augmenter à chaque retournement ; on pourra observer ce phénomène sur le mot $\sigma_2 \sigma_4 \sigma_6 \sigma_5^{-1} \sigma_3^{-1} \sigma_1^{-1}$. Signalons que des techniques analogues de retournement des mots permettent la construction de formes normales et de structures automatiques pour d'autres groupes dont les groupes d'Artin de groupe de Coxeter fini (voir [D98], [DP]).

Soit maintenant S un système AD simplifiable et n un entier ≥ 2 . L'action à droite $(\vec{a}, w) \mapsto \vec{a}w$ du monoïde B_n^+ sur le produit S^n , donnée par la formule (2.1), se prolonge à certains éléments (\vec{a}, w) de $S^n \times B_n$ de façon que les règles de coloriage restent vérifiées. C'est par exemple le cas de $(\vec{b}u, u^{-1})$ avec $b \in S^n$ et $u \in B_n^+$. Dans ce cas, on dira que l'élément $\vec{a}w$ est défini.

Lemme 3.— *Soit S un système AD simplifiable et $\vec{a} \in S^n$. Si le mot de tresse w se transforme par retournement en w' et si $\vec{a}w'$ est défini, alors $\vec{a}w$ est défini.*

Les lemmes 2 et 3 impliquent la proposition suivante selon laquelle on peut toujours trouver un ensemble de couleurs dans un système AD simplifiable pour colorier une tresse donnée en suivant la règle énoncée au n° 2.1.

Proposition 2.— *Soit S un système AD simplifiable. Pour tout mot de tresse w , il existe $\vec{a} \in S^n$ tel que $\vec{a}w$ soit défini. Si w et w' représentent la même tresse et que $\vec{a}w$ et $\vec{a}w'$ soient définis, alors $\vec{a}w = \vec{a}w'$.*

DÉMONSTRATION.— Contentons-nous d'établir la première assertion. D'après le lemme 2, w se transforme par retournement en $w' = u^{-1}v$ où u et v sont des mots positifs. Posons $\vec{a} = \vec{b}u$ où \vec{b} est n'importe quel élément de S^n . Alors $\vec{a}w' = \vec{b}v$ est défini. On termine en appliquant le lemme 3.

Pour établir la seconde assertion, on utilise une version à droite du retournement des mots de tresse et l'hypothèse que le système est simplifiable. /

2.5. Démonstration du théorème 1. Supposons que nous disposions d'un système autodistributif S acyclique et monogène (c'est-à-dire engendré par un seul élément) tel que l'ordre \prec associé (défini au n° 2.3) soit total. On vérifie que $a \prec b$ implique $c * a \prec c * b$, d'où l'on déduit que S est simplifiable.

Puisque S possède un ordre total, on peut munir chaque puissance S^n de S de l'ordre total lexicographique \prec_ℓ défini par $(a_1, \dots, a_n) \prec_\ell (b_1, \dots, b_n)$ s'il existe $i \geq 1$ tel que $a_j = b_j$ pour tout $j < i$ et $a_i \prec b_i$. Nous utilisons l'ordre total \prec_ℓ sur S^n pour définir une partie "positive" P de B_n comme suit. Si w est un mot de tresse, alors, d'après la proposition 2, il existe $\vec{a} \in S^n$ tel que $\vec{a}w$ soit défini. Nous dirons que la tresse représentée par w est dans P si $\vec{a} \prec_\ell \vec{a}w$.

Proposition 3.— *Une tresse est dans P si et seulement si elle est σ -positive. La relation $\beta < \beta'$ définie sur B_n par $\beta^{-1}\beta' \in P$ est une relation d'ordre total invariant par multiplication à gauche, qui coïncide avec l'ordre de Dehornoy du n° 1.3.*

DÉMONSTRATION.— Montrons que, si un mot de tresse est σ -positif, alors il représente une tresse du sous-ensemble P défini ci-dessus. Un mot σ_i -positif w est de la forme $w = w_0\sigma_i w_1\sigma_i \dots \sigma_i w_k$ où w_0, w_1, \dots, w_k sont des mots dans les lettres σ_j et σ_j^{-1} avec $j > i$. Soit $\vec{a} = (a_1, \dots, a_n) \in S^n$ tel que $\vec{a}w = (b_1, \dots, b_n) \in S^n$ soit défini. La formule (2.1) implique que $b_j = a_j$ pour tout $j < i$ et que b_i est de la forme $b_i = (\dots((a_1 * c_1) * c_2) * \dots) * c_k$ où $c_1, c_2, \dots, c_k \in S$ et k est le nombre d'occurrences de σ_i dans w . Par définition de l'ordre \prec de S , il en résulte que $a_i \prec b_i$. Par conséquent, $\vec{a} \prec_\ell \vec{a}w$ et $w \in P$.

La réciproque, à savoir qu'une tresse dans P est σ -positive, utilise la notion de tresse spéciale que nous introduirons au n° 3.3, et sera démontrée au n° 3.4.

Une fois l'équivalence précédente établie, il est clair que la relation $<$ est bien définie et que c'est l'ordre total du n° 1.3, dont l'existence est ainsi démontrée. /

3. SYSTÈME AUTODISTRIBUTIF LIBRE ET TRESSSES SPÉCIALES

Nous étudions maintenant deux systèmes autodistributifs particuliers. Le premier est libre et fournit le type de systèmes nécessaires à la démonstration du théorème 1, telle qu'elle est donnée au n° 2.5. Le second est constitué de tresses.

3.1. Le système autodistributif libre sur un générateur. Il est caractérisé par la propriété universelle suivante.

Proposition 4.— *Il existe un système autodistributif D_1 engendré par un singleton $\{x\}$ tel que, pour tout système autodistributif S et tout élément s de S , il existe un unique morphisme $f : D_1 \rightarrow S$ de systèmes AD vérifiant $f(x) = s$. Le système D_1 est unique à isomorphisme près.*

DÉMONSTRATION.— L'unicité est claire. Pour l'existence, on considère le magma libre M sur un générateur x (voir [Bo], I, § 7, n° 1). Les éléments de M sont des parenthésages complets de puissances strictement positives de la lettre x . Notons $*$ la loi de composition du magma. Sur M on considère la relation d'équivalence \equiv_{AD} invariante par composition et engendrée par les paires $(t_1*(t_2*t_3), (t_1*t_2)*(t_1*t_3))$. On définit D_1 comme l'ensemble des classes d'équivalence M/\equiv_{AD} . On vérifie aisément que D_1 est un système AD satisfaisant à la propriété universelle ci-dessus. \surd

Rappelons la relation binaire \prec introduite au n° 2.3. Le théorème suivant, établi par Dehornoy dans [D94], fournit l'exemple qui nous manquait au n° 2.5 pour construire un ordre total sur les groupes de tresses.

Théorème 3.— *Le système autodistributif D_1 est acyclique et la relation \prec le munit d'un ordre total.*

Le système libre D_1 est un ensemble de classes d'équivalence de parenthésages pour la relation \equiv_{AD} définie dans la preuve de la proposition 4. Avant de démontrer le théorème 3, établissons quelques propriétés de ces parenthésages.

Pour tout entier $n \geq 1$ définissons les puissances droites $x^{[k]}$ de x par $x^{[1]} = x$ et $x^{[k]} = x * x^{[k-1]}$ si $k > 1$. La relation d'équivalence \equiv_{AD} vérifie la propriété fondamentale d'absorption suivante.

Lemme 4.— *Pour tout parenthésage t , on a $x^{[k+1]} \equiv_{AD} t * x^{[k]}$ pour tout k suffisamment grand.*

DÉMONSTRATION.— On procède par récurrence sur la longueur de t . Si $t = x$, alors $x^{[k+1]} = t * x^{[k]}$ par définition des puissances droites. Si $t = t_1 * t_2$, et si

$x^{[k+1]} \equiv_{\text{AD}} t_1 * x^{[k]}$ et $x^{[k+1]} \equiv_{\text{AD}} t_2 * x^{[k]}$ pour tout k suffisamment grand, alors

$$\begin{aligned} x^{[k+2]} &\equiv_{\text{AD}} t_1 * x^{[k+1]} \equiv_{\text{AD}} t_1 * (t_2 * x^{[k]}) \\ &\equiv_{\text{AD}} (t_1 * t_2) * (t_1 * x^{[k]}) \equiv_{\text{AD}} t * x^{[k+1]}. \end{aligned}$$

La troisième équivalence est une conséquence de la définition de \equiv_{AD} . ✓

Nous dirons qu'un parenthésage t' est une *expansion* d'un parenthésage t si on l'obtient en remplaçant à partir de t des sous-expressions de la forme $t_1 * (t_2 * t_3)$ par $(t_1 * t_2) * (t_1 * t_3)$, c'est-à-dire en appliquant l'identité (AD), exclusivement de la gauche vers la droite. Il est clair que $t' \equiv_{\text{AD}} t$ si t' est une expansion de t . On a la réciproque suivante.

Lemme 5.— *Si t_1 et t_2 sont des parenthésages tels que $t_1 \equiv_{\text{AD}} t_2$, alors ils ont une expansion commune.*

Définissons maintenant les sous-termes gauches $G_k(t)$ d'un parenthésage t . Si $t = x$, seul $G_0(t) = t$ est défini. En général, posons $G_0(t) = t$ et, si $t = t_1 * t_2$ et $k \geq 1$, posons $G_k(t) = G_{k-1}(t_1)$ si ce dernier est défini. On a $G_\ell(G_k(t)) = G_{k+\ell}(t)$ lorsque ces expressions sont définies. Notons que l'élément de D_1 représenté par un sous-terme gauche $G_k(t)$, où $k \geq 1$, est plus petit pour l'ordre \prec que l'élément représenté par t .

Lemme 6.— *Si t' est une expansion de t , alors, pour tout k tel que $G_k(t)$ soit défini, il existe un entier $k' \geq k$ tel que $G_{k'}(t')$ soit une expansion de $G_k(t)$.*

3.2. Démonstration du théorème 3. Si le système libre D_1 n'était pas acyclique, il en serait de même de tout système AD, ce qui contredirait la proposition 1.

Montrons maintenant que deux éléments quelconques de D_1 sont comparables pour l'ordre \prec . Plus précisément, soit y et z deux éléments de D_1 que nous représentons respectivement par des parenthésages t_1 et t_2 . Il s'agit d'établir que soit $y = z$, soit $y \prec z$, soit $z \prec y$. D'après le lemme 4, il existe un entier $k \geq 1$ tel que $t_1 * x^{[k]} \equiv_{\text{AD}} t_2 * x^{[k]}$. Le lemme 5 nous permet d'affirmer que les parenthésages $t_1 * x^{[k]}$ et $t_2 * x^{[k]}$ ont une expansion commune t' . Le lemme 6 implique l'existence d'entiers naturels p et q tels que $G_p(t')$ soit une expansion de $G_1(t_1 * x^{[k]}) = t_1$ et que $G_q(t')$ soit une expansion de $G_1(t_2 * x^{[k]}) = t_2$. Les parenthésages $G_p(t')$ et $G_q(t')$ représentent respectivement les éléments y et z de D_1 . Trois cas se présentent à nous :

- (i) si $p = q$, alors $G_p(t') = G_q(t')$, et donc $y = z$;
- (ii) si $p > q$, alors $G_p(t') = G_{p-q}(G_q(t'))$, et donc $y \prec z$.
- (iii) si $p < q$, alors $G_q(t') = G_{q-p}(G_p(t'))$, et donc $z \prec y$. ✓

3.3. Une loi autodistributive sur les tresses. Pour achever la démonstration de la proposition 3 du n° 2.5, nous allons introduire ce que Dehornoy appelle des tresses spéciales. Dans ce but, considérons la limite inductive B_∞ des groupes de tresses B_n pour les inclusions évidentes ; une présentation du groupe B_∞ est obtenue en prenant une famille infinie de générateurs σ_i indexée par les entiers ≥ 1 et les relations (1.1) et (1.2). Soit T l'endomorphisme injectif de B_∞ défini par $T(\sigma_i) = \sigma_{i+1}$ pour tout $i \geq 1$. Posons

$$(3.1) \quad \alpha * \beta = \alpha T(\beta) \sigma_1 T(\alpha^{-1}).$$

Proposition 5.— *La formule (3.1) munit B_∞ d'une structure de système AD simplifiable et acyclique.*

DÉMONSTRATION.— Nous laissons au lecteur le soin de vérifier que B_∞ est un système AD simplifiable. Pour établir son acyclicité, on observe que l'homomorphisme de groupes $B_\infty \rightarrow \text{Aut}(F_\infty)$ qui est la limite inductive des homomorphismes $B_n \rightarrow \text{Aut}(F_n)$ mentionnés au n° 2.2 (a) transporte la loi (3.1) sur la loi (2.2). Comme le système $(\text{Aut}(F_\infty), *)$ est acyclique, il en est de même de $(B_\infty, *)$. \sphericalangle

Dans le système autodistributif B_∞ , considérons le sous-système B_{sp} engendré par la tresse unité 1. Les éléments de B_{sp} sont appelés *tresses spéciales*. Exemples de tresses spéciales : $1*1 = \sigma_1$, $1*(1*1) = \sigma_2\sigma_1$, $(1*1)*1 = \sigma_1^2\sigma_2^{-1}$, $1*(1*(1*1)) = \sigma_3\sigma_2\sigma_1$.

Proposition 6.— *Le système autodistributif B_{sp} est acyclique et la relation \prec le munit d'un ordre total.*

DÉMONSTRATION.— Le système B_{sp} hérite l'acyclicité de B_∞ . La relation \prec est donc une relation d'ordre sur B_{sp} . Il reste à montrer que deux éléments quelconques de B_{sp} sont comparables pour \prec . Comme B_{sp} est engendré par le singleton $\{1\}$, c'est un quotient du système libre D_1 . Par le théorème 3, deux éléments quelconques de D_1 sont comparables pour \prec , donc, par projection, il en est de même dans B_{sp} . \sphericalangle

L'argument précédent montre que, dans tout système AD acyclique et monogène, la relation \prec est une relation d'ordre total. En fait, comme l'a montré Laver [L92], tout système AD acyclique monogène est isomorphe à D_1 . En particulier, $B_{\text{sp}} \cong D_1$.

L'intérêt des tresses spéciales vient de la possibilité d'exprimer toute tresse en termes de produit de tresses spéciales.

Lemme 7.— *Pour tout $\beta \in B_n$, il existe $2n$ tresses spéciales $\alpha_1, \dots, \alpha_n, \gamma_1, \dots, \gamma_n$ telles que*

$$(3.2) \quad \beta = T^{n-1}(\alpha_n)^{-1} \dots T(\alpha_2)^{-1} \alpha_1^{-1} \gamma_1 T(\gamma_2) \dots T^{n-1}(\gamma_n).$$

DÉMONSTRATION.— Comme B_{sp} est un système AD simplifiable, nous pouvons appliquer la proposition 2, qui fournit une action partielle de B_n sur B_{sp}^n . En particulier, pour toute tresse $\beta \in B_n$, il existe des n -uples $(\alpha_1, \dots, \alpha_n)$ et $(\gamma_1, \dots, \gamma_n) \in B_{\text{sp}}^n$ tels que $(\alpha_1, \dots, \alpha_n) \beta = (\gamma_1, \dots, \gamma_n)$. On vérifie alors facilement l'égalité des produits

$$\alpha_1 T(\alpha_2) \dots T^{n-1}(\alpha_n) \beta = \gamma_1 T(\gamma_2) \dots T^{n-1}(\gamma_n)$$

dans le groupe B_∞ , d'où la formule (3.2). /

3.4. Fin de la preuve de la proposition 3. Supposons que α et γ soient deux tresses spéciales. On a ou bien $\alpha \prec \gamma$, ou bien $\alpha = \gamma$, ou bien $\gamma \prec \alpha$. Revenant à la définition de la relation \prec et de la loi (3.1), on voit que les trois relations précédentes entraînent respectivement que la tresse $\alpha^{-1}\gamma$ est σ_1 -positive, égale à 1, ou σ_1 -négative. Si maintenant β est une tresse quelconque, on considère une décomposition de type (3.2) de β : si i est le plus petit indice tel que $\alpha_i \neq \gamma_i$, alors β est σ_i -positive ou σ_i -négative suivant que $\alpha_i \prec \gamma_i$ ou $\gamma_i \prec \alpha_i$, ce qui achève la preuve de la proposition 3. /

4. RAPPORT AVEC LA THÉORIE DES ENSEMBLES

Dans ce numéro nous expliquons comment des développements en théorie des ensembles ont mené à la construction d'un ordre total sur les tresses.

4.1. Grands cardinaux et extensions de ZF. On sait depuis le théorème d'incomplétude de Gödel que le système ZF d'axiomes de Zermelo-Fraenkel est incomplet, c'est-à-dire qu'il existe des énoncés qui sont indécidables dans ZF. C'est le cas de l'axiome du choix et de l'hypothèse du continu. Dès lors, le but principal de la théorie des ensembles consiste à étudier, voire classifier, les extensions du système ZF dans lesquels les énoncés ouverts deviennent décidables. Gödel a émis l'idée que ces extensions pourraient être classifiées à l'aide d'axiomes de grands cardinaux comme celui qui postule l'existence de \aleph_1 . L'un des premiers axiomes considérés a été celui qui affirme

l'existence d'un cardinal inaccessible.² Non seulement cet axiome est indémontrable dans ZF, mais encore — à la différence de l'axiome du choix ou de l'hypothèse du continu — sa cohérence, c'est-à-dire le fait qu'il ne soit pas contradictoire, ne peut y être établie.

Dans les années 1960 une hiérarchie totalement ordonnée d'axiomes de grands cardinaux est apparue en théorie des ensembles, s'imposant comme les bonnes extensions de ZF (voir la monographie [Kan]). Nous allons nous intéresser au n° 4.2 à un axiome particulier de cette hiérarchie.

4.2. Plongements élémentaires. Un ensemble infini se caractérise par l'existence d'une injection non bijective. Une telle injection ne préserve pas nécessairement toutes les structures de l'ensemble considéré : ainsi l'application $n \mapsto n + 1$ de l'ensemble des entiers naturels dans lui-même préserve l'ordre naturel, mais non l'addition. Pour définir des notions d'infini plus fortes, il est naturel de postuler l'existence d'une injection non bijective d'un ensemble dans lui-même, préservant *toutes* les propriétés définissables par les opérations ensemblistes de base. Appelons *plongement élémentaire* une telle injection et *autosimilaire* un ensemble pour lequel il existe un plongement élémentaire. On démontre assez facilement qu'un ensemble autosimilaire est forcément très grand : son cardinal doit être inaccessible, ce qui entraîne que l'existence d'un tel ensemble ne peut être démontrée dans ZF.

Il est usuel en théorie des ensembles de considérer des ensembles particuliers appelés rangs, qui ont la propriété technique que toute fonction d'un rang R dans lui-même est (essentiellement) un élément de R . Les axiomes affirmant l'existence de plongements élémentaires d'un rang dans un autre ou d'un rang dans lui-même ont été au cœur de la théorie des ensembles dans les années 1980. Un des résultats majeurs obtenus à l'aide de tels axiomes a été la démonstration en 1984 par Martin et Steel de la propriété de détermination pour les ensembles projectifs de réels, un énoncé technique qui, d'une certaine façon, décrit complètement la structure fine de la droite réelle (voir [D89]). L'axiome qui nous intéresse est le suivant.

Axiome A. — *Il existe un rang autosimilaire.*

4.3. Le système autodistributif d'un rang autosimilaire. Notons E_R l'ensemble des plongements élémentaires du rang autosimilaire R dont nous venons de postuler l'existence. L'ensemble E_R est trivialement muni d'une structure de monoïde pour la composition des plongements. Mais, en vertu de l'axiome A et des propriétés

² α est inaccessible si tout ensemble de parties d'un ensemble de cardinalité $< \alpha$ et toute union d'ensembles de cardinalité $< \alpha$ sont de cardinalité $< \alpha$.

spécifiques des rangs, il est aussi muni d'une autre opération binaire qu'on peut appeler *l'application*. En effet, nous avons dit qu'une fonction définie sur un rang R peut être vue comme un élément de R . Si donc i et j sont des plongements élémentaires de R , i s'appliquant par hypothèse aux éléments de R et j pouvant être vu comme un tel élément, nous pouvons appliquer i à j . On montre que le nouvel objet $i(j)$ ainsi obtenu, héritant de toutes les propriétés définissables de j , est aussi un plongement élémentaire de R . La loi de composition $(i, j) \mapsto i(j)$ ainsi définie sur E_R est autodistributive. L'argument est le suivant : si y est l'image de x par l'application f , alors $i(y)$ est l'image de $i(x)$ par $i(f)$ chaque fois que i est un plongement élémentaire ; en d'autres termes, on a $i(f(x)) = i(f)(i(x))$, identité valide, en particulier, lorsque f et x sont des plongements élémentaires.

Laver [L92] a établi en 1989 le résultat suivant.

Proposition 7.— *Si j est un plongement élémentaire du rang R , alors le sous-système autodistributif de E_R engendré par j est acyclique.*

Avec ce résultat, Laver a construit le premier exemple d'un système AD acyclique, mais un exemple dont l'existence dépend d'un axiome indémontrable de théorie des ensembles. La situation ainsi créée était étrange : dans des travaux indépendants, Dehornoy et Laver avaient, à cette époque, partiellement résolu le problème des mots de l'identité (AD), en construisant, sous l'hypothèse de l'existence d'un système AD acyclique, un algorithme permettant de reconnaître si deux parenthésages donnés sont ou non équivalents modulo \equiv_{AD} . La proposition 7 prouvait donc la décidabilité du problème des mots en question à partir de l'axiome indémontrable A, établissant un lien paradoxal entre le problème des mots de (AD), qui ne met en jeu que des manipulations de mots finis, et un rang autosimilaire, qui est un objet gigantesque.

Une solution a été proposée vers la fin 1991 par Dehornoy qui, à partir d'une étude fine de l'autodistributivité dont nous donnerons une idée au numéro suivant, a montré par une démonstration directe qu'un système AD libre est acyclique, et a déduit de son analyse les liens avec les tresses (voir [D92a], [D92b], [D94]).

4.4. Le groupe de l'autodistributivité. L'étude algébrique des systèmes AD libres, en particulier celle du système D_1 à un générateur, ont fait l'objet de nombreux travaux de Laver et Dehornoy. Nous ne donnerons ici qu'un aperçu de l'approche de Dehornoy et de la façon dont elle mène aux tresses.

Pour étudier D_1 , il est commode de représenter les parenthésages par des arbres binaires avec racine. On représente la variable x par l'arbre réduit à sa racine ; si t_1 et t_2 sont des parenthésages, on représente le parenthésage $t_1 * t_2$ par l'arbre obtenu

en greffant l'arbre associé à t_1 à gauche d'une nouvelle racine, et l'arbre associé à t_2 à droite. Tout sommet d'un arbre binaire a une *adresse* qui est une suite finie de 0 et de 1 : l'adresse de la racine est la suite vide \emptyset , et, en s'éloignant de la racine, on passe de l'adresse d'un sommet à celui de son "successeur" gauche en ajoutant 0, et à celle de son "successeur" droit en ajoutant 1.

Au n° 3.1 nous avons défini la notion d'expansion d'un parenthésage. Cette définition se transpose facilement aux arbres binaires. Nous pouvons décrire plus finement la situation en considérant le cas où l'identité (AD) est appliquée une seule fois, toujours de la gauche vers la droite, et en prenant en compte l'adresse où elle l'est dans l'arbre correspondant. Nous noterons ∇_α l'opération consistant à appliquer (AD) à l'adresse α (l'opération ∇_α n'est pas partout définie sur M). Il est facile de vérifier les relations suivantes :

$$(4.1) \quad \nabla_{\alpha 0 \beta} \nabla_{\alpha 1 \gamma} = \nabla_{\alpha 1 \gamma} \nabla_{\alpha 0 \beta},$$

$$(4.2) \quad \nabla_{\alpha 0 \beta} \nabla_\alpha = \nabla_\alpha \nabla_{\alpha 1 0 \beta} \nabla_{\alpha 0 0 \beta},$$

$$(4.3) \quad \nabla_{\alpha 1 0 \beta} \nabla_\alpha = \nabla_\alpha \nabla_{\alpha 0 1 \beta},$$

$$(4.4) \quad \nabla_{\alpha 1 1 \beta} \nabla_\alpha = \nabla_\alpha \nabla_{\alpha 1 1 \beta},$$

$$(4.5) \quad \nabla_{\alpha 1} \nabla_\alpha \nabla_{\alpha 1} \nabla_{\alpha 0} = \nabla_\alpha \nabla_{\alpha 1} \nabla_\alpha,$$

où α , β et γ sont des adresses quelconques.

Dans [D94] Dehornoy introduit alors le groupe G_{AD} engendré par les générateurs ∇_α (indexés par les adresses α) et les relations (4.1)–(4.5), et montre que ce groupe opère sur le magma libre M de telle sorte que les orbites de l'action sont exactement les classes d'équivalence d'éléments de M pour la relation \equiv_{AD} . Il établit que le système D_1 est acyclique en étudiant le groupe G_{AD} à l'aide d'une méthode de retournement des mots semblable à celle décrite au n° 2.4 pour les tresses, et en construisant un préordre sur G_{AD} . Une étape essentielle consiste à reprendre la démonstration du lemme d'absorption 4, qui se lit, dans le groupe G_{AD} , comme l'existence d'une loi binaire $*$ donnée par la formule

$$(4.6) \quad a * b = a T(b) \nabla_{\emptyset} T(a)^{-1},$$

où T est cette fois-ci l'endomorphisme qui envoie ∇_α sur $\nabla_{1\alpha}$ pour tout α .

Le lien avec les tresses résulte de l'observation suivante : l'application π définie par $\pi(\nabla_\alpha) = 1$ si l'adresse α contient au moins un 0, et $\pi(\nabla_\alpha) = \sigma_{i+1}$ si α est une suite constituée de i fois le chiffre 1, induit un homomorphisme surjectif de groupes de G_{AD} sur le groupe de tresses B_∞ . En outre, π transporte la loi (4.6) de G_{AD} sur la loi autodistributive (3.1) de B_∞ .

Signalons ici la parenté du groupe G_{AD} avec le groupe F de Thompson, qui en est l'exacte contrepartie lorsqu'on remplace l'identité AD par l'associativité. Dans cette correspondance, la relation (4.5) est l'analogue de la relation du pentagone de Mac Lane–Stasheff. On trouvera un exposé de synthèse sur le groupe F dans [CPF].

4.5. Systèmes autodistributifs finis. En étudiant les quotients finis du système AD associé à un plongement élémentaire, Laver [L95] a montré que, pour tout entier $n \geq 1$, il existe une unique structure autodistributive A_n sur l'ensemble fini $\{1, 2, \dots, 2^n\}$ tel que $k * 1 \equiv k + 1$ modulo 2^n (on trouvera plus bas les tables de multiplication de A_1 , A_2 et A_3). Les ensembles A_n forment un système projectif de systèmes AD et jouent un rôle-clé parmi les systèmes autodistributifs. En particulier, Drápal [Dr] a montré que tout système AD monogène fini se construit simplement à partir de systèmes A_n . Chaque ligne de la table de multiplication de A_n est la répétition périodique d'une suite croissante de nombres allant jusqu'à 2^n . De par l'existence du système projectif mentionné ci-dessus, la période p_n de la première ligne de la table de A_n est une suite croissante au sens large. En traduisant une propriété non triviale des ordinaux et des plongements élémentaires, Laver [L95] a démontré la proposition suivante, qui implique notamment que le sous-système AD engendré par 1 dans la limite projective des A_n est libre.

Proposition 8.— *S'il existe un rang autosimilaire, alors la suite $(p_n)_n$ tend vers l'infini.*

Pour l'instant aucune tentative pour démontrer ce résultat finitiste sans faire appel à la théorie des ensembles n'a abouti. On sait seulement que la première valeur de n pour laquelle p_n dépasse 32 est gigantesque [Do], [DJ]. Les systèmes A_n sont des objets d'une très grande complexité combinatoire, et restent encore très mal connus. Une question évidente, mais complètement ouverte, est de se demander s'ils sont

susceptibles d'applications topologiques.

A_1	1 2	A_2	1 2 3 4	A_3	1 2 3 4 5 6 7 8
1	2 2	1	2 4 2 4	1	2 4 6 8 2 4 6 8
2	1 2	2	3 4 3 4	2	3 4 7 8 3 4 7 8
		3	4 4 4 4	3	4 8 4 8 4 8 4 8
		4	1 2 3 4	4	5 6 7 8 5 6 7 8
				5	6 8 6 8 6 8 6 8
				6	7 8 7 8 7 8 7 8
				7	8 8 8 8 8 8 8 8
				8	1 2 3 4 5 6 7 8

Tables de multiplication de A_1 , A_2 et A_3

5. APPROCHES GÉOMÉTRIQUES

Récemment, deux constructions géométriques de l'ordre de Dehornoy ont été proposées par des topologues.

5.1. La définition de Fenn, Greene, Rolfsen, Rourke et Wiest. Notons D_n le disque unité fermé du corps \mathbf{C} des nombres complexes, muni de n points marqués P_1, \dots, P_n intérieurs au disque et situés sur le diamètre réel $\Gamma = [-1, 1]$. Ordonnons les points marqués de sorte que le diamètre Γ soit la juxtaposition des $n + 1$ segments $[P_0, P_1], [P_1, P_2], \dots, [P_{n-1}, P_n], [P_n, P_{n+1}]$ que nous numérotions de 1 à $n + 1$ (pour simplifier, on a posé $P_0 = -1$ et $P_{n+1} = 1$).

On sait (voir [BZ], chap. 10) que le groupe de tresses B_n est isomorphe au groupe des classes d'isotopie des homéomorphismes de D_n qui fixent le bord du disque point par point et permutent les points marqués. Considérons l'image de Γ par un tel homéomorphisme φ : la courbe simple $\varphi(\Gamma)$ sépare D_n en deux composantes connexes et est l'union de $n + 1$ "petites courbes" $\varphi([P_i, P_{i+1}])$ dont les extrémités sont les points P_0, P_1, \dots, P_{n+1} . On appelle $\varphi(\Gamma)$ le *diagramme de courbes* de φ . Suivant un procédé bien connu et utilisé par exemple dans [Mo], on rectifie $\varphi(\Gamma)$ en raccourcissant les petites courbes de manière à faire coïncider le maximum d'entre elles avec des segments $[P_i, P_{i+1}]$ et à éliminer les intersections inutiles avec Γ . Il est montré dans [F+] que deux tels homéomorphismes de D_n représentent la même tresse dans B_n si et seulement leurs diagrammes de courbes rectifiés sont isotopes *via* une isotopie qui

fixe les petites courbes coïncidant avec des segments $[P_i, P_{i+1}]$. Pour $i = 1, \dots, n-1$, on peut donc considérer l'ensemble Π_i des éléments de B_n correspondant aux diagrammes de courbes rectifiés pour lesquels $\varphi([P_{j-1}, P_j]) = [P_{j-1}, P_j]$ pour tout $j < i$ et $\varphi([P_{i-1}, P_i])$ monte vers le demi-plan supérieur de \mathbf{C} . On pose $\Pi = \bigcup_{i=1}^{n-1} \Pi_i$.

Fenn *et al.* [F+] ont établi en 1997 que la relation $\beta <_{\Pi} \beta'$ définie par $\beta^{-1}\beta' \in \Pi$ est une relation d'ordre total invariant par multiplication à gauche sur le groupe B_n . Ils ont aussi montré qu'une tresse est dans Π_i si et seulement si elle est σ_i -positive, ce qui entraîne que l'ordre $<_{\Pi}$ coïncide avec celui défini au n° 1.3. On obtient ainsi une définition géométrique de l'ordre de Dehornoy.

Utilisant les mêmes techniques, Rourke et Wiest [RW] (voir aussi [SW]) ont montré que le groupe des difféotopies ("mapping class group" en anglais) d'une surface compacte (non nécessairement orientable) à bord non vide est un groupe ordonnable. Ils ont aussi construit un algorithme qui décide en temps quadratique si un élément différent de l'identité est positif ou négatif pour l'ordre construit.

La figure 3 montre les diagrammes de courbes de cinq tresses. On voit, par exemple, sur ces diagrammes que $\sigma_1\sigma_2\sigma_1^{-1}$ et $\sigma_1\sigma_2\sigma_1^{-2}$ sont σ_1 -positifs.

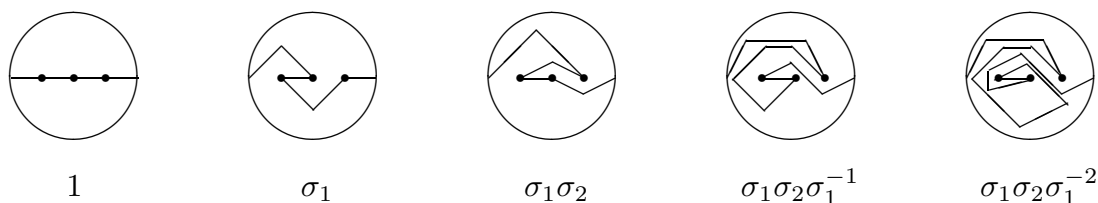


Figure 3

5.2. L'approche à la Nielsen. En septembre 1998, W. P. Thurston [Th] a expliqué comment les travaux classiques de Nielsen sur les surfaces permettaient de construire un ordre invariant sur B_n à partir de la proposition suivante.

Proposition 9.— *Le groupe B_n opère sur la droite réelle par des homéomorphismes croissants de sorte que l'action soit libre en au moins un point réel.*

Avant d'établir la proposition 9, montrons comment on en déduit un ordre total sur B_n . Soit $x \in \mathbf{R}$ un point pour lequel l'action de la proposition 9 est libre (c'est-à-dire un point dont le stabilisateur est trivial). Pour $\beta, \beta' \in B_n$, posons $\beta <_x \beta'$ si $\beta(x) < \beta'(x)$ pour l'ordre naturel de \mathbf{R} . La liberté de l'action en x implique que la relation $<_x$ est un ordre, et c'est alors clairement un ordre total. Comme B_n opère par des homéomorphismes croissants, l'ordre $<_x$ est invariant par multiplication à gauche.

Short et Wiest [SW] ont classifié les ordres $<_x$ à conjugaison près (il y en a une infinité non dénombrable) et montré que l'on retrouve ainsi l'ordre de Dehornoy.

5.3. Démonstration de la proposition 9. Soit S une surface orientée fermée de genre un marquée en n points P_1, \dots, P_n et C une courbe simple séparant S en une surface A de genre un et une surface de genre zéro contenant les points marqués. Le groupe de tresses B_n s'identifie aussi au groupe des classes d'isotopie de difféomorphismes de S préservant l'orientation, induisant l'identité sur A et permutant les points marqués.

Munissons $S - \{P_1, \dots, P_n\}$ d'une métrique hyperbolique complète pour laquelle C est une géodésique et les points marqués sont des "cusps." Fixons un point-base x_0 sur C et le revêtement universel $(D, 0)$ de $(S - \{P_1, \dots, P_n\}, x_0)$. La métrique hyperbolique identifie $(D, 0)$ avec le disque unité de \mathbf{C} . Tout difféomorphisme φ de S qui fixe x_0 et permute les points marqués se relève en un difféomorphisme $\tilde{\varphi}$ de D fixant 0. Dans [Ni], § 10, Nielsen a montré que $\tilde{\varphi}$ se prolonge en un homéomorphisme $\partial\tilde{\varphi}$ du bord ∂D de D , c'est-à-dire du cercle unité, que $\partial\tilde{\varphi}$ conserve l'orientation si φ la conserve, et que $\partial\tilde{\varphi}$ ne dépend que de la classe d'isotopie de φ . On obtient ainsi une action de B_n sur le cercle, préservant l'orientation. Pour obtenir une action sur \mathbf{R} , il suffit de remarquer que l'action précédente fixe un point, par exemple l'une des extrémités d'un relèvement de C passant par 0.

L'existence d'un point de \mathbf{R} pour lequel l'action est libre se déduit également d'un résultat de Nielsen qui dit que, si φ n'est pas isotope à l'identité, alors l'ensemble des points fixes de $\partial\tilde{\varphi}$ est un fermé d'intérieur vide (voir [Ni], § 14). Comme une union dénombrable de fermés d'intérieur vide est d'intérieur vide (théorème de Baire), il résulte que l'action de B_n sur \mathbf{R} est libre sur une partie G_δ -dense, donc en au moins un point. /

Remarque. La lecture de [Ni] montre que Nielsen avait déjà établi l'existence d'une action de B_n sur le cercle, préservant l'ordre cyclique, fixant un point et ayant une orbite libre.

BIBLIOGRAPHIE

- [Ad] S. I. ADYAN - *Fragments of the word Δ in a braid group*, Mat. Zametki **36** (1984), 25–34 (= Math. Notes **36** (1984), 505–510).
- [A26] E. ARTIN - *Theorie der Zöpfe*, Abh. Math. Sem. Univ. Hamburg **4** (1926), 47–72.
- [A47] E. ARTIN - *Theory of braids*, Ann. of Math. **48** (1947), 101–126.

- [Bo] N. BOURBAKI - *Algèbre*, chapitres 1 à 3, Hermann, Paris, 1970.
- [B+] *Braids*, Actes Conf. Santa Cruz (1986) publiés par Joan S. Birman et Anatoly Libgober, Contemp. Math., **78**, Amer. Math. Soc., Providence, RI, 1988.
- [Br] E. BRIESKORN - *Automorphic sets and braids and singularities*, publié dans [B+], pages 45–115.
- [Bu] S. BURCKEL - *The wellordering on positive braids*, J. Pure Appl. Algebra **120** (1997), 1–17.
- [BZ] G. BURDE, H. ZIESCHANG - *Knots*, Studies in Math. **5**, Walter de Gruyter, Berlin-New York, 1985.
- [CFP] J. W. CANNON, W. J. FLOYD, W. R. PARRY - *Introductory notes on Richard Thompson's groups*, Enseign. Math. (2) **42** (1996), 215–256.
- [Ca] P. CARTIER - *Développements récents sur les groupes de tresses ; applications à la topologie et l'algèbre*, Séminaire Bourbaki, exposé **716** (novembre 1989), Astérisque, vol. 189–190, Soc. Math. France, Paris (1991), 17–67.
- [D89] P. DEHORNOY - *La détermination projective (d'après Martin, Steel et Woodin)*, Séminaire Bourbaki, exposé **710** (juin 1989), Astérisque, vol. 177–178, Soc. Math. France, Paris (1989), 261–276.
- [D92a] P. DEHORNOY - *Preuve de la conjecture d'irréflexivité pour les structures distributives libres*, C. R. Acad. Sci. Paris Sér. I Math. **314** (1992), 333–336.
- [D92b] P. DEHORNOY - *Deux propriétés des groupes de tresses*, C. R. Acad. Sci. Paris Sér. I Math. **315** (1992), 633–638.
- [D94] P. DEHORNOY - *Braid groups and left distributive operations*, Trans. Amer. Math. Soc. **345** (1994), 115–150.
- [D95] P. DEHORNOY - *From large cardinals to braids via distributive algebra*, J. Knot Theory Ramifications **4** (1995), 33–79.
- [D96] P. DEHORNOY - *Une autre application de la théorie des ensembles*, Gaz. Math. **69** (1996), 3–20.
- [D97a] P. DEHORNOY - *Groups with a complemented presentation*, J. Pure Appl. Algebra **116** (1997), 115–137.
- [D97b] P. DEHORNOY - *A fast method for comparing braids*, Adv. Math. **125** (1997), 200–235.
- [D98] P. DEHORNOY - *Gaussian groups are torsion free*, J. of Algebra **210** (1998), 291–297.
- [D00] P. DEHORNOY - *Braids and self-distributivity*, disponible sur la Toile à l'URL <http://www.math.unicaen.fr/~dehornoy/book.html>, à paraître dans Progress in Math., Birkhäuser, Basel, Boston.

- [DP] P. DEHORNOY, L. P. PARIS - *Gaussian groups and Garside groups, two generalizations of Artin groups*, Proc. London Math. Soc. **79** (1999), 569–604.
- [Do] R. DOUGHERTY - *Critical points in an algebra of elementary embeddings*, Ann. Pure Appl. Logic **65** (1993), 211–241.
- [DJ] R. DOUGHERTY, T. JECH - *Finite left-distributive algebras and embedding algebras*, Adv. Math. **130** (1997), 201–241.
- [Dr] A. DRÁPAL - *Finite left distributive algebras with one generator*, J. Pure Appl. Algebra **121** (1997), 233–251.
- [EM] E. A. ELRIFAI, H. R. MORTON - *Algorithms for positive braids*, Quart. J. Math. Oxford Ser. (2) **45** (1994), 479–497.
- [E+] D. B. A. EPSTEIN, J. W. CANNON, D. F. HOLT, S. V. F. LEVY, M. S. PATERSON, W. P. THURSTON - *Word processing in groups*, Jones and Bartlett Publ., Boston, Mass., 1992.
- [F+] R. FENN, M. T. GREENE, D. ROLFSEN, C. ROURKE, B. WIEST - *Ordering the braid groups*, Pacific J. Math. **191** (1999), 49–74..
- [FR] R. FENN, C. ROURKE - *Racks and links in codimension two*, J. Knot Theory Ramifications **1** (1992), 343–406.
- [Ga] F. A. GARSIDE - *The braid group and other groups*, Quart. J. Math. Oxford **20** (1969), 235–254.
- [Joy] D. JOYCE - *A classifying invariant of knots, the knot quandle*, J. Pure Appl. Algebra **23** (1982), 37–65.
- [Kan] A. KANAMORI - *The higher infinite. Large cardinals in set theory from their beginnings*. Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1994.
- [Kau] L. H. KAUFFMAN - *Knots and physics*, Series on Knots and Everything, vol. 1, World Scientific Publishing Co., Inc., River Edge, NJ, 1991.
- [KR] D. M. KIM, D. ROLFSEN - *Ordering groups of pure braids and hyperplane arrangements*, prépublication Univ. British Columbia, 1998.
- [La] D. M. LARUE - *On braid words and irreflexivity*, Algebra Universalis **31** (1994), 104–112.
- [L92] R. LAVER - *The left distributive law and the freeness of an algebra of elementary embeddings*, Adv. Math. **91** (1992), 209–231.
- [L95] R. LAVER - *On the algebra of elementary embeddings of a rank into itself*, Adv. Math. **110** (1995), 334–346.
- [L96] R. LAVER - *Braid group actions on left distributive structures, and well orderings in the braid groups*, J. Pure Appl. Algebra **108** (1996), 81–98.

- [Mk] G. S. MAKANIN - *The conjugacy problem in the braid group*, Dokl. Akad. Nauk SSSR **182** (1968), 495–496 (= Soviet Math. Dokl. **9** (1968), 1156–1157).
- [Mt] S. V. MATVEEV - *Distributive groupoids in knot theory*, Mat. Sb. (N.S.) **119** (**161**) (1982), 78–88 (= Math. USSR Sbornik **47** (1984), 73–83).
- [Mo] L. MOSHER - *Mapping class groups are automatic*, Ann. of Math. (2) **142** (1995), 303–384.
- [MR] R. BOTTO MURA, A. H. RHEMTULLA - *Orderable groups*, Lecture Notes in Pure and Applied Mathematics, vol. **27**, Marcel Dekker, Inc., New York, Basel, 1977.
- [Ni] J. NIELSEN, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Math. **50** (1927), 189–358 ; traduit en anglais par John Stillwell dans Jakob Nielsen, *Collected mathematical papers*, Birkhäuser, Boston, Basel, Stuttgart, 1986.
- [Pa] D. S. PASSMAN - *The algebraic structure of group rings*, Pure and Applied Mathematics, Wiley-Interscience, New York, London, Sydney, 1977.
- [RZ] D. ROLFSEN, J. ZHU - *Braids, orderings and zero divisors*, J. Knot Theory Ramifications **7** (1998), 837–841.
- [RW] C. ROURKE, B. WIEST - *Order automatic mapping class groups*, prépublication University of Warwick 1998, à paraître dans Pacific J. Math.
- [SW] H. SHORT, B. WIEST - *Orderings of mapping class groups after Thurston*, prépublication CMI, Université de Provence 1999, math.GT/9907104.
- [Th] W. P. THURSTON - courrier électronique à Bertold Wiest, 4 septembre 1998.
- [W] B. WIEST - *Dehornoy's ordering of the braid groups extends the subword ordering*, Pacific J. Math. **191** (1999), 183–188.

Christian KASSEL

Institut de Recherche Mathématique Avancée
Université Louis Pasteur - C.N.R.S.

7 rue René Descartes

67084 Strasbourg Cedex, France

E-mail : kassel@math.u-strasbg.fr

Fax : +33 (0)3 88 61 90 69

<http://www-irma.u-strasbg.fr/~kassel/>

**EXPOSANTS CRITIQUES POUR LE MOUVEMENT BROWNIEN
ET LES MARCHES ALÉATOIRES**
[d'après Kenyon, Lawler et Werner]

par **Jean-François LE GALL**

INTRODUCTION

Cet exposé présente plusieurs travaux récents qui étudient certains exposants asymptotiques pour le mouvement brownien ou les marches aléatoires dans le plan. La première partie, issue des articles de Lawler et Werner, est consacrée aux exposants d'intersection, qui donnent l'ordre de grandeur de la probabilité pour deux ou plusieurs trajectoires de mouvements browniens, ou de marches aléatoires, de ne pas se rencontrer sur un intervalle de temps long. Si l'existence de ces exposants et parfois leurs valeurs exactes ont été prédites depuis longtemps par les physiciens théoriciens, leur construction rigoureuse ne date que de quelques années, et c'est seulement avec les travaux récents de Lawler et Werner que leur étude mathématique a connu des progrès vraiment significatifs. Un pas décisif a été l'obtention de relations "en cascade" entre les différents exposants (Théorème 1.3). Ces relations permettent ensuite de montrer que la classe importante des exposants dans un demi-espace peut s'écrire en termes d'une seule fonction réelle U pour laquelle on dispose d'une conjecture très plausible. La fonction U possède une propriété d'universalité remarquable puisqu'elle apparaît dès que l'on cherche à définir des exposants asymptotiques pour des mesures définies sur une classe de sous-ensembles compacts du plan et vérifiant une propriété convenable d'invariance conforme (Théorème 1.7). Après la soumission de la première version de cet exposé, nous avons reçu un travail tout récent de Lawler, Schramm et Werner [26] qui résout certaines des conjectures relatives aux exposants d'intersection (une brève discussion de ces derniers résultats est donnée à la fin de la première partie).

La deuxième partie a pour objectif principal de présenter un résultat de Kenyon (Théorème 2.1) donnant la valeur exacte de l'exposant de croissance de la marche aléatoire à boucles effacées dans le plan. La marche aléatoire à boucles effacées est un chemin aléatoire auto-évitant (i.e. sans auto-intersection) et est à ce titre un objet d'intérêt pour les physiciens théoriciens qui étudient les modèles de polymères. Le calcul de l'exposant de croissance applique des méthodes que Kenyon a développées pour l'étude asymptotique du nombre de pavages par dominos de certains sous-ensembles du plan, lorsque le pas du

réseau devient petit. Un résultat significatif de Kenyon dans cette direction figure aussi à la fin de cet exposé (Théorème 2.7). Ce résultat est susceptible de plusieurs interprétations, puisque les pavages par dominos des régions considérées sont en bijection avec les arbres couvrants du graphe associé, et qu'on sait depuis Kirchhoff que le nombre d'arbres couvrants est égal au produit des valeurs propres non nulles du Laplacien discret sur le graphe.

Les deux parties sont indépendantes, et les techniques impliquées sont très différentes. Cependant il existe des liens étroits entre les objets étudiés : l'article de Lawler et Werner qui fournit l'essentiel de la matière de la première partie donne aussi des conjectures précises sur le comportement asymptotique de la marche aléatoire à boucles effacées. En outre, les propriétés d'invariance conforme, pour les trajectoires browniennes dans la première partie ou la fonction de couplage asymptotique dans la seconde, jouent un rôle fondamental dans tout l'exposé.

1. EXPOSANTS D'INTERSECTION DU MOUVEMENT BROWNIEN

1.1. La définition des exposants

Soient des entiers $k \geq 2$ et $n_1, \dots, n_k \geq 1$ et soit une famille $(B_j^i; 1 \leq i \leq k, 1 \leq j \leq n_i)$ de $n_1 + \dots + n_k$ mouvements browniens plans indépendants, dont le point de départ est uniformément distribué sur le cercle unité. Pour tout $R > 1$, soit

$$T_j^i(R) = \inf\{t \geq 0 : |B_j^i(t)| = R\}$$

le temps de sortie du mouvement brownien B_j^i hors du disque de rayon R centré en l'origine. Pour $a \geq 0$ on note $B_j^i[0, a] = \{B_j^i(t); 0 \leq t \leq a\}$ l'ensemble des points visités par B_j^i sur l'intervalle de temps $[0, a]$.

PROPOSITION 1.1. — *Il existe un réel $\zeta(n_1, \dots, n_k) > 0$, tel que, pour $R \rightarrow \infty$,*

$$(1) \quad P\left[\left(\bigcup_{j=1}^{n_i} B_j^i[0, T_j^i(R)]\right) \cap \left(\bigcup_{j=1}^{n_{i'}} B_j^{i'}[0, T_j^{i'}(R)]\right) = \emptyset, \text{ pour tous } i \neq i'\right] \approx R^{-\zeta(n_1, \dots, n_k)}.$$

La notation $f(R) \approx g(R)$ signifie que $\frac{\log f(R)}{\log g(R)} \rightarrow 1$ quand $R \rightarrow \infty$.

Le nombre $\zeta(n_1, \dots, n_k)$ est appelé *exposant d'intersection* pour k paquets de respectivement n_1, \dots, n_k mouvements browniens plans. La première construction mathématique de ces exposants est due à Burdzy et Lawler [2] (voir aussi [6]).

Esquisons la preuve de la proposition. Soient z_j^i , $1 \leq i \leq k, 1 \leq j \leq n_i$ des points du cercle unité. Ecrivons $P^{(z_j^i)}$ pour la probabilité sous laquelle les processus B_j^i sont des mouvements browniens plans indépendants et B_j^i part de z_j^i . Soit \mathcal{A}_R l'événement

$$\mathcal{A}_R = \left\{ \left(\bigcup_{j=1}^{n_i} B_j^i[0, T_j^i(R)] \right) \cap \left(\bigcup_{j=1}^{n_{i'}} B_j^{i'}[0, T_j^{i'}(R)] \right) = \emptyset, \text{ pour tous } i \neq i' \right\}.$$

Posons

$$f(R) = \sup_{(z_j^i)} P^{(z_j^i)}(\mathcal{A}_R),$$

où le supremum porte sur tous les choix possibles de familles (z_j^i) de points sur le cercle unité. Une application de la propriété de Markov forte du mouvement brownien, puis de la propriété d'invariance par changement d'échelle, montre que, pour tous $R, R' > 1$,

$$f(RR') \leq f(R)f(R').$$

Donc $\log f$ est fonction sous-additive de $\log R$, et la limite

$$\lim_{R \rightarrow \infty} \frac{\log f(R)}{\log R} = -\zeta(n_1, \dots, n_k)$$

existe. Des estimations simples montrent que $0 < \zeta(n_1, \dots, n_k) < \infty$ (pour voir que $\zeta(n_1, \dots, n_k)$ est fini, on observe que \mathcal{A}_R est réalisé dès que les k paquets de mouvements browniens sont contenus dans des cônes disjoints basés en l'origine).

Il est ensuite facile d'obtenir l'énoncé de la proposition. La quantité $P(\mathcal{A}_R)$ qui figure à gauche dans (1) est évidemment majorée par $f(R)$. Inversement, si on observe que la loi du point d'atteinte par B_j^i du cercle de rayon 2 centré en 0 est majorée par une constante fois la probabilité uniforme sur ce cercle, on obtient aisément l'estimation $f(2R) \leq C P(\mathcal{A}_R)$, C étant une constante indépendante de R .

REMARQUE 1.2. — Modulo un certain travail technique, on peut remplacer les temps de sortie de grands disques centrés en l'origine par des instants déterministes qu'on fait tendre vers l'infini. Par exemple, on peut définir l'exposant $\zeta(1, 1)$ en disant que si B et B' sont deux mouvements browniens plans indépendants issus de points distincts, on a

$$P[B[0, t] \cap B'[0, t] = \emptyset] \approx t^{-\zeta(1,1)/2}$$

quand $t \rightarrow \infty$. On peut aussi [2] remplacer les mouvements browniens par des marches aléatoires : si S et S' sont deux marches aléatoires simples (i.e. à plus proches voisins) indépendantes dans \mathbb{Z}^2 , issues de points distincts, on a

$$P[\{S_k; 0 \leq k \leq n\} \cap \{S'_k; 0 \leq k \leq n\} = \emptyset] \approx n^{-\zeta(1,1)/2}$$

quand $n \rightarrow \infty$. La définition des exposants qui résulte de la Proposition 1.1 est cependant la plus facile à justifier.

Dans toute la classe des exposants $\zeta(n_1, \dots, n_k)$ un seul a pu être calculé exactement: $\zeta(2, 1) = 2$. Cependant, Lawler et Werner [24] ont montré qu'il existe des relations remarquables et inattendues entre ces exposants. L'objectif principal des paragraphes qui suivent est d'expliquer la démarche qui conduit à ces relations. Parallèlement, nous donnerons aussi certains résultats qui justifient l'importance du calcul des exposants.

Un rôle crucial dans l'approche de [24] est joué par une seconde classe d'exposants, les exposants d'intersection dans un demi-plan. Soit $H = \{z : \text{Im}z > 0\}$ le demi-plan supérieur. Des arguments très semblables à ceux de la preuve précédente montrent que

$$P[\mathcal{A}_R \cap \{B_j^i[0, T_j^i(R)] \subset H \text{ pour tous } i, j\}] \approx R^{-\xi(n_1, \dots, n_k)}$$

avec un exposant $\xi(n_1, \dots, n_k) \geq \zeta(n_1, \dots, n_k)$. Remarquons que cela a un sens de définir l'exposant ξ lorsque $k = 1$ (alors \mathcal{A}_R n'intervient pas), et des estimations faciles montrent que $\xi(n_1) = n_1$.

Bien qu'il n'y ait jusqu'à présent pas de calcul mathématique rigoureux des exposants, les physiciens théoriciens ont proposé des valeurs exactes pour certains d'entre eux. Sur la base de simulations et de considérations non-rigoureuses de théorie des champs conforme, Duplantier et Kwon [11] (voir aussi [12]) ont conjecturé que

$$\xi(1^{\otimes k}) = \frac{k(2k+1)}{3}, \quad \zeta(1^{\otimes k}) = \frac{4k^2-1}{12},$$

où $\alpha^{\otimes k}$ désigne le k -uplet dont toutes les composantes sont égales à α . Bien que la valeur donnée ci-dessus pour ζ lorsque $k = 1$ ($\zeta(1) = \frac{1}{4}$) n'ait a priori pas de sens, Duplantier a suggéré que cette valeur correspond à l'exposant de disconnexion que nous introduirons plus loin. Après avoir eu connaissance du travail de Lawler et Werner [24], Duplantier [8],[9] a justifié et étendu les conjectures précédentes dans le cadre (non rigoureux pour les mathématiciens) de la théorie de la gravité quantique.

1.2. L'extension aux valeurs réelles des paramètres

Dans ce paragraphe, nous montrons comment on étend la définition des exposants ζ ou ξ au cas où certains des paramètres n_1, \dots, n_k prennent des valeurs réelles. Par souci de simplicité, nous définissons d'abord $\zeta(1, \lambda)$ pour tout $\lambda \geq 0$. Nous considérons deux mouvements browniens plans B et W partant comme ci-dessus d'un point uniformément réparti sur le cercle unité, définis respectivement sous les probabilités P et P' . Notons T_R , resp. S_R , pour le temps d'atteinte par B , resp. W , du cercle de rayon R . Par définition,

$$P \otimes P'[B[0, T_R] \cap W[0, S_R] = \emptyset] \approx R^{-\zeta(1,1)}.$$

En intégrant d'abord par rapport à P' , on peut aussi écrire

$$P \otimes P'[B[0, T_R] \cap W[0, S_R] = \emptyset] = E[Z_R(B)]$$

où $Z_R(B) = P'[B[0, T_R] \cap W[0, S_R] = \emptyset]$ est une fonction du mouvement brownien B (en fait de l'ensemble aléatoire $B[0, T_R]$). Si W est un autre mouvement brownien plan défini sous P' (indépendant de W et avec la même loi de départ), le même argument montre aussi que

$$E[Z_R(B)^2] = P \otimes P'[B[0, T_R] \cap (W[0, S_R] \cup \tilde{W}[0, \tilde{S}_R]) = \emptyset] \approx R^{-\zeta(1,2)}$$

et plus généralement, pour tout entier $n \geq 1$,

$$E[Z_R(B)^n] \approx R^{-\zeta(1,n)}.$$

Ceci suggère de définir pour tout réel $\lambda \geq 0$ l'exposant $\zeta(1, \lambda)$ par :

$$E[Z_R(B)^\lambda] \approx R^{-\zeta(1,\lambda)},$$

avec la convention inhabituelle que lorsque $\lambda = 0$, $0^0 = 0$. Un argument de sous-additivité montre à nouveau qu'il existe un tel exposant $\zeta(1, \lambda)$.

La même idée permet pour tout entier $k \geq 1$, pour tous les entiers $n_1, \dots, n_k \geq 1$ et tous les réels $\lambda_1, \dots, \lambda_k \geq 0$ de définir un exposant $\zeta(n_1, \lambda_1, \dots, n_k, \lambda_k)$ qui étend la définition précédente. Remarquons qu'il est (pour l'instant) nécessaire d'alterner entre valeurs réelles et entières des paramètres. En effet, on conditionne d'abord par rapport aux k paquets de respectivement n_1, \dots, n_k mouvements browniens, puis on considère k variables Z_R^1, \dots, Z_R^k qui sont (toutes) non nulles seulement sur l'ensemble où les k paquets sont d'intersection vide deux à deux, et de plus sont ordonnés dans le sens trigonométrique autour de l'origine. La variable Z_R^i est alors la probabilité conditionnelle qu'un mouvement brownien indépendant reste "entre le i -ième et le $i + 1$ -ième paquet" jusqu'au temps d'atteinte du rayon R . Finalement, l'exposant $\zeta(n_1, \lambda_1, \dots, n_k, \lambda_k)$ est défini par

$$E[(Z_R^1)^{\lambda_1} \dots (Z_R^k)^{\lambda_k}] \approx R^{-\zeta(n_1, \lambda_1, \dots, n_k, \lambda_k)}.$$

Une extension analogue vaut pour le cas des exposants ξ dans un demi-espace. En fait pour ces exposants il est naturel de définir $\xi(\lambda_0, n_1, \lambda_1, \dots, n_k, \lambda_k)$ pour n_1, \dots, n_k entiers (non-nuls) et $\lambda_0, \lambda_1, \dots, \lambda_k$ réels positifs. La méthode est exactement celle décrite ci-dessus, mais on impose aux paquets de mouvements browniens de rester dans le demi-plan supérieur H et en plus de Z_R^1, \dots, Z_R^k , on introduit aussi Z_R^0 qui est la probabilité pour un mouvement brownien indépendant de rester (jusqu'au temps d'atteinte du rayon R) entre la demi-droite réelle positive et le premier paquet de n_1 mouvements browniens. Par convention on prend $\xi(\lambda) = \lambda$ pour tout $\lambda \geq 0$, ce qui est cohérent avec le cas où λ est entier.

On notera dans la suite $\xi(1, \lambda) = \xi(0, 1, \lambda)$.

1.3. Propriétés trajectorielles du mouvement brownien

Avant d'aller plus loin, expliquons pourquoi le calcul des exposants est important en liaison avec l'étude de la trajectoire brownienne. Remarquons d'abord que les exposants $\zeta(p, 0)$ pour p entier strictement positif, ont une interprétation remarquable. En effet, avec les notations ci-dessus, on a

$$P[Z_R(B) > 0] \approx R^{-\zeta(1,0)}.$$

Or l'événement $\mathcal{B}_R = \{Z_R(B) > 0\}$ est réalisé ssi la trajectoire $B[0, T_R]$ ne disconnecte pas le cercle unité de l'infini. On appelle donc $\zeta(1, 0)$ l'exposant de disconnexion pour une

trajectoire brownienne. On a une interprétation analogue pour $\zeta(p, 0)$, qui est l'exposant de disconnexion pour p trajectoires browniennes indépendantes. Lawler a montré que ces exposants sont liés à une conjecture fameuse de Mandelbrot [28] sur la dimension de Hausdorff du contour de la courbe brownienne. Le contour, noté \mathcal{C} , de $B[0, 1]$ est par définition la frontière de la composante connexe non bornée de $c \setminus B[0, 1]$. La conjecture de Mandelbrot est $\dim \mathcal{C} = \frac{4}{3}$, et Lawler [22] a montré que

$$\dim \mathcal{C} = 2 - \zeta(2, 0)$$

(un premier résultat dans cette direction avait été obtenu dans [3]). L'idée de la preuve est relativement simple : si $z = B_t$, avec $t \in]0, 1[$, est un point de \mathcal{C} , la courbe de B considérée avant et après l'instant t fournit localement deux trajectoires browniennes indépendantes issues de z qui ne disconnectent pas le point z de l'infini.

Un second lien entre exposants d'intersection et propriétés trajectoires de la courbe brownienne concerne les points de coupure. Un point $z \in \mathcal{C}$ est appelé point de coupure de B si $z = B_s$ pour une valeur de $s \in]0, 1[$ telle que $B[0, s] \cap B[s, 1] = \{B_s\}$. Si \mathcal{D} désigne l'ensemble des points de coupure, Lawler [21] a montré que

$$\dim \mathcal{D} = 2 - \zeta(1, 1).$$

Voir encore [4] pour d'autres applications des exposants d'intersection.

1.4. Les relations en cascade

THÉORÈME 1.3 ([24]). — *Soient $n_1, \dots, n_k \geq 1$ des entiers et $\lambda_0, \lambda_1, \dots, \lambda_k$ des réels positifs. Alors, pour tout entier $j \in \{1, \dots, k-1\}$,*

$$\xi(\lambda_0, n_1, \lambda_1, \dots, n_k, \lambda_k) = \xi(\lambda_0, n_1, \lambda_1, \dots, n_j, \xi(\lambda_j, n_{j+1}, \lambda_{j+1}, \dots, n_k, \lambda_k))$$

et

$$\zeta(n_1, \lambda_1, \dots, n_k, \lambda_k) = \zeta(n_1, \lambda_1, \dots, n_j, \xi(\lambda_j, n_{j+1}, \lambda_{j+1}, \dots, n_k, \lambda_k)).$$

REMARQUE 1.4. — La première relation ne met en jeu que les exposants ξ , alors que la seconde mélange ξ et ζ . D'une certaine manière, les exposants dans un demi-espace sont plus fondamentaux.

PREUVE (esquisse) — Nous allons traiter en détail un cas particulier du théorème, à savoir la relation $\xi(1, 1, 1) = \xi(1, \xi(1, 1))$. Les mêmes arguments montreraient aussi bien l'égalité $\zeta(1, 1, 1) = \zeta(1, \xi(1, 1))$, et la preuve de la forme générale des relations en cascade repose sur des idées très semblables.

Nous considérons trois mouvements browniens plans indépendants W, B, B' issus respectivement de y, z_1, z_2 avec $y = e^{i\gamma}, z_1 = e^{i\theta_1}, z_2 = e^{i\theta_2}$, où $0 < \gamma < \theta_1 < \theta_2 < \pi$. Par commodité, nous supposons que W, B, B' sont en fait des mouvements browniens conditionnés à atteindre le cercle de rayon R avant de pénétrer dans le disque unité ouvert (bien qu'il s'agisse d'un conditionnement par un ensemble de probabilité nulle, on

vérifie assez facilement que les exposants peuvent être définis aussi bien en termes de ces processus conditionnés). On a alors, avec des notations évidentes

$$P[W[0, S_R] \cap B[0, T_R] = \emptyset, B[0, T_R] \cap B'[0, T'_R] = \emptyset, \\ W[0, S_R] \subset H, B[0, T_R] \subset H, B'[0, T'_R] \subset H] \approx R^{-\xi(1,1,1)}.$$

Observer que l'événement $\{W[0, S_R] \cap B'[0, T'_R] = \emptyset\}$ est automatiquement réalisé sur l'ensemble que nous considérons.

Nous allons maintenant conditionner par rapport à la trajectoire $W[0, S_R]$ (en supposant que $W[0, S_R] \subset H$). Notons $H_R = \{z \in H : 1 < |z| < R\}$ et U_R la composante connexe de $H_R \setminus W[0, S_R]$ dont la frontière contient z_1 et z_2 . Alors,

$$P[W[0, S_R] \cap B[0, T_R] = \emptyset, B[0, T_R] \cap B'[0, T'_R] = \emptyset, \\ W[0, S_R] \subset H, B[0, T_R] \subset H, B'[0, T'_R] \subset H] \\ = E[1_{\{W[0, S_R] \subset H\}} P[B[0, T_R] \cap B'[0, T'_R] = \emptyset, B[0, T_R] \subset U_R, B'[0, T'_R] \subset U_R \mid W]].$$

On peut trouver un réel $L_R > 1$ (dépendant de U_R et donc de W) et une bijection conforme φ de U_R sur H_{L_R} , qui se prolonge par continuité à \bar{U}_R , de telle manière que :

- φ envoie $\partial^1 U_R := \{z \in \partial U_R : |z| = 1\}$ sur $\partial^1 H_{L_R} := \{z \in \partial H_{L_R} : |z| = 1\}$;
- φ envoie $\partial^2 U_R := \{z \in \partial U_R : |z| = R\}$ sur $\partial^2 H_{L_R} := \{z \in \partial H_{L_R} : |z| = L_R\}$.

Le nombre L_R s'interprète en termes de la distance extrémale entre $\partial^1 U_R$ et $\partial^2 U_R$ dans U_R , voir [1].

L'invariance conforme du mouvement brownien plan montre alors que la probabilité conditionnelle

$$P[B[0, T_R] \cap B'[0, T'_R] = \emptyset, B[0, T_R] \subset U_R, B'[0, T'_R] \subset U_R \mid W]$$

est égale à la probabilité que deux mouvements browniens indépendants issus de $\partial^1 H_{L_R}$ ne se rencontrent pas et ne quittent pas H_{L_R} avant d'atteindre le cercle de rayon L_R . Par définition des exposants dans un demi-espace, nous savons que cette probabilité se comporte comme $L_R^{-\xi(1,1)}$.

De même par invariance conforme, la quantité

$$P[B[0, T_R] \subset U_R \mid W]$$

est égale à la probabilité qu'un mouvement brownien issu de $\partial^1 H_{L_R}$ ne quitte pas H_{L_R} avant d'atteindre le cercle de rayon L_R . Une estimation facile montre que cette probabilité se comporte comme L_R^{-1} .

En mettant bout à bout les considérations précédentes, on obtient

$$R^{-\xi(1,1,1)} \approx E[1_{\{W[0, S_R] \subset H\}} L_R^{-\xi(1,1)}] \\ \approx E[1_{\{W[0, S_R] \subset H\}} P[B[0, T_R] \subset U_R \mid W]^{\xi(1,1)}] \\ \approx R^{-\xi(1, \xi(1,1))}$$

où la dernière ligne découle directement de la définition des exposants $\xi(1, \lambda) = \xi(0, 1, \lambda)$ pour λ réel. On obtient la relation $\xi(1, 1, 1) = \xi(1, \xi(1, 1))$ comme annoncé. \triangleright

En utilisant aussi un argument de symétrie, on déduit assez facilement du Théorème 1.3 des *relations de commutation* pour les exposants d'intersection. Par exemple, pour toutes permutations ρ et σ de $\{1, \dots, k\}$, on a

$$\zeta(n_1, \lambda_1, \dots, n_k, \lambda_k) = \zeta(n_{\rho(1)}, \lambda_{\sigma(1)}, \dots, n_{\rho(k)}, \lambda_{\sigma(k)}),$$

et on a un énoncé analogue pour les exposants ξ .

1.5. Le théorème principal

Les relations en cascade permettent d'étendre la définition des exposants à toutes les valeurs réelles des exposants. Commençons par le cas des exposants dans un demi-espace. On montre que la fonction $\lambda \rightarrow \xi(1, \lambda)$ est strictement croissante et continue de $[0, \infty[$ sur $[1, \infty[$ (la continuité sur $]0, \infty[$ découle du fait que cette fonction est concave). Alors, si α et α' sont deux réels positifs quelconques, on définit $\xi(\alpha, \alpha')$ comme étant l'unique réel positif tel que $\xi(1, \xi(\alpha, \alpha')) = \xi(\alpha, 1, \alpha')$ (ce choix est forcé si l'on veut conserver les relations de commutation et les relations en cascade décrites ci-dessus). Puis on définit simplement par récurrence

$$\xi(\alpha_1, \dots, \alpha_k) = \xi(\alpha_1, \xi(\alpha_2, \dots, \alpha_k))$$

pour tous réels $\alpha_1, \dots, \alpha_k \geq 0$. Il est ensuite possible de vérifier que cette définition est cohérente avec le cas où les α_j sont entiers, puis que les relations en cascade ($\xi(\alpha_1, \dots, \alpha_k) = \xi(\alpha_1, \dots, \alpha_j, \xi(\alpha_{j+1}, \dots, \alpha_k))$) ainsi que les relations de commutation ($\xi(\alpha_1, \dots, \alpha_k) = \xi(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(k)})$) sont valides.

On procède (presque) de même pour les exposants ζ . Il y a cependant une difficulté. On ne définit $\zeta(\alpha_1, \dots, \alpha_k)$, pour $k \geq 2$ et $\alpha_1, \dots, \alpha_k$ réels positifs, que si deux au moins des nombres α_i sont supérieurs ou égaux à 1. Si $\alpha_j \geq 1$, on définit β_j par la relation $\alpha_j = \xi(1, \beta_j)$ puis on pose

$$\zeta(\alpha_1, \dots, \alpha_k) = \zeta(1, \xi(\alpha_1, \dots, \alpha_{j-1}, \beta_j, \alpha_{j+1}, \dots, \alpha_k)).$$

La condition que deux des nombres α_i sont supérieurs ou égaux à 1 intervient pour montrer que cette définition de $\zeta(\alpha_1, \dots, \alpha_k)$ est bien cohérente avec celle introduite précédemment. Les relations en cascade ($\zeta(\alpha_1, \dots, \alpha_k) = \zeta(\alpha_1, \dots, \alpha_j, \xi(\alpha_{j+1}, \dots, \alpha_k))$) et les relations de commutation ($\zeta(\alpha_1, \dots, \alpha_k) = \zeta(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(k)})$) restent vraies.

THÉORÈME 1.5 ([24]). — *Il existe une fonction $U : [0, \infty[\rightarrow [0, \infty[$ continue et strictement croissante, avec $U(0) = 0$, $U(\infty) = \infty$, telle que, pour tous $k \geq 1$ et $\alpha_1, \dots, \alpha_k \geq 0$,*

$$\xi(\alpha_1, \dots, \alpha_k) = U^{-1}(U(\alpha_1) + \dots + U(\alpha_k)).$$

De plus, il existe une fonction croissante continue $\eta : [\xi(1, 1), \infty[\longrightarrow [0, \infty[$ telle que, pour tout entier $k \geq 2$ et tous $\alpha_1, \dots, \alpha_k \geq 0$ avec $\alpha_1 \geq 1$ et $\alpha_2 \geq 1$, on a

$$\zeta(\alpha_1, \dots, \alpha_k) = \eta(\xi(\alpha_1, \dots, \alpha_k)).$$

La fonction U peut être obtenue par la formule

$$U^2(\alpha) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \xi(\alpha^{\otimes N}).$$

La preuve de la première partie du théorème utilise fortement les relations en cascade entre les exposants. Un autre ingrédient important est le comportement asymptotique de l'exposant d'intersection pour k paquets de mouvements browniens, lorsque le nombre de mouvements browniens à l'intérieur de chaque paquet tend vers l'infini [33].

L'expression de ξ en fonction de ζ est une conséquence presque directe des relations en cascade. Rappelons que la fonction $\kappa(\lambda) = \xi(1, \lambda)$ est bijective de $[0, \infty[$ sur $[1, \infty[$ et, en supposant $\alpha_1 \geq 1$ notons $\beta_1 = \kappa^{-1}(\alpha_1)$. Alors, en utilisant d'abord la construction donnée ci-dessus de $\zeta(\alpha_1, \dots, \alpha_k)$, puis en appliquant deux fois les relations en cascade, on a

$$\begin{aligned} \zeta(\alpha_1, \dots, \alpha_k) &= \zeta(1, \xi(\beta_1, \alpha_2, \dots, \alpha_k)) \\ &= \zeta(1, \kappa^{-1}(\xi(1, \beta_1, \alpha_2, \dots, \alpha_k))) \\ &= \zeta(1, \kappa^{-1}(\xi(\alpha_1, \dots, \alpha_k))), \end{aligned}$$

ce qui donne la dernière formule du théorème avec $\eta(\lambda) = \zeta(1, \kappa^{-1}(\lambda))$.

REMARQUE. Du point de vue de la physique théorique [8],[9], les formules du Théorème 1.5 trouvent une interprétation dans le cadre de la théorie (non rigoureuse) de la gravité quantique. A titre d'exemple, la fonction U^{-1} est la transformation qui fait passer les exposants du demi-plan "quantique" (où les exposants ont une structure additive) vers le demi-plan "standard".

1.6. Conjectures

Les prédictions de [11] pour les valeurs des exposants $\zeta(1^{\otimes N})$ et $\xi(1^{\otimes N})$ conduisent à conjecturer les formes explicites suivantes pour les fonctions U et η qui interviennent dans le Théorème 1.5 :

$$U(x) = \frac{\sqrt{24x+1}-1}{\sqrt{24}}, \quad \eta(\alpha) = \frac{(\sqrt{24\alpha+1}-1)^2-4}{48}.$$

Il est très plausible que la dernière formule du Théorème 1.5 reste vraie sous la seule condition que $\max(\alpha_i) \geq 1$ (qui est nécessaire pour définir $\zeta(\alpha_1, \dots, \alpha_p)$). En particulier on devrait avoir, pour tout entier $p \geq 1$, $\zeta(p, 0) = \eta(\xi(p, 0)) = \eta(p)$, et la fonction η s'interpréterait comme un exposant de disconnexion généralisé. La valeur $\eta(1) = \frac{1}{4}$ est bien cohérente avec les prédictions de Duplantier et Kwon [11]. De même la valeur

$\eta(2) = \frac{2}{3}$ est cohérente avec la conjecture de Mandelbrot sur la dimension de Hausdorff du contour \mathcal{C} de la courbe brownienne.

Les conjectures pour U et η et le Théorème 1.5 conduisent aux valeurs suivantes pour les exposants ζ et ξ :

$$\zeta(\alpha_1, \dots, \alpha_k) = \frac{(\sqrt{24\alpha_1 + 1} + \dots + \sqrt{24\alpha_k + 1} - k)^2 - 4}{48}$$

$$\xi(\alpha_1, \dots, \alpha_k) = \frac{(\sqrt{24\alpha_1 + 1} + \dots + \sqrt{24\alpha_k + 1} - (k - 1))^2 - 1}{24}.$$

Il est rassurant de voir qu'on retrouve la seule valeur connue $\zeta(2, 1) = 2$!

1.7. Prolongements

Très récemment, Lawler et Werner [25] ont montré une propriété d'universalité remarquable des exposants ξ , ou de manière équivalente de la fonction U du Théorème 1.5. Ces exposants interviennent non seulement dans l'étude des trajectoires browniennes mais bien plus généralement dès que l'on s'intéresse à des mesures définies sur une certaine classe de sous-ensembles compacts du disque unité et vérifiant une propriété d'invariance conforme. Dans ce paragraphe, nous donnons une présentation succincte d'un résultat majeur de [25].

On note D le disque unité ouvert dans le plan. Un *pont* est par définition un sous-ensemble compact K simplement connexe de \bar{D} tel que $\partial K \cap \partial D$ a exactement deux composantes connexes. Le pont est *orienté* lorsque l'on a choisi l'une de ces deux composantes connexes, qui sera notée $d_1 = d_1(K)$, l'autre étant $d_2 = d_2(K)$. On note \mathcal{K} l'ensemble des ponts orientés. Si A_1 et A_2 sont deux arcs (fermés) disjoints sur ∂D , on note $\mathcal{K}(A_1, A_2)$ l'ensemble des ponts orientés tels que $d_1 \subset A_1$ et $d_2 \subset A_2$. Lorsque $A_2 = V_\delta^+ := \{e^{i\theta}; -\delta \leq \theta \leq \delta\}$ et $A_1 = V_\delta^- := -A_2$, on note simplement $\mathcal{K}_\delta = \mathcal{K}(V_\delta^-, V_\delta^+)$. Enfin, si $D' \subset D$ et $A_1 \cup A_2 \subset \partial D'$, on note $\mathcal{K}(A_1, A_2; D')$ l'ensemble des ponts orientés de $\mathcal{K}(A_1, A_2)$ qui ont la propriété additionnelle que $K \cap D \subset D'$.

DÉFINITION 1.6. — *Une mesure μ sur \mathcal{K} est dite complètement invariante conforme (CIC) si elle possède les deux propriétés suivantes.*

- *Il existe un exposant $\alpha = \alpha(\mu) > 0$ tel que, quand $\delta \rightarrow 0$,*

$$\mu(\mathcal{K}_\delta) \approx \delta^{2\alpha}.$$

- *Soit D' un sous-ensemble ouvert simplement connexe de D et soient A_1, A_2 deux arcs (fermés non vides) disjoints contenus dans $\partial D \cap \partial D'$. Il existe une unique valeur $\delta = \delta(A_1, A_2; D')$ et une bijection conforme $\Phi : D' \rightarrow D$ telles que $\Phi(A_1) = V_\delta^-$ et $\Phi(A_2) = V_\delta^+$. Alors, si $\mu_{A_1, A_2}^{D'}$ désigne la restriction de μ à $\mathcal{K}(A_1, A_2; D')$, on a*

$$\Phi(\mu_{A_1, A_2}^{D'}) = \mu_{V_\delta^-, V_\delta^+}^D.$$

Dans la dernière formule, $\Phi(\mu_{A_1, A_2}^{D'})$ désigne la mesure image de $\mu_{A_1, A_2}^{D'}$ par l'application Φ , qui agit de manière évidente sur $\mathcal{K}(A_1, A_2; D')$.

La propriété la plus importante est la deuxième. L'exposant $\alpha(\mu)$ qui intervient dans la première est appelé exposant de croisement de μ . Il mesure la difficulté pour un pont distribué selon μ de passer par des "petites" portes d'entrée et de sortie. On vérifie aisément qu'une mesure CIC est nécessairement de masse totale infinie.

EXEMPLE. — La mesure d'excursion du mouvement brownien plan à l'intérieur du cercle unité permet de construire un premier exemple de mesure CIC, que nous noterons μ_0 et qui fait le lien avec les paragraphes précédents. Le compact K est obtenu sous μ_0 comme complémentaire de la composante connexe non bornée du complémentaire de la trajectoire sous la mesure d'excursion. Dans ce cas, les arcs d_1 et d_2 sont réduits à des points, et on voit facilement que l'exposant de croisement vaut 1.

Nous allons maintenant définir un exposant d'intersection pour une famille finie de mesures CIC. Nous devons auparavant introduire une notation supplémentaire. Si K est un pont orienté, on note $D^- = D^-(K)$ et $D^+ = D^+(K)$ les composantes connexes de $D \setminus K$ numérotées de sorte que les arcs $d_1, \partial D^- \cap \partial D, d_2$ et $\partial D^+ \cap \partial D$ sont dans l'ordre trigonométrique. Soient alors μ_1, \dots, μ_p des mesures CIC, et supposons que les ponts orientés K_1, \dots, K_p sont choisis selon la mesure produit $\mu_1 \otimes \dots \otimes \mu_p$. On note \mathcal{O} l'événement sur lequel les ponts (orientés) K_1, \dots, K_p sont ordonnés, au sens où

$$K_j \cap D \subset D^-(K_{j+1}), \quad j = 1, \dots, p-1$$

(de manière imagée, le pont K_1 est situé "en-dessous" du pont K_2 , lequel est lui-même en-dessous de K_3 , etc.). L'exposant d'intersection $\alpha = \alpha(\mu_1, \dots, \mu_p)$ est alors défini par

$$\mu_1 \otimes \dots \otimes \mu_p((\mathcal{K}_\delta)^p \cap \mathcal{O}) \approx \delta^{2\alpha}$$

quand $\delta \rightarrow 0$, lorsqu'un tel exposant α existe.

Autrement dit, on évalue la "probabilité" que les ponts "indépendants" K_1, \dots, K_p aient des arcs d'entrée et de sortie proches (puisque contenus dans V_δ^- , respectivement V_δ^+) mais néanmoins ne se rencontrent pas (en fait on leur demande même d'être dans un ordre bien défini, mais cela ne fait pas de différence pour le comportement logarithmique de la probabilité). Lorsque μ_1, \dots, μ_p sont toutes égales à la mesure μ_0 construite à partir de la mesure d'excursion brownienne, on vérifie que $\alpha(\mu_0, \dots, \mu_0)$ existe, et

$$\alpha(\mu_0, \dots, \mu_0) = \xi(1, \dots, 1).$$

Le théorème qui suit généralise considérablement cette observation.

THÉORÈME 1.7 ([25]). — Soient μ_1, \dots, μ_p des mesures CIC. Alors l'exposant d'intersection $\alpha(\mu_1, \dots, \mu_p)$ existe, et

$$\alpha(\mu_1, \dots, \mu_p) = \xi(\alpha(\mu_1), \dots, \alpha(\mu_p)).$$

L'énoncé du Théorème 1.7 ne fait plus intervenir de mouvement brownien, si on oublie la manière dont les exposants ξ ont été introduits ci-dessus. Cependant la preuve de ce théorème fait une utilisation intensive de la mesure d'excursion brownienne. L'intérêt du Théorème 1.7 est qu'il est susceptible de s'appliquer aux mesures CIC autres que la mesure μ_0 , que l'on s'attend à obtenir par des passages à la limite convenables à partir de modèles probabilistes comme les marches aléatoires auto-évitantes ou les amas de percolation (tout récemment, Schramm [30] a discuté l'existence de modèles continus pour la marche aléatoire à boucles effacées qui sera étudiée dans la partie suivante). L'article [25] contient une série de conjectures dans cette direction, dont plusieurs rejoignent les prédictions des physiciens théoriciens.

Après la rédaction de la première version de cet exposé, nous avons eu connaissance d'un travail tout récent de Lawler, Schramm et Werner [26] qui démontre certaines des conjectures mentionnées dans le paragraphe 1.6 ci-dessus. L'idée est d'utiliser le processus de croissance introduit par Schramm [30] pour lequel des calculs explicites des exposants sont possibles. Grâce aux résultats d'universalité décrits dans le paragraphe 1.7, cela permet aussi de calculer au moins certaines valeurs des exposants d'intersection browniens, en accord avec les conjectures du paragraphe 1.6. Des résultats même plus complets que ceux de [26], incluant une preuve de la conjecture de Mandelbrot $\dim \mathcal{C} = \frac{4}{3}$, ont été annoncés par Lawler, Schramm et Werner.

2. MARCHES ALÉATOIRES À BOUCLES EFFACÉES ET PAVAGES PAR DOMINOS

2.1. Marches aléatoires à boucles effacées et arbres couvrants.

La marche aléatoire à boucles effacées a été introduite par Lawler [19] en 1980. Commençons par rappeler sa définition. Soit $\gamma = (\gamma(0), \gamma(1), \dots, \gamma(N))$ un chemin déterministe dans \mathcal{Z}^d ou sur un graphe quelconque. Si ce chemin est auto-évitant, c'est-à-dire si les $\gamma(i)$ sont distincts, on arrête là la construction. Sinon, on choisit j minimal tel qu'il existe $i < j$ avec $\gamma(j) = \gamma(i)$, on remplace γ par $\gamma' = (\gamma(0), \dots, \gamma(i), \gamma(j+1), \dots, \gamma(N))$ et on recommence. La construction s'arrête au bout d'un nombre fini d'étapes et donne alors un chemin auto-évitant ayant mêmes points de départ et d'arrivée que γ .

Si on considère maintenant un chemin infini $\gamma = (\gamma(0), \gamma(1), \dots)$ ayant la propriété de ne passer qu'un nombre fini de fois en chaque sommet du graphe, la même méthode fournit un chemin infini auto-évitant. Ceci s'applique à une trajectoire de marche aléatoire simple dans \mathcal{Z}^d pour $d \geq 3$ et le chemin aléatoire ainsi obtenu est appelé marche aléatoire à boucles effacées dans \mathcal{Z}^d . En revanche, dans le graphe \mathcal{Z}^2 (ou dans $\mathcal{N} \times \mathcal{Z}$), la construction précédente ne marche pas directement à cause de la récurrence de la marche aléatoire simple. Néanmoins, on peut l'appliquer d'abord à la marche aléatoire issue de 0 arrêtée au premier temps de sortie d'une grande boîte centrée en 0, puis faire tendre la taille de

la boîte vers l'infini pour obtenir encore la marche aléatoire à boucles effacées dans \mathcal{Z}^2 , ou dans $\mathcal{N} \times \mathcal{Z}$ (voir le dernier chapitre de [20]). L'article récent de Lawler [23] donne un aperçu très complet des problèmes ouverts et des conjectures sur les marches aléatoires à boucles effacées.

Le théorème suivant, dû à Kenyon [17], résout l'une des plus importantes de ces conjectures.

THÉORÈME 2.1. — *Pour la marche aléatoire à boucles effacées dans $\mathcal{N} \times \mathcal{Z}$ issue de l'origine, le nombre moyen de points visités qui sont à distance inférieure à N de l'origine est $N^{\frac{5}{4}+o(1)}$. Plus précisément, la probabilité qu'un point de la forme $(x, y) = re^{i\theta}$ soit visité se comporte quand $r \rightarrow \infty$ comme*

$$r^{-\frac{3}{4}(1+o(1))} ((\cos \theta)^{\frac{1}{4}} + o(1)),$$

où le terme $o(1)$ dans l'exposant ne dépend pas de θ .

On s'attend à ce qu'un résultat analogue soit vrai pour la marche aléatoire à boucles effacées dans \mathcal{Z}^2 , même si cette extension présente probablement des difficultés techniques. Le théorème suggère fortement que la moyenne à l'instant n du carré de la norme d'une marche aléatoire à boucles effacées dans \mathcal{Z}^2 se comporte asymptotiquement comme $n^{\frac{8}{5}}$. Cette valeur de l'exposant $\gamma = \frac{8}{5}$ a été prédite par les physiciens théoriciens (Majumdar [27], voir aussi [7]). Lawler [20] donne la minoration $\gamma \geq \frac{3}{2}$.

La formulation du théorème précédent qui figure dans [17] est différente et utilise la notion d'arbre couvrant uniforme. Si G est un graphe connexe quelconque, on appelle arbre couvrant de G un sous-graphe connexe de G qui est un arbre et qui contient pour chaque sommet de G au moins une arête incidente à ce sommet. Lorsque G est fini, on définit un arbre couvrant uniforme comme étant un sous-graphe aléatoire de G dont la loi est la probabilité uniforme sur l'ensemble (fini) de tous les arbres couvrants de G .

Cette définition peut être étendue à des graphes infinis. Considérons le cas où $G = \mathcal{Z}^d$. Pour tout entier $n \geq 1$, notons D_n le cube de côté $2n$ centré en l'origine dans \mathcal{Z}^d . Soit μ_n la probabilité uniforme sur l'ensemble des arbres couvrants de D_n , vu ici comme contenu dans l'ensemble des sous-graphes de \mathcal{Z}^d , qui est muni de la topologie-produit (on identifie un sous-graphe à un élément de $\{0, 1\}^{\mathcal{A}}$, si \mathcal{A} est l'ensemble des arêtes de \mathcal{Z}^d). On montre alors [29] que μ_n converge étroitement vers une mesure de probabilité μ portée par les sous-graphes de \mathcal{Z}^d . De plus, lorsque $d \leq 4$, la mesure μ est portée par les arbres couvrants de \mathcal{Z}^d , et on appelle arbre couvrant uniforme sur \mathcal{Z}^d un arbre aléatoire dont la loi est μ . La même construction s'applique au graphe $\mathcal{N} \times \mathcal{Z}$.

On montre aussi [29], toujours en dimension $d \leq 4$, que la mesure μ est portée par les arbres qui ont un seul "bout", c'est-à-dire tels que la suppression d'un sommet divise l'arbre en deux composantes dont une seule est infinie. Il est alors immédiat qu'il existe dans l'arbre couvrant uniforme un seul chemin (sans recoupement) joignant 0 à ∞ .

PROPOSITION 2.2 ([29]). — *Pour l'arbre aléatoire couvrant uniforme dans \mathcal{Z}^d , $d \leq 4$ ou dans $\mathcal{N} \times \mathcal{Z}$, l'arc joignant 0 à l'infini a la loi d'une marche aléatoire à boucles effacées issue de 0.*

2.2. Arbres couvrants et pavages par dominos

Grâce à la Proposition 2.2, on peut ramener la preuve du Théorème 2.1 à celle d'un énoncé équivalent sur les arbres couvrants. Nous allons voir qu'on peut introduire encore une autre formulation mettant en jeu les pavages par dominos de certaines régions du plan.

Soit T le pavage "en échiquier" de \mathcal{R}^2 par des carrés unité, chaque carré étant centré en un point du réseau \mathcal{Z}^2 . Nous supposons que le carré centré en l'origine est blanc, et nous notons W_0 l'ensemble des carrés (blancs) dont les coordonnées du centre sont paires, W_1 l'ensemble des autres carrés blancs. Nous notons aussi B_0 l'ensemble des carrés noirs dont les coordonnées du centre sont égales à $(1,0)$ modulo 2, B_1 l'ensemble des autres carrés noirs.

Un *polyomino* est une réunion finie de carrés de T dont la frontière est une courbe fermée simple (pour simplifier, nous ne considérons ici que des polyominos simplement connexes). Un polyomino est dit *pair* si tous les carrés de coin, que ce coin soit convexe ou concave, sont de type B_1 . Enfin un polyomino *temperlien* est obtenu à partir d'un polyomino pair en enlevant un (seul) carré noir d adjacent à la frontière de ce dernier (d sera appelé le *point de base* du polyomino). A cause de la définition d'un polyomino pair, d est forcément de type B_1 (dans un polyomino pair les seuls carrés noirs adjacents à la frontière sont de type B_1). Remarquons qu'un polyomino temperlien contient le même nombre de carrés blancs et noirs.

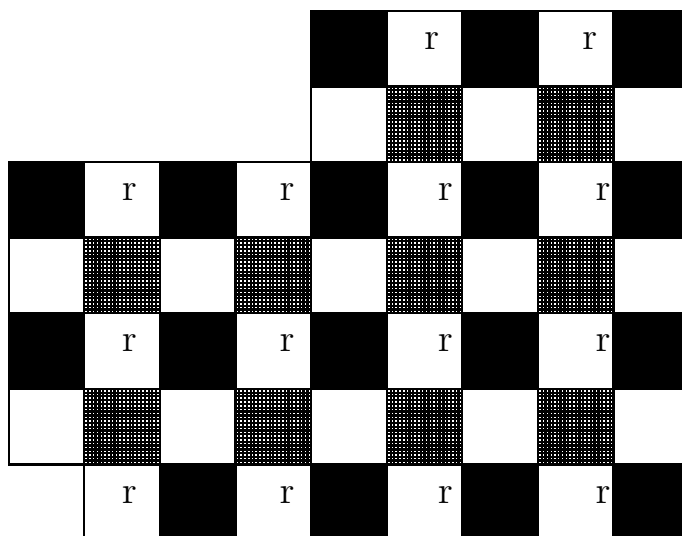


Figure 1

La Figure 1 donne un exemple de polyomino temperlien, pour lequel le carré de base (qui ne fait pas partie du polyomino) est en bas à gauche. Sur cette figure, les carrés noirs sont de type B_1 , les carrés gris de type B_0 , les carrés blancs pointés de type W_1 et enfin les carrés blancs de type W_0 .

Soit P un polyomino temperlien, et P' le polyomino pair dont est issu P . On note $\mathbf{B}_1(P)$ le graphe dont les sommets sont les carrés de type B_1 contenus dans P' , deux sommets étant reliés par une arête si les centres des carrés sont à distance 2. Remarquons que d est un sommet du graphe $\mathbf{B}_1(P)$, et qu'à chaque arête de $\mathbf{B}_1(P)$ est associé un unique carré blanc de P . Le dual du graphe $\mathbf{B}_1(P)$ est le graphe $\mathbf{B}_0(P)$ dont les sommets sont les carrés de type B_0 dans P .

On appelle *domino* un rectangle 1×2 ou 2×1 dont les sommets sont des points de \mathbb{Z}^2 . Il est facile de vérifier qu'un polyomino temperlien est toujours pavable par dominos. Le théorème suivant est dû à Temperley [31] dans le cas d'un rectangle (voir [5] et [18] pour des généralisations incluant celle-ci).

THÉORÈME 2.3. — *Les pavages par dominos de P sont en bijection avec les arbres couvrants du graphe $\mathbf{B}_1(P)$.*

De manière intuitive, si l'on part d'un arbre couvrant dont on oriente les arêtes en partant du point de base d , on construit le pavage par dominos correspondant de la manière suivante. On "jette" pour chaque arête de l'arbre un domino dont le carré noir est le sommet terminal de l'arête et le carré blanc est le carré associé à l'arête. Pour compléter le pavage, il suffit ensuite de "boucher les trous", ou, ce qui revient au même, de jeter pareillement les dominos le long de l'arbre dual. La Figure 2 donne un exemple de pavage par dominos du polyomino de la Figure 1 et de l'arbre associé dont les sommets sont les points de $\mathbf{B}_1(P)$.

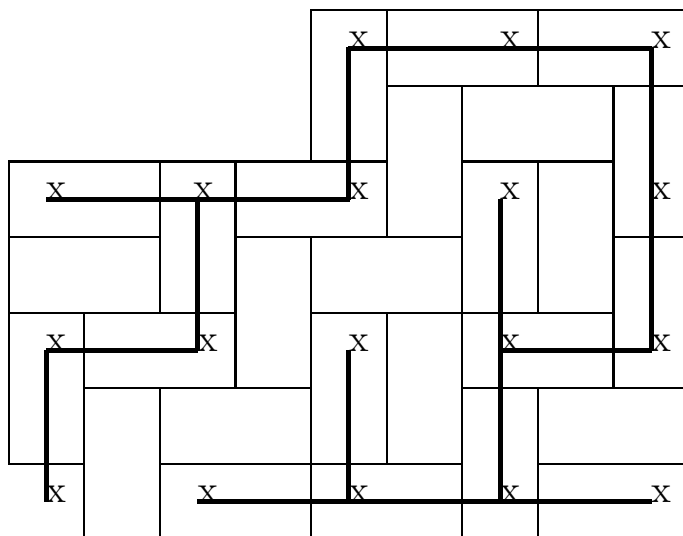


Figure 2

Considérons maintenant, pour un entier pair n , le polyomino temperlien P rectangulaire défini comme la réunion des carrés unité centrés aux points de $\{1, 2, \dots, 2n + 1\} \times \{-n, -n + 1, \dots, n - 1, n\}$, à l'exception du carré d centré en $(1, 0)$. Soit Q l'ensemble obtenu à partir de P en enlevant un carré noir b adjacent à la frontière de P (donc de type B_1) ainsi qu'un carré blanc w (qu'on peut voir comme une arête de $\mathbf{B}_1(P)$).

LEMME 2.4 ([17]). — *Les pavages par dominos de Q sont en bijection avec les arbres couvrants de $\mathbf{B}_1(P)$ pour lesquels l'arc de b à d contient l'arête w .*

Ce lemme nous permet d'esquisser la méthode d'approche du Théorème 2.1 [17]. Nous voulons estimer la probabilité que dans l'arbre couvrant uniforme sur $\mathcal{N} \times \mathcal{Z}$, l'arc joignant 0 à ∞ contienne l'arête w fixée. Le lemme précédent nous permettra d'estimer la probabilité que pour l'arbre couvrant uniforme dans une boîte rectangulaire arbitrairement grande, l'arc joignant le point de base à un autre point de la frontière contienne l'arête w . Cette probabilité s'exprimera d'après le Lemme 2.4 et le Théorème 2.3 comme le rapport du nombre de pavages par dominos de Q sur le nombre de pavages par dominos de P . Dans les sections suivantes, nous expliquons comment on estime le nombre de pavages par dominos d'un polyomino temperlien. Ceci ne s'applique pas directement à Q qui n'est pas temperlien, mais la méthode peut être adaptée.

2.3. Matrice de Kasteleyn et fonction de couplage

Dans ce paragraphe, nous considérons un polyomino temperlien P et nous notons M l'ensemble des carrés de P . L'ensemble M est muni d'une structure de graphe évidente, où deux carrés sont adjacents ssi ils ont un côté en commun. La *matrice de Kasteleyn* de P est la matrice symétrique $K = (K(v, v'), v, v' \in M)$ telle que $K(v, v') = 0$ si v et v' ne sont pas adjacents et, pour $v \in M \cap (W_0 \cup W_1)$ et $v' \in M$,

$$K(v, v') = \begin{cases} 1 & \text{si } v' = v + (1, 0) \\ i & \text{si } v' = v + (0, 1) \\ -1 & \text{si } v' = v + (-1, 0) \\ -i & \text{si } v' = v + (0, -1). \end{cases}$$

Kasteleyn [14] a montré (avec une définition différente de K) que le nombre de pavages par dominos de P est la racine carrée du module du déterminant de K . En général, sauf dans le cas du rectangle où une formule explicite existe, il est cependant difficile d'utiliser ce résultat pour estimer le nombre de pavages par dominos de P .

La *fonction de couplage* de P est par définition la matrice inverse $C = K^{-1}$.

PROPOSITION 2.5 ([15]). — *Si u, v sont deux carrés adjacents de P , la probabilité qu'un pavage aléatoire uniforme de P contienne le domino $\{u, v\}$ est égale à $|C(u, v)|$.*

Plus généralement [15], la probabilité qu'un pavage aléatoire uniforme contienne une famille fixée de dominos disjoints s'exprime comme la valeur absolue du déterminant d'une sous-matrice de C .

Nous verrons dans la partie suivante que si l'on se donne une suite de polyominos temperliens P_ε de $\varepsilon\mathbb{Z}^2$ qui approchent une région du plan, il est possible d'obtenir des informations précises sur le comportement asymptotique des fonctions de couplage C_ε associées. Afin de mieux comprendre l'origine de ces informations, donnons quelques propriétés simples de la fonction de couplage.

Si on énumère les éléments de M en écrivant successivement les carrés de type W_0, W_1, B_0 et enfin B_1 , la matrice K est de la forme

$$K = \begin{pmatrix} 0 & 0 & K_1 & iK_2 \\ 0 & 0 & iK_3 & K_4 \\ K_1^t & iK_3^t & 0 & 0 \\ iK_2^t & K_4^t & 0 & 0 \end{pmatrix}$$

où les matrices K_1, K_2, K_3, K_4 sont réelles. Il est facile de voir que l'inverse de K doit être aussi de cette forme. En particulier, $C(v, v') = 0$ si v et v' sont tous les deux blancs, ou tous les deux noirs, $C(v, v')$ est réel quand $v' = v + (1, 0)$ modulo 2 et imaginaire pur quand $v' = v + (0, 1)$ modulo 2.

Par définition de C , on a $KC(v, v') = \delta_v(v')$, où δ_v est la fonction indicatrice du singleton $\{v\}$. Notons K^* la matrice conjuguée de K . On vérifie facilement que la matrice K^*K agit sur les fonctions définies sur $\mathbf{B}_0(P)$, et que pour une telle fonction f on a pour tout $v \in \mathbf{B}_0(P)$,

$$(2) \quad K^*Kf(v) = 4f(v) - f(v + (2, 0)) - f(v + (-2, 0)) - f(v + (0, 2)) - f(v + (0, -2)).$$

Cette formule est vraie même si les quatre voisins (dans B_0) de v ne sont pas tous dans $\mathbf{B}_0(P)$, dans quel cas il faut affecter la valeur 0 à la fonction f aux points extérieurs à $\mathbf{B}_0(P)$. La formule précédente s'interprète en disant que $K^*K = \Delta$ est (4 fois) le Laplacien discret sur $\mathbf{B}_0(P)$, avec conditions de Dirichlet au bord. Ce qui précède reste vrai si on remplace $\mathbf{B}_0(P)$ par $\tilde{\mathbf{B}}_1(P) = \mathbf{B}_1(P) \setminus \{d\}$, mais la formule (2) doit être modifiée si v est adjacent à la frontière de P . Dans ce cas, le terme de droite de (2) est $f(v)$ multiplié par le nombre de voisins de v dans $\mathbf{B}_1(P)$ moins la somme des valeurs de f en ces voisins (en prenant $f(d) = 0$).

En appliquant K^* à l'égalité $KC(v, v') = \delta_v(v')$, on trouve pour $v \in W_0 \cup W_1$,

$$\Delta C(v, \cdot) = \delta_{v+(1,0)} - \delta_{v+(-1,0)} - i\delta_{v+(0,1)} + i\delta_{v+(0,-1)}.$$

Fixons $v \in W_0 \cap M$. La restriction, notée f , de $C(v, \cdot)$ à $\mathbf{B}_0(P)$ est à valeurs réelles, et l'égalité précédente montre que cette fonction est harmonique (au sens discret) sur $\mathbf{B}_0(P) \setminus \{v - (1, 0), v + (1, 0)\}$, avec conditions frontière de type Dirichlet. En revanche la restriction, notée ig , de $C(v, \cdot)$ à $\tilde{\mathbf{B}}_1(P)$ est à valeurs imaginaires pures et est harmonique

sur $\tilde{\mathbf{B}}_1(P) \setminus \{v + (0, 1), v + (0, -1)\}$. Les conditions frontière sont maintenant de type Neumann : la valeur de g en un point frontière v' est la moyenne de ses valeurs aux voisins de v' dans $\mathbf{B}_1(P)$, en prenant $g(d) = 0$.

De plus, on peut dire que g est la conjuguée harmonique de f , au sens où on a les équations de Cauchy-Riemann discrètes

$$\begin{aligned}\partial_x f(v') &= \partial_y g(v') && \text{pour } v' \in (M \cap W_0) \setminus \{v\} \\ \partial_y f(v') &= -\partial_x g(v') && \text{pour } v' \in M \cap W_1,\end{aligned}$$

avec la notation $\partial_x f(v) = f(v + (1, 0)) - f(v + (-1, 0))$, $\partial_y f(v) = f(v + (0, 1)) - f(v + (0, -1))$.

On a des résultats analogues pour la fonction $C(v, \cdot)$ lorsque $v \in W_1$.

2.4. Comportement asymptotique de la fonction de couplage

Soit U un polygone rectiligne de \mathcal{R}^2 (U est un domaine simplement connexe et sa frontière est composée de segments parallèles aux axes), et soit $d_0 \in \partial U$. Considérons pour chaque $\varepsilon > 0$ un polyomino temperlien P_ε de $\varepsilon\mathcal{Z}^2$, et supposons que ces polyominos convergent vers U de la manière évidente (en particulier, P_ε a autant de coins que U , et chaque coin de P_ε converge vers le coin correspondant de U). On suppose aussi que les points de base d_ε de P_ε convergent vers le point d_0 .

Notons C_ε la fonction de couplage de P_ε et M_ε l'ensemble des carrés de P_ε .

THÉORÈME 2.6 ([16]). — *Soient, pour chaque $\varepsilon > 0$, $u_\varepsilon, v_\varepsilon, w_\varepsilon, x_\varepsilon$ des éléments de M_ε appartenant respectivement à W_0, W_1, B_0, B_1 , et convergeant respectivement vers des points distincts $u, v, w, x \in U$. Alors,*

$$\begin{aligned}\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} C_\varepsilon(u_\varepsilon, w_\varepsilon) &= \operatorname{Re} F_0(u, w) \\ \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} C_\varepsilon(u_\varepsilon, x_\varepsilon) &= i \operatorname{Im} F_0(u, x) \\ \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} C_\varepsilon(v_\varepsilon, w_\varepsilon) &= \operatorname{Re} F_1(v, w) \\ \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} C_\varepsilon(v_\varepsilon, x_\varepsilon) &= i \operatorname{Im} F_1(v, x)\end{aligned}$$

où les fonctions F_0 et F_1 sont caractérisées par les propriétés suivantes.

Pour $v \in U$, la fonction $F_0(v, \cdot)$ est méromorphe sur U , sa partie réelle s'annule sur ∂U , elle a un zéro en $z = d_0$, un pôle simple de résidu $\frac{1}{\pi}$ en $z = v$ et pas d'autre pôle sur \bar{U} .

De même, la fonction $F_1(v, \cdot)$ est méromorphe sur U , sa partie imaginaire s'annule sur ∂U , elle a un zéro en $z = d_0$, un pôle simple de résidu $\frac{1}{\pi}$ en $z = v$ et pas d'autre pôle sur \bar{U} .

REMARQUE — Un résultat analogue est vrai pour des ouverts U plus généraux : voir [16].

Les propriétés des fonctions F_0 et F_1 , et le fait qu'on obtienne tantôt $\operatorname{Re} F$ tantôt $\operatorname{Im} F$, viennent directement des propriétés des fonctions de couplage décrites dans la partie précédente. Donnons rapidement l'idée de la preuve du théorème. On note $\mathbf{B}'_0(P_\varepsilon)$ le graphe obtenu en ajoutant à $\mathbf{B}_0(P_\varepsilon)$ tous les carrés noirs (de type B_0) qui sont adjacents à un carré blanc à la frontière de P_ε , et les arêtes reliant deux carrés noirs si le carré blanc situé entre eux est dans P_ε . Soit alors $G_\varepsilon(w, w')$ la fonction de Green de $\mathbf{B}'_0(P_\varepsilon)$: pour $w \in \mathbf{B}_0(P_\varepsilon)$, la fonction $G_\varepsilon(w, \cdot)$ vérifie $\Delta G_\varepsilon(w, \cdot) = \delta_w$ sur $\mathbf{B}_0(P_\varepsilon)$, et $G_\varepsilon(w, w') = 0$ pour $w' \in \mathbf{B}'_0(P_\varepsilon) \setminus \mathbf{B}_0(P_\varepsilon)$. Par abus de notation, écrivons $\operatorname{Re} C_\varepsilon(u_\varepsilon, \cdot)$ pour la restriction de $C_\varepsilon(u_\varepsilon, \cdot)$ à $\mathbf{B}_0(P_\varepsilon)$ (qui est à valeurs réelles) et prolongeons-la à $\mathbf{B}'_0(P_\varepsilon)$ en lui donnant la valeur 0 sur $\mathbf{B}'_0(P_\varepsilon) \setminus \mathbf{B}_0(P_\varepsilon)$. Nous avons vu que, sur $\mathbf{B}_0(P_\varepsilon)$,

$$\Delta \operatorname{Re} C_\varepsilon(u_\varepsilon, \cdot) = \delta_{u_\varepsilon + \varepsilon} - \delta_{u_\varepsilon - \varepsilon}.$$

Puisque $\operatorname{Re} C_\varepsilon(u_\varepsilon, \cdot)$ s'annule à la frontière, il en découle que

$$\operatorname{Re} C_\varepsilon(u_\varepsilon, w) = G_\varepsilon(u_\varepsilon + \varepsilon, w) - G_\varepsilon(u_\varepsilon - \varepsilon, w).$$

On peut alors déduire la première convergence du théorème de résultats sur le comportement asymptotique des fonctions de Green discrètes.

Les fonctions de couplage asymptotiques F_0 et F_1 possèdent des propriétés d'invariance conforme qui sont très importantes pour les applications. Posons $F_+^U = F_0 + F_1$ et $F_-^U = F_0 - F_1$. Alors, $F_+^U(v, z)$ est méromorphe en v et en z , et $F_-^U(v, z)$ est méromorphe en z et anti-méromorphe en v . De plus, si $f : V \rightarrow U$ est une bijection conforme envoyant le point de base de V sur le point de base de U , on a :

$$\begin{aligned} F_+^V(v, z) &= f'(v) F_+^U(f(v), f(z)), \\ F_-^V(v, z) &= \overline{f'(v)} F_-^U(f(v), f(z)). \end{aligned}$$

2.5. L'asymptotique du nombre de pavages par dominos

Dans ce paragraphe, nous décrivons brièvement un résultat important de [17] donnant l'asymptotique du logarithme du nombre $N(P_\varepsilon)$ de pavages par dominos de P_ε (sous les hypothèses de la partie précédente). Ce résultat ne peut pas être appliqué directement à la preuve du Théorème 2.1, qui nécessite d'estimer le rapport $N(Q_\varepsilon)/N(P_\varepsilon)$ où Q_ε n'est pas temperlien. Cependant, les techniques impliquées dans cette preuve sont proches de celles que nous décrivons très succinctement ci-dessous.

Nous notons A_ε l'aire de P_ε , $\operatorname{perim}(P_\varepsilon)$ son périmètre, et V le nombre de sommets du polygone V .

THÉORÈME 2.7 ([17]). — Soit $N(P_\varepsilon)$ le nombre de pavages par dominos de P_ε . Alors,

$$\log N(P_\varepsilon) = \frac{c_0 A_\varepsilon}{\varepsilon^2} + \frac{c_1 \operatorname{perim}(P_\varepsilon)}{\varepsilon} - \frac{1}{2} \log \frac{1}{\varepsilon} - \frac{V-4}{36} h(\varepsilon) + c_2(U) + o(1),$$

où $c_0 = \frac{G}{\pi}$, si $G = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \dots$ est la constante de Catalan, $c_1 = \frac{G}{2\pi} + \frac{\log(\sqrt{2}-1)}{4}$, $c_2(U)$ est une constante dépendant seulement de U , et la fonction h est indépendante de U et vérifie $h(\varepsilon) = \log \frac{1}{\varepsilon} + o(\log \frac{1}{\varepsilon})$.

REMARQUE. — Le premier terme du développement ci-dessus est essentiellement contenu dans le travail de Burton et Pemantle [5].

Dans le cas particulier où U est un rectangle, le développement du théorème peut être obtenu à partir des expressions explicites dues à Kasteleyn [13] et Temperley et Fischer [32] (voir [10] pour une dérivation du développement dans ce cas). Pour se ramener au cas du rectangle, on utilise une technique de découpage. Considérons un segment (vertical ou horizontal) γ' qui traverse U , touche ∂U seulement en ses points de départ et d'arrivée, et évite le point de base d_0 de U . Pour chaque $\varepsilon > 0$, considérons aussi une bande γ_ε de largeur ε formée de carrés de P_ε , qui traverse P_ε en restant à une distance $O(\varepsilon)$ de γ' . On peut imposer que les carrés noirs contenus dans γ_ε soient de type B_0 , et que si l'une ou l'autre extrémité de γ' est à un coin concave de U , l'extrémité correspondante de γ_ε soit au coin correspondant de P_ε .

La longueur de γ_ε doit alors être un multiple impair de ε . Si nous retirons la bande γ_ε de P_ε , il reste deux polyominos disjoints. Celui, noté P_1 , qui contient le carré de base de P_ε est temperlien. L'autre, noté P_2 , devient temperlien si on lui enlève un seul carré s de type B_1 adjacent à l'un des points terminaux de γ_ε . Remarquons que la réunion $\gamma_\varepsilon \cup s$ a un seul pavage par dominos. Le nombre de pavages par dominos de P_ε est égal au produit du nombre de pavages de P_1 et du nombre de pavages de P_2 , divisé par la probabilité que le pavage de $\gamma_\varepsilon \cup s$ apparaisse dans un pavage de P_ε tiré au hasard uniformément.

Nous pouvons répéter cette procédure à partir de P_1 et P_2 de manière à découper le polyomino initial en des morceaux de plus en plus simples. Au bout d'un nombre fini d'itérations on arrive à des rectangles temperliens pour lesquels on peut utiliser la formule exacte. Tout le problème est donc d'estimer la probabilité que dans un pavage aléatoire uniforme de P_ε , apparaisse le pavage de $\gamma_\varepsilon \cup s$. Notons a_1, \dots, a_N les dominos qui forment ce pavage. Les a_i sont mis bout-à-bout (à partir de la frontière) sauf pour a_N qui est perpendiculaire aux autres. La probabilité que tous les a_i soient présents s'écrit comme le produit

$$\prod_{j=1}^N P(a_j \mid a_1, \dots, a_{j-1})$$

des probabilités conditionnelles que a_j soit présent sachant que a_1, \dots, a_{j-1} le sont. Pour j fixé, cette dernière probabilité conditionnelle est la probabilité que a_j soit présent dans un pavage aléatoire de la région $P_\varepsilon^{(j)} := P_\varepsilon \setminus \{a_1, \dots, a_{j-1}\}$, qui est elle-même un polyomino temperlien. Soit $C_\varepsilon^{(j)}$ la fonction de couplage de $P_\varepsilon^{(j)}$, et soient $u_\varepsilon^{(j)}, v_\varepsilon^{(j)}$ les deux carrés qui composent le domino a_j . Alors, la probabilité que a_j soit présent dans un pavage aléatoire de $P_\varepsilon^{(j)}$ est d'après la Proposition 2.5 égale à $|C_\varepsilon^{(j)}(u_\varepsilon^{(j)}, v_\varepsilon^{(j)})|$. Le comportement asymptotique de ces quantités peut être décrit en termes des fonctions de couplage asymptotiques

F_0, F_1 de la partie 4, correspondant à la région U privée d'une coupure le long (d'une partie) du segment γ' . A l'aide d'une analyse détaillée de ces asymptotiques et d'une utilisation intensive des propriétés d'invariance conforme, cela conduit à des estimations suffisamment précises pour établir le Théorème 2.7.

BIBLIOGRAPHIE

- [1] L.V. AHLFORS – *Conformal Invariants, Topics in Geometric Function Theory*. McGraw-Hill, New York 1973
- [2] K. BURDZY, G.F. LAWLER – *Non-intersection exponents for random walk and Brownian motion. Part I: Existence and an invariance principle*. Probab. Th. Rel. Fields **84** (1990), 393-410.
- [3] K. BURDZY, G.F. LAWLER – *Non-intersection exponents for random walk and Brownian motion. Part II: Estimates and applications to a random fractal*. Ann. Probab. **18** (1990), 981-1009.
- [4] K. BURDZY, W. WERNER – *No triple point of planar Brownian motion is accessible*. Ann. Probab. **24** (1996), 125-147.
- [5] R. BURTON, R. PEMANTLE – *Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer impedances*. Ann. Probab. **21** (1993), 1329-1371.
- [6] M. CRANSTON, T. MOUNTFORD – *An extension of a result by Burdzy and Lawler*. Probab. Th. Rel. Fields **89** (1991), 487-502.
- [7] B. DUPLANTIER – *Loop-erased self-avoiding walks in two dimensions: exact critical exponents and winding numbers*. Physica A **191** (1992), 516-522.
- [8] B. DUPLANTIER – *Random walks and quantum gravity in two dimensions*. Phys. Rev. Lett. **81** (1998), 5489-5492.
- [9] B. DUPLANTIER – *Two-dimensional copolymers and exact conformal multifractality*. Phys. Rev. Lett. **82** (1999), 880-883.
- [10] B. DUPLANTIER, F. DAVID – *Exact partition functions and correlation functions of multiple Hamiltonian walks on the Manhattan lattice*. J. Stat. Phys. **51** (1988), 327-434.
- [11] B. DUPLANTIER, K.-H. KWON – *Conformal invariance and intersections of random walks*. Phys. Rev. Lett. **61** (1988), 2514-2517.
- [12] B. DUPLANTIER, G.F. LAWLER, J.F. LE GALL, T.J. LYONS – *The geometry of the Brownian curve*. Bull. Sci. math. **117** (1993), 91-106.
- [13] P. KASTELEYN – *The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice*. Physica **27** (1961), 1209-1225.

- [14] P. KASTELEYN – *Graph theory and crystal physics*. In: *Graph Theory and Theoretical Physics*. Academic Press, London 1967
- [15] R. KENYON – *Local statistics of lattice dimers*. Ann. Inst. Henri Poincaré Probab. Stat. **33** (1997), 591-618.
- [16] R. KENYON – *Conformal invariance of domino tilings*. Preprint.
- [17] R. KENYON – *The asymptotic determinant of the discrete Laplacian*. Preprint.
- [18] R. KENYON, J. PROPP, D. WILSON – *Trees and matchings*. Preprint.
- [19] G.F. LAWLER – *A self-avoiding random walk*. Duke Math. J. **47** (1980), 655-693.
- [20] G.F. LAWLER – *Intersections of Random Walks*. Birkhäuser, Boston 1991
- [21] G.F. LAWLER – *Hausdorff dimension of cut points for Brownian motion*. Electron. J. Probab. **1** (1996), no 2.
- [22] G.F. LAWLER – *The dimension of the frontier of planar Brownian motion*. Electronic Comm. Probab. **1** (1996), no 5.
- [23] G.F. LAWLER – *Loop-erased random walk*. In: *Perplexing Problems in Probability*. Birkhäuser, Boston 1999
- [24] G.F. LAWLER, W. WERNER – *Intersection exponents for planar Brownian motion*. Preprint.
- [25] G.F. LAWLER, W. WERNER – *Universality for conformally invariant intersection exponents*. Preprint.
- [26] G.F. LAWLER, O. SCHRAMM, W. WERNER – *Values of Brownian intersection exponents I: Half-plane exponents*. Preprint.
- [27] S.N. MAJUMDAR – *Exact fractal dimension of the loop-erased self-avoiding walk in two dimensions*. Phys. Rev. Lett. **68** (1992), 2329-2331.
- [28] B.B. MANDELBROT – *The Fractal Geometry of Nature*. Freeman 1982.
- [29] R. PEMANTLE – *Choosing a spanning tree for the integer lattice uniformly*. Ann. Probab. **19** (1991), 1559-1574.
- [30] O. SCHRAMM – *Scaling limits of loop-erased random walks and uniform spanning trees*. Preprint.
- [31] H. TEMPERLEY – *Combinatorics: Proceedings of the British Combinatorial Conference 1973*. London Math. Soc. Lecture Notes Series **13** (1974), 202-204.
- [32] H. TEMPERLEY, M. FISHER – *The dimer problem in statistical mechanics – an exact result*. Phil. Mag. **6** (1961), 1061-1063.
- [33] W. WERNER – *Asymptotic behaviour of disconnection and intersection exponents*. Probab. Th. Rel. Fields **108** (1997), 131-152.

866-23

DMA – École normale supérieure

45, rue d'Ulm

75230 PARIS Cedex 05

E-mail : legall@dmi.ens.fr

LA CORRESPONDANCE DE MCKAY

by Miles REID

1. COMMENT C'EST

1.1. Model case: the binary dihedral group BD_{4n}

For $G \subset \mathrm{SL}(2, \mathbb{C})$ a finite group, the quotient variety $X = \mathbb{C}^2/G$ is called a *Klein quotient singularity*. I draw the quotient map $\pi: \mathbb{C}^2 \rightarrow X$ and the minimal resolution of singularities $Y \rightarrow X$ together in the diagram:

$$\begin{array}{ccc} & & \mathbb{C}^2 \\ & & \downarrow \pi \\ Y & \xrightarrow{\varphi} & X \end{array}$$

This situation has been well studied, since Klein around 1870 and Coxeter and Du Val in the 1930s: the subgroup G is classified as cyclic, binary dihedral or a binary group corresponding to one of the Platonic solids; the quotient singularity is a hypersurface $X \subset \mathbb{C}^3$ with defining equation one of a list of simple functions. The resolution Y is a surface with $K_Y = \varphi^* K_X$, and the exceptional locus $\varphi^{-1}(0) \subset Y$ of the resolution consists of a bunch of -2 -curves E_i (that is, $E_i \cong \mathbb{P}_{\mathbb{C}}^1$ and E_i has self-intersection $E_i^2 = -2$), and the intersection $E_i E_j$ is given by one of the Dynkin diagrams A_n, D_n, E_6, E_7, E_8 . To avoid writing out lists, let us simply discuss the binary dihedral group

$$G = \mathrm{BD}_{4n} = \langle \alpha, \beta \rangle, \quad \text{where} \quad \alpha = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

where $\varepsilon = \exp \frac{2\pi i}{2n}$. If u, v are coordinates on \mathbb{C}^2 , the G -invariant polynomials are

$$\mathbb{C}[x, y, z]/(z^2 - yx^2 + 4y^{n+1}), \quad \text{where} \quad x = u^{2n} + v^{2n}, y = u^2 v^2, z = uv(u^{2n} - v^{2n});$$

thus the quotient variety is the singularity $X : (z^2 = yx^2 - 4y^{n+1}) \subset \mathbb{C}^3$ of type D_{n+2} , and the quotient morphism $\pi: (u, v) \mapsto (x, y, z)$. The resolution of singularities $Y \rightarrow X$ has

exceptional locus consisting of -2 -curves E_1, \dots, E_{n+2} forming the D_{n+2} configuration:

(1.1)

The classical McKay correspondence begins in the late 1970s with the observation that the same graph arises in connection with the representation theory of G . For a group G and a given representation Q , the *McKay graph* (or McKay quiver) has a node for each irreducible representation, and an edge $V \rightarrow V'$ whenever V' is a direct summand of $V \otimes Q$. In our case, BD_{4n} has the 2-dimensional representations

$$V_i \cong \mathbb{C}^2, \quad \text{with action } \alpha = \begin{pmatrix} \varepsilon^i & 0 \\ 0 & \varepsilon^{-i} \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ (-1)^i & 0 \end{pmatrix} \quad \text{for } i = 0, \dots, n.$$

This is irreducible for $0 < i < n$, and splits into 2 eigenlines when $i = 0$ or n . The inclusion $G \subset \text{SL}(2, \mathbb{C})$ provides the *given* representation $Q = V_1$. It is a simple exercise [Homework] to write down the action of G on a basis $\{e_i \otimes e'_j\}$ of $Q \otimes V_i$ to get $V_i \otimes Q = V_{i-1} \oplus V_{i+1}$ for $0 < i < n$, so that the McKay graph of BD_{4n} is the extended Dynkin diagram \tilde{D}_{n+2} :

(1.2)

Here $\mathbf{1}$ is the trivial 1-dimensional representation.

This example, and the other $\text{SL}(2, \mathbb{C})$ cases observed by McKay, suggest that there is a one-to-one correspondence between the components of the exceptional locus of $Y \rightarrow X$ in (1.1) and the nontrivial irreducible representations of $G \subset \text{SL}(2, \mathbb{C})$ in (1.2). This talk explains this coincidence in several different ways, and discusses higher dimensional generalisations.

1.2. General assumption

I use the following diagram throughout:

(1.3)
$$\begin{array}{ccc} & M & \\ & \downarrow \pi & \\ Y & \xrightarrow{\varphi} & X = M/G \end{array}$$

Here M is a quasiprojective algebraic manifold with $K_M = 0$ and G a finite automorphism group of M that acts trivially on a global basis $s_M \in H^0(K_M)$. The object of study is the quotient variety $X = M/G$ and its resolutions $Y \rightarrow X$, sometimes assumed to have

$K_Y = 0$. An important motivating case is a finite subgroup $G \subset \mathrm{SL}(3, \mathbb{C})$ acting on $M = \mathbb{C}^3$.

1.3. Definition–Reassurance

The quotient varieties $X = M/G$ occurring here are singular. The theory of minimal models of higher dimensional algebraic varieties (Mori theory) has a whole battery of definitions that deal systematically with singular varieties; here I only need one small item: the orbifolds X here have trivial canonical class $K_X = 0$ (or trivial Serre–Grothendieck dualising sheaf $\omega_X = \mathcal{O}_X$). In concrete terms, this means the following: X is a complex n -fold (algebraic or analytic variety), nonsingular in codimension 1, and its nonsingular locus $\mathrm{NonSing} X$ has an everywhere nondegenerate holomorphic n -form s_X (deduced from s_M). So s_X is a complex volume element at every nonsingular point of X , or in other words, it is a global basis of $\Omega_{\mathrm{NonSing} X}^n$. A resolution of singularities $\varphi: Y \rightarrow X$ is *crepant* if $K_Y = \varphi^*K_X$ or $\omega_Y = \varphi^*\omega_X$, which simply means that Y is a nonsingular n -fold with $K_Y = 0$ or $\omega_Y = \Omega_Y^n = \mathcal{O}_Y \cdot s_Y$, where $s_Y = \varphi^*s_X$. More generally, an arbitrary proper birational map $\varphi: V \rightarrow X$ has a *discrepancy divisor* $\Delta_\varphi = \sum a_i E_i$ defined by $K_V = \varphi^*K_X + \sum a_i E_i$ with $a_i \geq 0$; a divisor E_i is *crepant* if $a_i = 0$. The discrepancy Δ_φ is the divisor of zeros on V of the basic n -form s_X on X , generalising the divisor of zeros of the Jacobian determinant; in Mori theory, it measures how far V is from minimal.

1.4. Summary and slogan

I start with a preview of different approaches to the McKay correspondence, which are treated in more detail in later sections. Each of these approaches gives a result in the case of a finite subgroup $G \subset \mathrm{SL}(3, \mathbb{C})$ acting on $M = \mathbb{C}^3$.

- (1) Gonzalez-Sprinberg and Verdier sheaves: the first direct link from the representation theory of G to the geometry of the resolution $Y \rightarrow X$ was the work of Gonzalez-Sprinberg and Verdier [GSpV]: for a Kleinian subgroup $G \subset \mathrm{SL}(2, \mathbb{C})$, they constructed *sheaves \mathcal{F}_ρ on Y , indexed by the irreducible representations of G , whose first Chern classes base the cohomology of Y* .
- (2) String theory: the first hint of a McKay correspondence in higher dimensions comes from work of the string theorists Dixon, Harvey, Vafa and Witten [DHVW] around 1985: if $G \subset \mathrm{SL}(3, \mathbb{C})$ and $Y \rightarrow X = \mathbb{C}^3/G$ is a crepant resolution of the quotient \mathbb{C}^3/G , *the Euler number of Y equals the number of conjugacy classes of G (or the number of its irreducible representations)*.
- (3) Explicit methods: the finite subgroups $G \subset \mathrm{SL}(3, \mathbb{C})$ are classified, and work in the early 1990s of Roan, Ito, Markushevich and others proved case-by-case *the existence of crepant resolutions, and the validity of the formula of [DHVW] for the Betti numbers of Y* .

- (4) Valuation theory: for a finite subgroup $G \subset \mathrm{SL}(n, \mathbb{C})$, the paper [IR] shows that G has a grading by *age*, analogous to the weight grading in Hodge theory, and proves that *the conjugacy classes of junior elements $g \in G$ (elements of age 1) correspond one-to-one with the crepant divisors of a resolution* (more precisely, their discrete valuations). This result holds for any $G \subset \mathrm{SL}(n, \mathbb{C})$ and is intrinsic, classification-free; but for $n \geq 4$ it only gives a small part of a McKay correspondence (so far).
- (5) Nakamura’s G -Hilbert scheme: a resolution of singularities $Y \rightarrow X$, even if it is a Mori minimal model theory, is not at all unique. Moreover, if $X = M/G$, the construction of a resolution Y need not have much to do with G . In 1995, Nakamura made the revolutionary suggestion that in many interesting cases, *the G -Hilbert scheme is a preferred resolution Y of X* (see [IN2], [N], [R]). When this holds, Y is a “very good” moduli space over M , and the general yoga of moduli suggests that there should be a “tautological” treatment of the geometry of Y (comparable to the cohomology of Grassmann varieties).
- (6) Fourier–Mukai transform: the derived category $D(V)$ of coherent sheaves on a variety V (considered up to isomorphism of triangulated categories) can be used as a geometric characteristic of V , in place of K theory or cohomology. The Fourier–Mukai transform is a general method for constructing isomorphisms of derived categories (see [Mu], [O], [BO1], [Br], [BrM]). Bridgeland and others [BKR] have recently used this technique to prove that, if $Y = G\text{-Hilb } M$ is a crepant resolution, then $D^G(M) = D(Y)$. This implies the corresponding result in K theory.
- (7) Motivic integration: the motivic integration of Batyrev, Denef and Loeser, and Kontsevich is a rigorous and comparatively simple mathematical trick that mimics some aspects of the path integrals of QFT. Very roughly, if $\varphi: Y \rightarrow X$ is a resolution of singularities, possibly far from minimal, with discrepancy divisor $K_Y - \varphi^*K_X = \sum a_i E_i$, the calculation amounts to defining the *stringy homology* of X by picking only $\frac{1}{a_i+1}$ th of the homology of E_i . Quite remarkably, *this is well defined, agrees with the predictions of [DHWV] mentioned in (2) above, and provides an exact form of the homological McKay correspondence for finite subgroups $G \subset \mathrm{SL}(n, \mathbb{C})$.*
- (8) Explicit methods (bis): for a finite group $G \subset \mathrm{SL}(3, \mathbb{C})$, the results of (6) (maybe also (7)) imply that Gonzalez-Sprinberg–Verdier sheaves \mathcal{F}_ρ base the K theory of the resolution $Y \rightarrow X$, so that their Chern classes or Chern characters base the cohomology. Reworking this in explicit terms presents a treasure chest of delightful computational problems – already the Abelian cases lead to lovely pictures (compare [R], [CR], [C2]).

I believe that many other approaches to the McKay correspondence remain to be discovered, and many interrelations between the different approaches; this problem area is

recommended to aficionados of noncommutative geometry, perverse sheaves, Gromov–Witten invariants, elliptic cohomology, Chow groups, etc. Here is an attempt to describe the subject in a single statement:

PRINCIPLE 1.1. — *Let M be an algebraic manifold, G a group of automorphisms of M , and $Y \rightarrow X$ a resolution of singularities of X . Then the answer to any well posed question about the geometry of Y is the G -equivariant geometry of M .*

I give two illustrations

- I. If $G \subset \mathrm{SL}(n, \mathbb{C})$ acts on \mathbb{C}^n and the quotient $X = \mathbb{C}^n/G$ has a crepant resolution $Y \rightarrow X$, the homology or K theory of Y is expected (or known) to be independent of Y . In this case, the principle says that the homology or K theory of Y is the representation theory of G (equal to the G -equivariant geometry of \mathbb{C}^n because \mathbb{C}^n is contractible).
- II. Let M be a Calabi–Yau n -fold and G a group of automorphisms of M that acts trivially on Ω_M^n . The stringy homology of $X = M/G$ (see Sections 3 and 4) is well defined by [DL1]. The principle says that it must agree with the G -equivariant homology of M . (I expand on what this means in Section 4.)

Viewed as an orbifold or stack, $X = M/G$ contains M and the G action, and you can of course derive tautological question-and-answer pairs from this (this is often popular as a source of questions after the talk). The content of my slogan is that *the equivariant geometry of M already knows about the crepant resolution $Y \rightarrow X$* . Minimal models exist for surfaces by classical work, and for 3-folds by Mori theory (or by explicit methods). Minimal models of orbifolds by finite subgroups $G \subset \mathrm{SL}(3, \mathbb{C})$ provide infinitely many examples of local models of Calabi–Yau 3-folds; calculating their Betti numbers or K theory in a priori terms is in no sense a tautology. If you prefer to think of the singular X as the fundamental object, and not resolve it (a perfectly sensible alternative), the content is that *X has invariants that can be described from the orbifold M/G , but are birationally invariant* under appropriate conventions about resolutions.

2. AGE AND DISCREPANCY

Let $G \subset \mathrm{SL}(n, \mathbb{C})$ be a finite group; any element $g \in G$ has finite order r , say. For any such r , I choose at the outset a primitive r th root of 1, say $\exp \frac{2\pi i}{r}$. A choice of eigenbasis diagonalises the action of $g \in G$ on $M = \mathbb{C}^n$, giving

$$(2.1) \quad g = \mathrm{diag}(\varepsilon^{a_1}, \dots, \varepsilon^{a_n}) \quad \text{with } 0 \leq a_i < r.$$

I write $g = \frac{1}{r}(a_1, \dots, a_n)$, possibly depending on the choices made. Now $\sum a_i \equiv 0 \pmod{r}$ because $g \in \mathrm{SL}(n, \mathbb{C})$. Following [IR], define the *age* of g by $\mathrm{age} g = \frac{1}{r} \sum a_i$. As we will see, this is an analog of *weight* in Hodge theory; Denef and Loeser [DL2] refer to it by the

long-winded but not inappropriate term *valuation theoretic weight*. Clearly, age g is an integer in the range $[0, \dots, n-1]$, and only the identity has age 0. The elements of age 1 are *junior*.

Junior elements of G give rise to crepant divisors of a resolution $V \rightarrow M/G$ by the following toric mechanism (for more details, and a picture, see [IR], 2.6–7). For $g \in G$ (not the identity), consider the a_i of (2.1), and suppose $(a_1, \dots, a_n) \in \mathbb{Z}^n$ is primitive. The coordinate subspace corresponding to the x_i with $a_i = 0$ is the fixed locus $\text{Fix } g$; it splits off as a direct product, and I assume that all $a_i > 0$ to short-cut some simple arguments. A useful example to bear in mind is when all the $a_i = 1$ (compare Example 4.1).

I view the integers (a_1, \dots, a_n) as *weights*. They define the grading $\text{wt } x_i = a_i$ on the coordinate ring $\mathbb{C}[x_1, \dots, x_n]$, or equivalently, the action $x_i \mapsto \lambda^{a_i} x_i$ of \mathbb{C}^* on $M = \mathbb{C}^n$ that defines the weighted projective space

$$\mathbb{P}(a_1, \dots, a_n) = (\mathbb{C}^n \setminus 0)/\mathbb{C}^*.$$

We obtain the weighted blowup $B_g \rightarrow M$ as the closed graph of the quotient map $M \dashrightarrow \mathbb{P}(a_1, \dots, a_n)$; it has the exceptional divisor $B_g \supset E_g = \mathbb{P}(a_1, \dots, a_n)$. Obviously g acts on B_g , and fixes E_g pointwise (because g acts on M as $\varepsilon \in \mathbb{C}^*$). Therefore $B_g \rightarrow B_g/\langle g \rangle$ is totally ramified along E_g .

THEOREM 2.1 ([IR], 2.6–7). — *Suppose that $V \rightarrow X$ is any resolution of singularities of the quotient $X = M/G$. Then V contains a divisor F_g rationally dominated by E_g under the rational map $B_g \rightarrow M \dashrightarrow V$. The discrepancy of F_g is given by $a_{F_g} = \text{age } g - 1$, and in particular*

$$F_g \text{ is crepant} \iff g \text{ is junior.}$$

Every crepant divisor of any resolution V occurs thus.

Discussion of proof. — Write $X_g = M/\langle g \rangle$ for the partial quotient. Then $B_g/\langle g \rangle \rightarrow X_g$ is a partial resolution, with the single exceptional divisor E_g . An easy toric calculation gives the discrepancy of $E_g \subset B_g$ or $E_g \subset B_g/\langle g \rangle$ (compare [YPG], 4.8): the standard basis of Ω_M^n is $s_M = dx_1 \wedge \dots \wedge dx_n$. For K_{B_g} , choose a Laurent monomial $y_1 = x^{\mathbf{m}}$ of weight 1 (recall that the a_i were coprime). Then y_1 is the defining equation of $E_g \subset B_g$ at a general point of E_g (away from all the coordinate hyperplanes), and y_1^r that of $E_g \subset B_g/\langle g \rangle$. Choosing Laurent monomials y_2, \dots, y_n forming a basis of the lattice of monomials of weight 0, we get that

$$s_{B_g} = dy_1 \wedge \dots \wedge dy_n \in \Omega_{B_g}^n$$

is the required basis. The discrepancy is the exponent a in $s_M = (\text{unit}) \cdot y_1^a s_{B_g}$, and is determined by weighty considerations: s_M has weight $\sum a_i$ and s_{B_g} weight 1, so $a = \sum a_i - 1$. On the quotient $B_g/\langle g \rangle$ we only have y_1^r , so we get the stated discrepancy $\frac{1}{r}(\sum a_i - 1 - (r-1)) = \text{age } g - 1$.

The quotient morphism $M \rightarrow X$ is a Galois cover with group G ; a cyclic subgroup $\langle g \rangle$ corresponds to an intermediate cover $M \rightarrow M/\langle g \rangle = X_g \rightarrow X$. The reduction to a cyclic group is in terms of ramification theory; see [IR], 2.6–7. Roughly, over the general point of any exceptional divisor F of $V \rightarrow X$, the Galois extension of function fields $k(X) \subset k(M)$ forms a tower, starting with a cyclic ramified cover. For a crepant exceptional divisor, the cyclic ramification can be chased back up to a conjugacy class of junior elements $g \in G$.

REMARK 2.2. — This argument works in all dimensions, but it only identifies the *divisors* of a crepant resolution Y , and thus only gives a basis of $H^2(Y, \mathbb{Q})$ or $H_{2n-2}(Y, \mathbb{Z})$ corresponding in McKay style to junior conjugacy classes of G . In 3 dimensions, we used Poincaré duality to bootstrap ourselves up to a basis of $H^*(Y, \mathbb{Q})$ in [IR]. Historically, this was the first intrinsic proof of the conjectured formula of [DHVW] for the Betti numbers of a crepant resolution.

As Brylinski [B] remarks (following Mumford), if $V \rightarrow X$ is any resolution, the group G can be viewed as the fundamental group of V minus the branch locus, so that an exceptional divisor F of a resolution V corresponds directly to a conjugacy class of G as a little anticlockwise loop around F ; for crepant divisors, this is of course the same relation as in [IR]. But I don't know how to use this idea to get a well defined relation between, say, codimension 2 cycles of Y and age 2 conjugacy classes of G .

3. L'INNOMABLE

This section is mainly for sociological and historical interest, but some harmless hilarity may derive from my garrulous display of incompetence and ignorance in physics.

A theoretical prediction of string theory: Fermionic strings propagate in 10-dimensional space-time. Indeed, a universe of any other dimension would have particles moving faster than the speed of light. Since this prediction, on the face of it, contradicts the empirically observed 4-dimensions of space-time, string theorists want 6 of the dimensions to be filled up with tiny Calabi–Yau 3-folds. (This means (i) a 6-dimensional Riemannian manifold with $SU(3)$ holonomy, or (ii) a complex manifold V with a Ricci flat Kähler metric and $H^1(V, \mathbb{R}) = 0$, or (iii) an algebraic manifold V with $K_V = 0$ and $H^1(V, \mathcal{O}_V) = 0$. It seems that the holonomy or Kähler conditions on V , together with some finite volume, are required by the physics, whereas making V nonsingular, compact, and a constant fibre over macroscopic space-time are just convenient choices when you try to guess a model.)

The two papers [DHVW] were concerned with trying to calculate string theory on examples of Calabi–Yau varieties obtained by dividing a 3-dimensional complex torus M by a finite group G preserving a basic holomorphic 3-form, so that the stabiliser subgroup at any point is a subgroup of $SL(3, \mathbb{C})$. A closed string on the quotient may lift either to a closed string on the cover, or to a path that goes from x to $g \cdot x$. The latter are called *twisted sectors*. The physicists need to take care of these in order to relate \int_X

to G -equivariant \int_M , and they are the key to the form of the McKay correspondence in Theorem 4.4, (4).

Taking limits is a tradition in physics, where the old is frequently the limit of the new: Newtonian mechanics is the limit of special relativity as $c \rightarrow \infty$, classical mechanics the limit of quantum physics as $\hbar \rightarrow 0$, groups and their Hopf algebras the limit of quantum groups as $q \rightarrow 1$. In string theory, if the scale (or radius of curvature) of the tiny Calabi–Yau tends to zero, the theory should approximate ordinary Lorentz 4-dimensional space-time, whereas letting it tend to macroscopic proportions would approximate flat Lorentz 10-dimensional space-time. In this context, the twisted sector near a point $x \in M^H$ plays the role of strings that are topologically nontrivial, but are allowed to remain of finite length (and so contribute to path integrals) as the scale becomes large. To calculate something called the *1-loop partition function*, DHVW considered mapping the elliptic curve $S^1 \times S^1$ (with parameters σ and τ along the copies of S^1) into X , or the σ, τ square into M with equivariant boundary conditions depending on g, h . Thinking about twisted sectors and limits led DHVW (I confess that their logic eludes me somewhat) to the formula

$$(3.1) \quad e_{\text{string}}(X) = e(M, G) := \frac{1}{|G|} \sum_{\substack{g, h \in G \\ \text{commuting}}} e(M^{(g, h)}).$$

Here $e(M, G)$ on the left-hand side is the G -equivariant Euler number of M ; on the right-hand side, the sum runs over all commuting pairs of elements of G , $\langle g, h \rangle$ is the Abelian group they generate, $M^{(g, h)}$ its fixed locus in M , and e is the usual Euler number. The formula is a replacement for the Euler number of the singular orbifold X . The papers [DHVW] contain more-or-less explicitly the conjecture that this number is the Euler number of a minimal resolution of singularities.

It is not hard (see [HH], [Roan] and [Homework]) to rearrange the sums in (3.1) to give

$$(3.2) \quad e_{\text{string}}(X) = e(M, G) = \sum_{[H] \subset G} e(X^H) \times \text{card}\{[h] \in H\},$$

where (i) the first sum runs over conjugacy classes of subgroups $H \subset G$; (ii) the stratum X^H is the set of $x \in X$ such that Stab_y is conjugate to H for any point $y \in M$ over x ; (iii) the second factor is the number of conjugacy classes in H . This means that $X^H \subset X$ contributes to $e(M, G)$ with multiplicity the representation theory of H .

REMARK 3.1. — The physicists want to do path integrals, that is, they want to integrate some “Action Man functional” over the space of all paths or loops $\gamma: [0, 1] \rightarrow Y$. This impossibly large integral is one of the major schisms between math and fizz. The physicists learn a number of computations in finite terms that approximate their path integrals, and when sufficiently skilled and imaginative, can use these to derive marvellous consequences; whereas the mathematicians give up on making sense of the space of paths, and not infrequently derive satisfaction or a misplaced sense of superiority from pointing out that

the physicists' calculations can equally well be used (or abused!) to prove $0 = 1$. Maybe it's time some of us also evolved some skill and imagination. The motivic integration treated in the next section builds a miniature model of the physicists' path integral, by restricting first to germs of holomorphic paths $\gamma: U \rightarrow Y$, where $0 \in U \subset \mathbb{C}$ is a neighbourhood of 0, then to formal power series $\gamma: \text{Spec } \mathbb{C}[[z]] \rightarrow Y$.

4. MOTIVIC INTEGRATION

The material in this section is due to Batyrev [Ba1], [Ba2], Denef and Loeser [DL1], [DL2] and Kontsevich [K]. I recommend Craw [C1] as a readable first introduction to these ideas.

Rather than trying to restrict to crepant resolutions, take an arbitrary normal crossing resolution $\varphi: Y \rightarrow X$, marked by the discrepancy divisor $D = \Delta_\varphi = \sum_{i \in I} a_i D_i$ (here I is the indexing set of the components D_i). The normal crossing divisor D defines a stratification of Y , with

$$\text{closed strata } D_J = \bigcap_{j \in J} D_j, \quad \text{and} \quad \text{open strata } D_J^\circ = D_J \setminus \bigcup_{J' \supsetneq J} D_{J'}$$

for $J \subset I$ (including, of course, $Y = D_\emptyset$ and $Y \setminus D = D_\emptyset^\circ$).

Motivic integration is discussed and defined below, but it is convenient to start from the answer: the *stringy motive* of (Y, D) , or of X itself, turns out to be

$$(4.1) \quad h_{\text{string}}(X) = h(Y, D) = \sum_{J \subset I} [D_J^\circ] \cdot \prod_{j \in J} \frac{\mathbb{L} - 1}{\mathbb{L}^{a_j+1} - 1}.$$

Here $\mathbb{L} = [\mathbb{A}_{\mathbb{C}}^1] = [\mathbb{C}]$ is the Tate motive, and the formula takes place in a certain ring of motives with formal power series in \mathbb{L}^{-1} adjoined. We will worry about the coefficient ring later, but in lucky cases it will happen that the cyclotomic polynomials in the denominators cancel out, leaving an integral motive (see Example 4.1 and [Homework] for examples). It follows from Theorem 4.4, (2) and (3) that $h(Y, D)$ is independent of the choice of the normal crossing resolution Y , so depends only on X . In the case when $D = aE$ has a single component with discrepancy a , it boils down to

$$(4.2) \quad [Y - E] + \frac{[E]}{1 + \mathbb{L} + \mathbb{L}^2 + \dots + \mathbb{L}^a} = [Y - E] + \frac{[E]}{[\mathbb{P}^a]}.$$

EXAMPLE 4.1. — Let $n = ab$, and consider the n -fold quotient singularity X of type $\frac{1}{b}(1, \dots, 1)$, that is, the quotient $\mathbb{C}^n/(\mathbb{Z}/b)$, with the diagonal action of $\varepsilon = \exp \frac{2\pi i}{b}$. It is the cone over the b th Veronese embedding of \mathbb{P}^{n-1} , so that its resolution $Y \rightarrow X$ has exceptional divisor $E = \mathbb{P}^{n-1}$ with $\mathcal{O}_E(E) = \mathcal{O}_{\mathbb{P}^{n-1}}(-b)$. The discrepancy is $a - 1$, to fit the adjunction formula, with $K_Y = (a - 1)E$, and $K_E = \mathcal{O}_E(aE) = \mathcal{O}_{\mathbb{P}^{n-1}}(-n)$.

Now whereas Y is homotopy equivalent to \mathbb{P}^{n-1} , so has n homology classes, one in each dimension $0, 2, \dots, 2(n - 1)$, the effect of dividing by $[\mathbb{P}^{a-1}]$ in (4.2) is to throw away most

of these, leaving only the b stringy homology classes in dimension $0, 2a, 4a, 2a(b-1)$. This is exactly what we need for the McKay correspondence: the b elements of \mathbb{Z}/b have age $0, a, 2a, \dots, a(b-1)$ and correspond to the stringy classes in dimension $2ia$.

EXAMPLE 4.2. — Consider the blowup $\sigma: Y_1 \rightarrow Y$ of a subvariety $C \subset Y$ that intersects all the strata of D transversally, and set $D_1 = \sigma^*D + (c-1)E$, where $E = \sigma^{-1}C$ is the exceptional divisor of the blowup and $c = \text{codim } C$. The coefficient is the discrepancy of E , so that $K_{Y_1} - D_1 = \sigma^*(K_Y - D)$. It is an exercise to see that

$$h(Y, D) = h(Y_1, D_1).$$

(This is rather trivial if $C \cap D = \emptyset$ in view of Grothendieck's *formule clef* for the motive of a blowup; see [Homework] for more hints.) This is good evidence for the birational invariance of $h(Y, D)$.

I now describe briefly the mechanics of motivic integration, following [C1]. Start from the Grothendieck ring $K_0(\mathcal{V})$ of classes of varieties under the equivalence relation $[V] = [V \setminus W] + [W]$. Addition and multiplication are quite harmless. The Tate motive is $\mathbb{L} = [\mathbb{A}_{\mathbb{C}}^1] = [\mathbb{C}]$. We formally adjoin \mathbb{L}^{-1} to $K_0(\mathcal{V})$, and make a fairly mild (\mathbb{L}^{-1}) -adic completion to give the value ring $R = \widehat{K}_0(\mathcal{V})[\mathbb{L}^{-1}]$. This value ring is the really clever thing about the whole construction. (Exercise: $(\mathbb{L}^a - 1)^{-1}$ can be written as a formal power series in \mathbb{L}^{-1} , so all the terms on the right-hand side of (4.1) are in R .)

Motivic integration takes place over the infinite jet space $J_{\infty}Y$, which coincides with the set $Y(\mathbb{C}[[z]])$ of points of Y with values in the formal power series ring $\mathbb{C}[[z]]$. An element $\gamma \in Y(\mathbb{C}[[z]])$ is a point $y = \gamma(0) \in Y$ together with a formal arc $\gamma: \text{Spec } \mathbb{C}[[z]] \rightarrow Y$ starting at y ; if convergent, γ is the Taylor series of a holomorphic germ $\tilde{\gamma}: (\mathbb{C}, 0) \rightarrow Y$. The infinite jet space $J_{\infty}Y$ is the profinite limit $\varprojlim_k J_k Y$ of the finite jet spaces $J_k Y$; recall that $J_0 Y = Y$, $J_1 Y$ is the total space of the tangent bundle T_Y , and $J_{k+1} \rightarrow J_k$ is a \mathbb{C}^n -fibre bundle.

The projection maps $\pi_k: J_{\infty}Y \rightarrow J_k$ of the profinite limit allow us to define a *cylinder set* in $J_{\infty}Y$ to be $\pi_k^{-1}(B_k)$ for a constructible set $B_k \subset J_k$. The measure on $J_{\infty}Y$ is initially defined on these, by setting¹

$$(4.3) \quad \mu(\pi_k^{-1}(B_k)) := [B_k] \cdot \mathbb{L}^{-nk} \in R.$$

It is straightforward to see that this is independent of k , and is a “finitely additive measure”.

As our measurable functions, consider an effective divisor D on Y , and define a function $F_D: J_{\infty}Y \rightarrow \mathbb{Z}_{\geq 0}$ by $F_D(\gamma) = D \cdot \gamma$ (intersection number). In other words, suppose $\gamma(0) = P \in Y$ and let g_D be the local defining equation of D at P ; then $F_D(\gamma)$ is the

¹The papers [DL1] and [C1] have the exponent $\mathbb{L}^{-n(k+1)}$. This is just a normalising convention, giving $h(Y, D) = [Y] \cdot \mathbb{L}^{-n}$ in Theorem 4.4, (1), and making the motive of Y 0-dimensional. I prefer my version.

order of $\gamma^*(g_D) \in \mathbb{C}[[z]]$. Since the first s coefficients of $\gamma^*(g_D)$ clearly only depend on $\pi_s(\gamma) \in J_s$, it is obvious that $F_D^{-1}(s)$ is a cylinder set.

The grand definition is now: for Y a nonsingular variety and D a normal crossing divisor, the motivic integral is

$$(4.4) \quad h(Y, D) = \int_{J_\infty Y} \mathbb{L}^{-F_D} := \sum_{s \in \mathbb{Z}_{\geq 0}} \mu(F_D^{-1}(s)) \cdot \mathbb{L}^{-s} \in R.$$

REMARK 4.3. — I omit some tricky details on convergence required to get a genuine measure (involving the $[\mathbb{L}^{-1}]$ -adic completion). To tell the truth, I don't know if they are at all essential. A basic point for applications is that the measure of $F_D^{-1}(s)$ tends to 0 as $s \rightarrow \infty$; this is plausible enough (because arcs γ with $\gamma \cdot D \geq s$ have codimension $\geq s$ in $J_\infty Y$), and is an intuitive reason behind birational invariance: the arcs in a Zariski closed subset of Y have measure zero.

THEOREM 4.4. — $h(Y, D)$ of (4.4) has the following properties:

- (1) If $D = 0$ then $h(Y, D) = [Y]$.
- (2) $h(Y, D)$ is calculated by the right-hand side of (4.1).
- (3) Birational invariance: let Y', D' and Y, D be pairs, and $\varphi: Y' \rightarrow Y$ a birational morphism such that $K_{Y'} - D' = \varphi^*(K_Y - D)$; then

$$h(Y', D') = h(Y, D).$$

- (4) If $X = M/G$ is as in Assumption 1.2, $Y \rightarrow X$ a normal crossing resolution, and D the discrepancy, then

$$(4.5) \quad h_{\text{string}}(X) = h(Y, D) = \sum_{[H] \subset G} [X^H] \cdot \sum_{[g] \in H} \mathbb{L}^{\text{age } g},$$

where the range of summation is as in (3.2), and the second sum is over conjugacy classes in H .

Discussion of proof. — I give some indications, leaving most of the proof as references to [DL1] and [DL2]. Alternatively, do them as exercises (see [Homework] for more hints). The key point of the proof is that, whatever its substance, (4.4) has the formal properties of an integral, and is subject to the same kind of *change of variables* formula. In the words of the Master:

“La théorie consiste pour l'essentiel dans des questions de variance”

([H], Introduction). Note first that the condition in (3) says that $D' - D = \text{div}(\text{Jac } \varphi)$ is the divisor of zeros of the Jacobian determinant of φ (I omit φ^* from now on). Composition defines a map $j_\varphi: J_\infty Y' \rightarrow J_\infty Y$, and, unless it falls entirely in the locus of indeterminacy of φ^{-1} , an arc in Y has a birational transform as an arc in Y' ; in other words, away from subsets of measure zero, j_φ is a bijection on the infinite jet spaces. For (3), it remains only to stratify the finite jet spaces $J_k Y'$ and $J_k Y$ so that the corresponding morphism

$j_k: J_k Y' \rightarrow J_k Y$ is a \mathbb{C}^t -bundle on each stratum with $F_{D'-D}(\gamma) = \text{div}(\text{Jac } \varphi) \cdot \gamma = t$ (see [DL1], Lemma 3.4 and [Homework]).

(2) is proved in [DL1], Proposition 6.3.2, [Ba2], Theorem 6.28, and worked out in detail in [C1], Theorem 1.16. The proof of (4) consists of two steps, relating to the two morphisms $\pi: M \rightarrow X$ and $\varphi: Y \rightarrow X$ of Assumption 1.2.

Step I. — We translate the twisted sectors of [DHVW] into the language of formal arcs, obtaining the stratification (4.6) below.

Let $y \in M^H$ be a point with $\text{Stab } y = H$ and $x = \pi(y) \in X^H$. As at the start of Section 2, suppose that r is an integer divisible by the order of each $g \in H$, and choose an r th root ε of 1 and an r th root $\zeta = z^{1/r}$ of the parameter used for formal arc, so that a formal arc γ at $x \in X$ parametrised by z lifts to a formal arc at $y \in M$ parametrised by ζ . Unless γ falls entirely in the branch locus of $\pi: M \rightarrow X$, there is a unique conjugacy class $g \in H$ defined by $\gamma(\varepsilon\zeta) = g\gamma(\zeta)$. Here g is the *twisted sector*, the conjugacy class of γ in the local fundamental group H (where γ is viewed as a little loop in X minus the branch locus).

This argument shows that, after we delete the subset of arcs falling entirely in the branch locus (which has infinite codimension, so measure zero) the infinite jet space $J_\infty X$ is a disjoint union

$$(4.6) \quad J_\infty X = \coprod_{[H] \subset G} \coprod_{[g] \in H} J_\infty^{H,g} Y,$$

where H, g are as in (3.2), and $J_\infty^{H,g} Y$ is the set of arcs with $\gamma(0) \in X^H$ in the twisted sector g .

Step II. — Using change of variables as in the proof of (3), one calculates that $J_\infty^{H,g} Y$ contributes $X^H \cdot L^{g^{-1}}$ to $h(Y, D)$ ([DL2], Lemma 4.3). The difference in appearance of the formulas here and in [DL2] are explained by two trivial shifts of notation: as explained in the footnote on page 10, my measure is \mathbb{L}^n times theirs; and they diagonalise g as ε^{e_i} with $1 \leq e_i \leq r$, defining $w(g) = \frac{1}{r} \sum e_i = n - \text{age}(g^{-1})$.

REMARK 4.5. — Statement (4) is an exact analogue of the [DHVW] formula (3.2), saying that the stratum X^H appears in the stringy homology of Y multiplied by the set of conjugacy classes in H .

As discussed in Definition 1.3, the discrepancy $D = \text{div } s_X$ is the divisor of zeros of s_X , the global basis of $\Omega_{\text{NonSing } X}^n$. In the normal course of events, integrating *functions* on Y requires a volume form; here we take s_X as a holomorphic volume form, viewing its zeros on D as scaling down the contribution from neighbourhood of the discrepant exceptional divisors. This is what produces a birationally invariant answer.

5. HILBERT SCHEMES OF G -ORBITS

This section explains the definition of the G -orbit Hilbert scheme $G\text{-Hilb } M$, and Nakamura's idea of using it to resolve certain quotient singularities. We know by general results (especially Hironaka's theorems) that the singularities of a quotient variety $X = M/G$ can be resolved somehow-or-other, but the construction of an actual resolution is messy, involves lots of choices, and will probably have almost nothing to do with the group action. Around 1995, Ito and Nakamura observed that in the case of $G \subset \mathrm{SL}(2, \mathbb{C})$, *the Hilbert scheme $G\text{-Hilb } \mathbb{C}^2$ of G -clusters is a crepant resolution of the quotient \mathbb{C}^2/G* . Nakamura conjectured that this continues to hold for $G \subset \mathrm{SL}(3, \mathbb{C})$, and this has since been confirmed and extended to some other cases by work of Bridgeland and others (see [BKR] and Theorem 6.1).

First, a *cluster* in a variety M (say, quasiprojective and nonsingular) is a 0-dimensional subscheme $Z \subset M$, defined by an ideal $\mathcal{I}_Z \subset \mathcal{O}_M$, so that the cokernel $\mathcal{O}_Z = \mathcal{O}_M/\mathcal{I}_Z$ is a finite dimensional \mathbb{C} -vector space. The *degree* of Z is the dimension of \mathcal{O}_Z . Like the intersection of two plane curves in Bezout's theorem, a cluster Z may consist of reduced points $Z = P_1 + \cdots + P_N$, or may have a nonreduced structure; in the latter case, we *keep track of the ideal* $\mathcal{I}_Z \subset \mathcal{O}_M$, as a way of using algebraic equations to keep information about the relative positions when some of the points P_i come together. For example,

$$(x^2, xy, y^2) \quad \text{and} \quad (x - ay - by^2, y^3) \quad \text{for any } a, b \in \mathbb{C}$$

are clusters of degree 3 supported at $0 \in \mathbb{C}^2$.

LEMMA 5.1. — *All clusters $Z \subset M$ of given degree N in M are parametrised by a quasiprojective scheme $\mathrm{Hilb}^N M$, which is a fine moduli space.*

Proof. — The assertion is quite elementary. M is quasiprojective; choose an embedding $M \subset \mathbb{P}^s$. Every ideal $\mathcal{I}_Z \subset \mathcal{O}_M$ of codimension N defines and is defined by a codimension N vector subspace

$$H^0(\mathbb{P}^s, \mathcal{I}_Z(N)) \subset H^0(\mathbb{P}^s, \mathcal{O}_{\mathbb{P}^s}(N)),$$

the forms of degree N vanishing on Z (same N). Subspaces of given codimension are parametrised by a Grassmann variety, and the condition that a space of forms defines a cluster of degree N in M is a locally closed condition. (It can be written in terms of rank of a matrix = N .) □

REMARK 5.2. — The map $\mathrm{Hilb}^N M \rightarrow S^N M$ to the symmetric product, defined at the level of sets by $Z \mapsto \mathrm{Supp } Z$, is a morphism of schemes, the Hilbert–Chow morphism (see [GIT], Chapter 5, §4). For a curve, $\mathrm{Hilb}^N C$ is just the symmetric product $S^N C$, which is itself already nonsingular. For a surface, the symmetric product $S^N S$ is singular at the diagonals, and $\mathrm{Hilb}^N S \rightarrow S^N S$ is a crepant resolution, in fact, a symplectic resolution;

see [IN2], §6. But $\text{Hilb}^N M$ is singular as soon as $\dim M \geq 3$ and $N = \deg Z \geq 4$, and usually even has components of excess dimension.

PROPOSITION–DEFINITION 5.3 (Ito and Nakamura). — *Let G be a finite group of order N acting faithfully on an algebraic manifold M ; consider the action of G on $\text{Hilb}^N M$ and its fixed locus $(\text{Hilb}^N M)^G$. This has a unique irreducible component that contains a general orbit $G \cdot y$ of G on M . This component is defined to be the G -Hilbert scheme, and denoted by $G\text{-Hilb } M$. The composite $G\text{-Hilb } M \hookrightarrow \text{Hilb}^N M \rightarrow S^N M$ induces a Hilbert–Chow morphism $G\text{-Hilb } M \rightarrow M/G$ which is proper and birational.*

A cluster $Z \in G\text{-Hilb}$ is G -invariant, and is called a G -cluster; its defining ideal \mathcal{I}_Z is G -invariant, and as a representation of G , the quotient $\mathcal{O}_Z = \mathcal{O}_M/\mathcal{I}_Z$ is the regular representation $\mathbb{C}[G]$.

See also [CR], 4.1 for a rival definition and a comparison between the two.

Proof. — The general orbit $G \cdot y$ consists of N points permuted simply transitively by G , so is a G -invariant cluster in $(\text{Hilb}^N M)^G$. These orbits fill out an irreducible open set in $(\text{Hilb}^N M)^G$, because a small G -invariant deformation of $G \cdot y$ is clearly still a set of N distinct points permuted by G and disjoint from any fixed locus. The closure of this component is $G\text{-Hilb } M$ by definition. The composite $G\text{-Hilb } M \hookrightarrow \text{Hilb}^N M \rightarrow S^N M$ is a morphism; by definition, a dense open set of $G\text{-Hilb } M$ consists of general orbits $G \cdot y$, and these maps to orbits in $S^N M$, that is, to M/G .

Finally, the quotient sheaves \mathcal{O}_Z for $Z \in G\text{-Hilb } M$ fit together as a locally free sheaf \mathcal{O}_Z over $G\text{-Hilb } M$, with a G -action that makes it the regular representation on a dense open set. Its isotypical decomposition under the idempotents of $\mathbb{C}[G]$ is a direct sum, so each component must also vary as a locally free sheaf, therefore $\mathcal{O}_Z \cong \mathbb{C}[G]$ for every $Z \in G\text{-Hilb } M$ (since $G\text{-Hilb } M$ is defined to be irreducible). \square

The G -Hilbert scheme is a crepant resolution for finite groups $G \subset \text{SL}(3, \mathbb{C})$. The general case of this is proved by Bridgeland and others [BKR] using derived category methods and a homological characterisation of regularity. For a diagonal Abelian group, $A\text{-Hilb } \mathbb{C}^3$ is a completely explicit construction of Nakamura (see [N] and [CR]): the monomial xyz is A -invariant, and every G -cluster Z is defined by 7 (possibly redundant) equations of the form

$$\begin{aligned} x^{a+1} &= \lambda y^d z^g & y^{d+1} z^{g+1} &= \alpha x^a \\ y^{b+1} &= \mu z^e x^h & z^{e+1} x^{h+1} &= \beta y^b & \text{and } xyz &= \xi \\ z^{c+1} &= \nu x^f y^i & x^{f+1} y^{i+1} &= \gamma z^c \end{aligned}$$

for appropriate exponents a, \dots, i and coefficients α, \dots, ξ satisfying $\alpha\lambda = \beta\mu = \gamma\nu = \xi$. The monomial basis of \mathcal{O}_Z forms a tripod shaped Newton polygon in the plane lattice \mathbb{Z}^2 of Laurent monomials modulo xyz ; this lattice is naturally the universal cover of the McKay quiver and the tripod is a choice of fundamental domain for the covering group

(see [N] and [R] for pictures). The explicit calculations remain an interesting challenge in the non-Abelian cases, e.g., in the trihedral case.

EXAMPLE 5.4. — These results are known to fail for finite $G \subset \mathrm{SL}(4, \mathbb{C})$. In the first place, most quotient singularities $X = \mathbb{C}^4/G$ do not have any crepant resolution. For example, the series of cyclic quotient singularities $\mathbb{C}^4/(\mathbb{Z}/r)$ of type $\frac{1}{r}(1, r-1, i, r-i)$ have no junior elements, so are terminal; compare Example 4.1. These examples motivated the initial exploration of stringy homology in [BD].

Next, even when a crepant resolution exists, the G -Hilbert scheme may be singular or discrepant or both. A simple example is the quotient singularity \mathbb{C}^4/G by the maximal diagonal subgroup $(\mathbb{Z}/2)^{\oplus 3} \subset \mathrm{SL}(4, \mathbb{C})$ of exponent 2. The junior simplex Δ has all the midpoints of the edges $\frac{1}{2}(1, 1, 0, 0)$ etc., as lattice points. This has several subdivisions into basic simplexes, giving crepant resolutions, but none that is symmetric under permuting the coordinates – the only symmetric thing you can do is chop off the 4 basic simplexes at the corners, leaving a terminal simplex of volume 2. On the other hand, $G\text{-Hilb } \mathbb{C}^4$ is obviously symmetric.

6. COHERENT DERIVED CATEGORY

Grothendieck and Verdier introduced the derived category $D(X)$ of coherent sheaves on a variety X in the 1960s as a technical convenience in homological algebra; it has enjoyed an unfortunate reputation for technicality and abstraction ever since then. Recently, however, it has been increasingly used as a geometric characteristic of X similar to K theory: whereas K theory works with the group of bundles or sheaves modulo the relation $F = F' + F''$ for every short exact sequence $0 \rightarrow F' \rightarrow F \rightarrow F'' \rightarrow 0$, the derived category $D(X)$ consists of complexes F^\bullet modulo the relation of quasi-isomorphism (defined at the start of the theory, and thankfully never referred to again). Following Mukai's pioneering work [Mu] for Abelian varieties, Orlov and Bondal [O], [BO1] have advocated the idea of considering the derived category $D(X)$ (up to isomorphism of triangulated categories) as a geometric characteristic of X . From this point of view, $D(X)$ behaves like an enriched version of K theory.

A variety X with $\pm K_X$ ample can be reconstructed from its derived category $D(X)$ (as a triangulated category) [BO1], but if $K_X = 0$ (notably for an Abelian variety or a K3 surface), the same triangulated category may occur as $D(X)$ for different X , or there may be infinitely many symmetries of $D(X)$ not arising from automorphisms of X . Isomorphisms $D(X) \cong D(Y)$ arise as *Fourier–Mukai transforms* $\Phi_{X \rightarrow Y}^{\mathcal{F}}$ corresponding to a sheaf \mathcal{F} on $X \times Y$, defined as the composite of the functors p_X^* , $\otimes \mathcal{F}$ and q_{Y*} (more precisely, their derived functors); for an up-to-date treatment, see [Br] and the references given there. In practice, Y is most frequently a moduli space of coherent sheaves on X and \mathcal{F} the universal sheaf over $X \times Y$, so that Y parametrises sheaves F_y on X ; in very

good cases, the apparatus of moduli functors, stable bundles, and deformation theory gives essentially for free that the F_y have orthonormality properties under Ext functors (formally analogous to those of trig functions in the theory of Fourier transform).

Let M be a nonsingular quasiprojective n -fold with $K_M = 0$, and G a finite group acting on M , with trivial action on K_M . Set $Y = G\text{-Hilb } M$. Since Y is a fine moduli space for G -clusters $Z \subset M$, there is a universal G -cluster $\mathcal{Z} \subset Y \times M$, fitting in a diagram

$$(6.1) \quad \begin{array}{ccc} \mathcal{Z} & \xrightarrow{q} & M \\ p \downarrow & & \downarrow \pi \\ Y & \xrightarrow{\varphi} & X \end{array}$$

Bridgeland and others [BKR] prove the following theorem.

THEOREM 6.1. — *Suppose that the inverse image of the diagonal $(\varphi \times \varphi)^{-1}(\Delta_X)$ has dimension $\leq n + 1$ (automatic for $n = 3$). Then Y is a crepant resolution of X and the Fourier–Mukai functor $\Phi = \mathbb{R}q_* \circ p^*: D(Y) \rightarrow D^G(M)$ is an equivalence of categories.*

Once we know that Y is a crepant resolution, ω_M is trivial as a G -sheaf and ω_Y is trivial, so that both the derived categories $D^G(M)$ and $D(Y)$ have Serre duality functors; the remainder of the proof is then standard Fourier–Mukai technology. However, the surprising thing here is Bridgeland’s derivation of the nonsingularity of Y from the famous theorem of commutative algebra known for a long time as Serre’s “Intersection conjecture”.

7. FIN DE PARTIE

Samuel Beckett’s play of the same title has the wonderful line:

“Personne au monde n’a jamais pensé aussi tordu que nous.”

This seems to reflect a truth about math research: progress beyond the obvious takes really twisted thinking. In this spirit, let me raise all the open questions I can think of.

There are two basic flavours of McKay correspondence:

- (1) conjugacy classes of $G \leftrightarrow$ homology of Y (or stringy homology); and
- (2) representations of $G \leftrightarrow$ derived category $D(Y)$ or K theory of Y .

Is there a “bivariant” version of the correspondence containing both (1) and (2) at the same time? For example, in some contexts, \mathcal{D} -modules or perverse sheaves manage to accommodate both coherent and topological cohomology. Note that (1) and (2) achieve a well posed question in completely different ways: (1) takes accounts of discrepancy systematically, whereas (2) currently only works under the very strict condition that $Y = G\text{-Hilb}$ is a crepant resolution.

The representation theory of finite groups has two ingredients, conjugacy classes and irreducible representations, and a character table, which is a nonsingular matrix making

them “dual” (I apologise to group theorists for this gratuitous vulgarity). Although in substance very different, the homology and K theory of a variety Y could be described in similar terms. In cases when McKay holds, is there any direct relation?

All the different approaches to McKay described here have one thing in common: none of them seems to say anything very useful about multiplicative structures. The following questions seem most likely to be approachable: can tensor product of G -modules and tensor product in K theory of Y be related? Can you reconstruct the McKay quiver in $D(Y)$ or K_0Y ?

Motivic integration takes a fraction of the homology of a discrepant exceptional divisor, say, half the homology of the exceptional \mathbb{P}^3 for the quotient singularity $\mathbb{C}^4/(\mathbb{Z}/2)$ (the cone on the second Veronese embedding $v_2(\mathbb{P}^3)$). In contrast, half of a derived category is something no-one has ever seen. In the case of $v_2(\mathbb{P}^3)$, the Gonzalez-Sprinberg–Verdier sheaves corresponding to the characters ± 1 are \mathcal{O}_Y and $\mathcal{O}_Y(1)$. Breaking up the derived category $D(Y)$ into two bits, one of which will correspond to the representations of $\mathbb{Z}/2$, doesn’t seem to make any sense. On the other hand, in this case we can extend the action of $\mathbb{Z}/2$ to the action $\frac{1}{4}(1, 1, 1, 1)$ of $\mathbb{Z}/4$, whose quotient does have a crepant resolution.

Another general problem area: resolutions of Gorenstein quotient singularities give a collection of examples of Calabi–Yau 3-folds with very nice properties: the homology of the resolution is well defined (independent of the choice of resolution), and the homology and K theory are closely related by something like a duality. Do these properties hold for Calabi–Yau 3-folds more generally? It seems very likely that birational Calabi–Yau 3-folds have isomorphic derived categories, but so far this only seems to be established when they are related by classic flops [BO2].

Part of motivic integration is the simple idea of using φ^*s_X as the volume form, even though it vanishes along the discrepancy divisor D (compare Remark 4.5). Maybe this idea can be used with differentials on X itself (not passing to $J_\infty X$) to get birationally invariant de Rham and Hodge cohomology?

Elliptic cohomology is another area of geometry with an alleged stringy interpretation – as the index of the Dirac operator on the space of loops. Could part of this theory have a rigorous treatment in terms of spaces of formal arcs, like motivic integration in Section 4? If we believe that the elliptic cohomology of M/G has a well defined answer (see Totaro [T] for some evidence) then Principle 1.1 predicts what the answer must look like in a whole pile of substantial cases.

Which Gorenstein quotient singularities admit crepant resolutions? Since 4-fold singularities usually do not have crepant resolutions, those that do are of particular interest; see [DHZ] for examples. How does this relate to complex symplectic geometry? The papers of Verbitsky [Vb] and Kaledin [Ka1], [Ka2] study crepant resolutions and related issues for symplectic quotient singularities. When crepant resolutions exist they are symplectic [Vb], therefore “semismall”, giving a complete and elegant solution to the homological

form (1) of the McKay correspondence [Ka2]. Is it possible that there is a “special” geometry in 3 complex dimensions (such as complexified imaginary quaternions), like symplectic or hyper-Kähler geometry for complex surfaces or 4-folds, that explain why crepant resolutions exist for 3-folds?

How should we interpret Nakamura’s results and conjectures on G -Hilb? If a crepant resolution exists, it would be exceedingly convenient to be able to describe it as a fine moduli space of something; G -clusters have no especially privileged role, but the requirement that the space be birational to M/G seems to impose some relation with the moduli space of group orbits. Nakamura and Nakajima have raised the question of whether the *other* crepant resolutions (after a flop) can also be interpreted as moduli, for example as Quot schemes; a single convincing example of this would add weight to their suggestion. Do the crepant resolutions in Example 5.4 have interpretations as moduli?

REFERENCES

- [Ba1] Victor Batyrev, Birational Calabi–Yau n -folds have equal Betti numbers, in *New trends in algebraic geometry*, Klaus Hulek and others (eds.), CUP, 1999, pp. 1–11
- [Ba2] Victor Batyrev, Stringy Hodge numbers of varieties with Gorenstein canonical singularities, in *Integrable systems and algebraic geometry (Kobe/Kyoto, 1997)*, 1–32, World Sci. Publishing, River Edge, NJ, 1998
- [Ba3] Victor Batyrev, Non-Archimedean integrals and stringy Euler numbers of log-terminal pairs, *J. Eur. Math. Soc.* **1** (1999), 5–33
- [BD] Victor Batyrev and Dimitrios Dais, Strong McKay correspondence, string-theoretic Hodge numbers and mirror symmetry, *Topology* **35** (1996), 901–929
- [BO1] A. Bondal and D. Orlov, Reconstruction of a variety from the derived category and groups of autoequivalences, Max Planck Inst. Bonn preprint MPI-97-36, math.AG/9712029, 20 pp.
- [BO2] A. Bondal and D. Orlov, Semi-orthogonal decomposition for algebraic varieties, preprint alg-geom/9506012
- [Br] T. Bridgeland, Equivalences of triangulated categories and Fourier–Mukai transforms, *Bull. London Math. Soc.* **31** (1999), 25–34
- [BrM] Tom Bridgeland and Antony Maciocia, Fourier–Mukai transforms for K3 fibrations, preprint math/9908022, 18 pp.
- [B] Jean-Luc Brylinski, A correspondence dual to McKay’s, preprint alg-geom/9612003
- [BKR] Tom Bridgeland, Alastair King and Miles Reid, Mukai implies McKay, preprint math/9908027, 17 pp.

- [C1] Alastair Craw, An introduction to motivic integration, preliminary draft available from www.maths.warwick.ac.uk/~craw, 23 pp.
- [C2] Alastair Craw, A -Hilb \mathbb{C}^3 and McKay correspondence, work in progress
- [CR] Alastair Craw and Miles Reid, How to calculate A -Hilb \mathbb{C}^3 , preprint math/9909085, 29 pp.
- [DHVW] L. Dixon, J. Harvey, C. Vafa and E. Witten, Strings on orbifolds. I, Nuclear Phys. B **261** (1985), 678–686. II, same J. **274** (1986), 285–314
- [DHZ] Dimitrios I. Dais, Martin Henk and Guenter M. Ziegler, All Abelian quotient c.i. singularities admit crepant resolutions, preprint alg-geom/9704007, 35 pp.
- [DL1] Jan Denef and François Loeser, Germs of arcs on singular algebraic varieties and motivic integration, Invent. Math. **135** (1999), 201–232
- [DL2] Jan Denef and François Loeser, Motivic integration, quotient singularities and the McKay correspondence, preprint math/9903187, 20 pp.
- [GSpV] G. Gonzalez-Sprinberg and J.-L. Verdier, Construction géométrique de la correspondance de McKay, Ann. Sci. École Norm. Sup. (4) **16** (1983), 409–449
- [H] R. Hartshorne, Residues and duality, L.N.M. **20** Springer, 1966
- [HH] F. Hirzebruch and Thomas Höfer, On the Euler number of an orbifold, Math. Ann. **286** (1990), 255–260
- [IN] Y. Ito and H. Nakajima, McKay correspondence and Hilbert schemes in dimension three, preprint math.AG/9803120, 35 pp., to appear in Topology
- [IN1] Yukari Ito and Iku Nakamura, McKay correspondence and Hilbert schemes, Proc. Japan Acad. Ser. A Math. Sci. **72** (1996), 135–138
- [IN2] Yukari Ito and Iku Nakamura, Hilbert schemes and simple singularities, in New trends in algebraic geometry, Hulek and others (eds.), CUP 1999, pp. 155–233.
- [IR] Yukari Ito and Miles Reid, The McKay correspondence for finite subgroups of $SL(3, \mathbb{C})$, in Higher-dimensional complex varieties (Trento, 1994), 221–240, de Gruyter, Berlin, 1996
- [K] Maxim Kontsevich, Motivic integration, Legendary lecture at Orsay, Thu 7th Dec 1995
- [Ka1] D. Kaledin, McKay correspondence for symplectic quotient singularities, preprint math/9907087, 28 pp.
- [Ka2] D. Kaledin, Dynkin diagrams and crepant resolutions of quotient singularities, preprint math/9903157, 30 pp.
- [Kr] P. B. Kronheimer, The construction of ALE spaces as hyper-Kähler quotients, J. Diff. Geom. **29** (1989), 665–683
- [GIT] D. Mumford, J. Fogarty and F. Kirwan, Geometric invariant theory (3rd edn.), Springer, 1994
- [Mu] S. Mukai, Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves, Nagoya Math. J. **81** (1981), 153–175

- [N] I. Nakamura, Hilbert schemes of Abelian group orbits, to appear in J. Alg. Geom.
- [O] D.O. Orlov, Equivalences of derived categories and K3 surfaces, Algebraic geometry, **7** J. Math. Sci. (New York) **84** (1997), 1361–1381, preprint alg-geom/9606006, 28 pp.
- [YPG] M. Reid, Young person’s guide to canonical singularities, in Algebraic Geometry, Bowdoin 1985, ed. S. Bloch, Proc. of Symposia in Pure Math. **46**, A.M.S. (1987), vol. 1, 345–414
- [R] M. Reid, McKay correspondence, in Proc. of algebraic geometry symposium (Kinosaki, Nov 1996), T. Katsura (ed.), 14–41, preprint alg-geom/9702016, 30 pp.
- [Homework] Homework sheets will be on my website www.maths.warwick.ac.uk/~miles, including examples, exercises, more hints, and errata to this lecture.
- [Roan] S.-S. Roan, Orbifold Euler characteristic, in Mirror symmetry, II, AMS 1997, pp. 129–140
- [T] Burt Totaro, Chern numbers for singular varieties and elliptic homology, to appear in Ann. of Math., preprint available from www.dpmms.cam.ac.uk/~bt219, 30 pp.
- [V] Cumrun Vafa, String vacua and orbifoldized LG models, Modern Phys. Lett. A **4** (1989), 1169–1185
- [Vb] M. Verbitsky, Holomorphic symplectic geometry and orbifold singularities, preprint math.AG/9903175, 17 pp.
- [Z] Eric Zaslow, Topological orbifold models and quantum cohomology rings, Comm. Math. Phys. **156** (1993), 301–331

Miles REID

Adresse courante : Math Inst.,

Univ. of Warwick, Coventry CV4 7AL

E-mail : Miles@Maths.Warwick.Ac.UK

LES VORTEX DE GINZBURG-LANDAU : LE MODÈLE STATIQUE

par **Tristan RIVIÈRE**

1. INTRODUCTION

1.1. Origines physiques du problème

Un des premiers modèles explicatifs de la supraconductivité (existence dans certains matériaux de courants permanents sans dissipation d'énergie) a été proposé durant les années 50 par V. Ginzburg et L. Landau à partir de la théorie de Landau sur les transitions de phases. Suivant ce modèle, la supraconductivité, dans un matériau occupant le domaine Ω de κ^3 , est caractérisée par une "fonction d'onde" $u : \Omega \rightarrow \mathcal{C}$, souvent appelée paramètre d'ordre. Dans la théorie quantique de J. Bardeen, L.N. Cooper et J. Schrieffer (théorie BCS) qui est venue justifier en 1957 le modèle phénoménologique de Ginzburg et Landau, le carré du module de ce paramètre d'ordre $|u|^2$ représente la densité locale de couplage d'électrons par paires, dites paires de Cooper, responsable du phénomène supraconducteur. Lorsque $|u| = 1$ la densité est maximale et $|u| = 0$ minimale.

La fonctionnelle d'énergie d'un supraconducteur proposée par Ginzburg et Landau est

$$\mathcal{J}(u, A) = \int_{\Omega} |\kappa^{-1} du - iAu|^2 + \frac{1}{2} |1 - |u|^2|^2 + |dA|^2 + 2 \int_{\Omega} dA \cdot h_e,$$

où A est la 1-forme potentiel vecteur associée au champ induit dA dans le supraconducteur ($du - iAu$ est donc une 1-forme à valeurs dans \mathcal{C}). h_e est la 2-forme représentant le champ extérieur auquel est soumis le supraconducteur. Il s'agit d'un paramètre du problème ainsi que la constante κ , dite constante de couplage, qui dépend du matériau et qui, comme nous le verrons plus bas, joue un rôle central dans la théorie. C'est une constante sans dimension, rapport de deux longueurs $\kappa = \frac{\lambda}{\xi}$, λ est la longueur de pénétration du champ extérieur h_e dans le matériau (voir plus bas) et ξ est la taille caractéristique d'un vortex (voir section 2). Notons que cette fonctionnelle est aussi la fonctionnelle d'action Yang-Mills-Higgs en théorie de jauge abélienne donnant la modélisation de l'interaction entre un champ magnétique classique et une particule de Higgs.

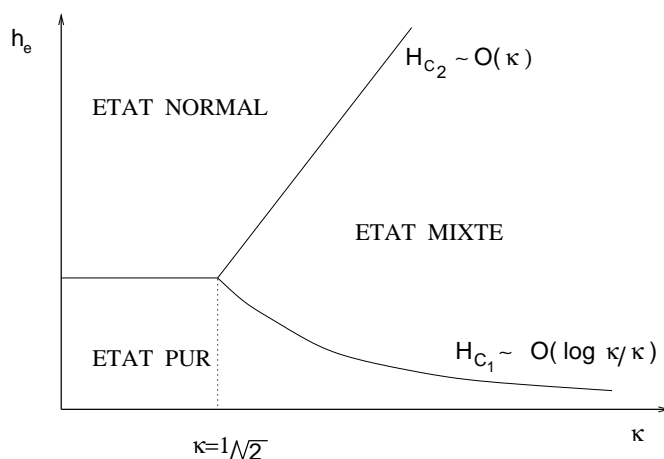


FIGURE 1. Diagramme de phases

Schématiquement les phénomènes observés sont les suivants. Lorsque le champ extérieur est nul le supraconducteur est dit dans l'*état pur* :

$$\begin{cases} |u| = 1 & \text{dans } \Omega \\ dA = 0 & \text{dans } \Omega. \end{cases}$$

La densité de paires de Cooper est maximale et le champ induit est nul. Lorsque le champ appliqué est suffisamment fort (dépendant du matériau) la supraconductivité disparaît :

$$\begin{cases} |u| = 0 & \text{dans } \Omega \\ dA = h_e & \text{dans } \Omega \end{cases}$$

la densité de paires de Cooper est alors minimale et le champ induit coïncide avec le champ extérieur. Le supraconducteur est alors dans l'*état normal*.

La nature du passage de l'*état pur* à l'*état normal* varie suivant le matériau et en particulier la valeur de κ . Il est observé que, pour $\kappa < 1/\sqrt{2}$ (supraconducteur de type I), le passage de l'état pur à l'état normal est brutal et s'effectue pour une intensité du champ extérieur indépendante de κ . Tandis que, pour $\kappa > 1/\sqrt{2}$ (supraconducteurs de type II), la transition de l'état pur à l'état normal, au fur et à mesure que le champ extérieur s'accroît, se fait par le passage dans une phase, dite *mixte*, où apparaissent des zones de plus en plus nombreuses d'état normal contenues dans des tubes (filaments de vorticité) autour desquels la phase de u effectue un ou plusieurs tours du cercle. Lorsque le matériau est homogène et le champ extérieur uniforme, ces tubes s'alignent dans le sens du champ pour former des réseaux périodiques dits d'Abrikosov du nom du physicien qui les a mis en évidence. Il est observé que le réseau dans l'état fondamental est triangulaire. On passe de l'*état pur* à l'*état mixte* pour un champ dit "premier champ critique" $H_{c_1} \simeq O\left(\frac{\log \kappa}{\kappa}\right)$ et on sort de l'*état mixte* pour rentrer dans l'*état normal* pour un champ dit "deuxième champ critique" $H_{c_2} \simeq O(\kappa)$. Le diagramme (figure 1) résume l'ensemble des observations mentionnées. Pour une présentation plus complète de la physique des supraconducteurs

nous renvoyons le lecteur aux références suivantes : [dG], [SST], [Ti]...

1.2. Les questions mathématiques de la supraconductivité

Les difficultés pour rendre mathématiquement rigoureuses les observations précédentes à partir du modèle de Ginzburg-Landau sont nombreuses. Une première réduction consiste à se ramener à une version 1 ou 2-dimensionnelle du modèle ($\Omega \subset \mathbb{R}^3$ et h_e ont alors les symétries correspondant à ces réductions : espace entre deux plans parallèles ou cylindre infini dans des champs uniformes...). Dans cet exposé nous n'aborderons pas les études en dimension 1 qui sont cependant nombreuses et qui permettent souvent d'analyser plus finement le diagramme des phases ([BH1], [BH2], [Af]... voir une présentation complète des résultats dans [AT]). Nous nous restreindrons ici au cas de la dimension 2, dimension minimale pour observer les vortex (les dimensions 3 et supérieures sont considérées dans [Ri2], [LR] et aussi dans [BBM]). Ω est donc un cylindre infini et h_e un champ uniforme orienté dans la direction de l'axe du cylindre. Ω désigne alors la section bidimensionnelle de ce cylindre et h_e étant une deux-forme constante sur cette section est souvent confondue avec le nombre donnant son intensité. Le but est donc de comprendre l'aspect des états fondamentaux de la fonctionnelle \mathcal{J} , ainsi que celui de ses points critiques en général, en fonction des différentes valeurs de (κ, h_e) dans le diagramme de phases figure 1. Par comprendre l'aspect des états fondamentaux de \mathcal{J} , on entend surtout l'identification du lieu d'annulation du paramètre d'ordre u d'une solution minimisant \mathcal{J} qui correspond à la coupe bidimensionnelle du réseau de vortex attendu dans la phase mixte.

Pour la commodité de l'analyse, on effectue le changement de variable $A \rightarrow \kappa A$ par rapport au modèle original, ce qui nous fait considérer désormais la fonctionnelle

$$\mathcal{G}_\kappa(u, A) = \int_\Omega |du - iAu|^2 + \frac{\kappa^2}{2}|1 - |u|^2|^2 + |dA|^2 + 2h_e \int_\Omega dA.$$

Celle-ci bénéficie de l'invariance de jauge $\mathcal{G}_\kappa(u, A) = \mathcal{G}_\kappa(e^{i\phi}u, A+d\phi)$ pour toute fonction ϕ sur Ω . Il est alors possible d'étendre le modèle au cas d'un domaine Ω qui soit une variété bidimensionnelle quelconque. (u, A) sont alors respectivement les sections et connexions d'un fibré en droite complexe E sur Ω que l'on munit d'un produit hermitien dont la partie réelle est (\cdot, \cdot) ou encore $|\cdot|^2$ pour la forme quadratique. $du - iAu$ est remplacé par la dérivation covariante $d_A u$ de u relativement à A et dA est la courbure de la connexion A . On notera par la suite $h = *dA$.

La section 2 est consacrée à l'étude de l'énergie libre de Ginzburg-Landau \mathcal{F} sans interactions avec le champ extérieur (i.e. $h_e = 0$). Dans 2.1, nous présentons les travaux de Jaffe et Taubes sur le cas dit intégrable ou non interactif $\kappa = \frac{1}{\sqrt{2}}$. Nous énonçons leurs conjectures pour $\kappa < \frac{1}{\sqrt{2}}$ et $\kappa > \frac{1}{\sqrt{2}}$. Dans 2.2, nous décrivons l'analyse asymptotique BBH (F. Béthuel, H. Brézis et F. Hélein [BBH]) dans la limite de London : $\kappa \rightarrow +\infty$ qui correspond au cas fortement répulsif. Dans cette limite, à laquelle nous nous restreindrons par la suite, les phénomènes de vorticités se dégagent plus nettement ; c'est aussi un cas

qui s'approche de la réalité de nombreux supraconducteurs de type II pour lesquels le paramètre non dimensionnel κ est très grand. Dans 2.3, nous revenons sur le rôle joué par l'énergie renormalisée W issue de l'analyse asymptotique BBH pour décrire l'espace des points critiques de \mathcal{F} . Enfin nous apportons des réponses aux conjectures de Jaffe et Taubes dans la limite de London et nous les étendons à des situations plus générales. La troisième partie est consacrée à l'étude de la fonctionnelle \mathcal{G} dans son intégralité comprenant le terme d'interaction avec le champ extérieur. La vorticité n'est alors plus un paramètre comme dans la section précédente mais devient une variable du problème. L'ensemble de ce qui y est présenté couvre une partie des résultats de la thèse de S. Serfaty, ainsi que ses travaux en commun avec E. Sandier.

2. ÉTUDE DE LA FONCTIONNELLE D'ÉNERGIE LIBRE \mathcal{F}

2.1. Le cas intégrable ou non interactif $\kappa = 1/\sqrt{2}$

Dans [JT], A. Jaffe et C. Taubes étudient les points critiques sur \mathfrak{r}^2 de la fonctionnelle d'énergie libre

$$\mathcal{F}_\kappa(u, A) = \int_{\mathfrak{r}^2} |d_A u|^2 + \frac{\kappa^2}{2} |1 - |u|^2|^2 + |dA|^2,$$

solutions des équations d'Euler

$$(1) \quad \begin{cases} d_A^* d_A u = \kappa^2 u(1 - |u|^2) \\ d^* dA = (iu, d_A u) \end{cases}$$

où d_A^* est l'opérateur agissant sur les 1-formes η donné par $d_A^* \eta = d^* \eta + iA \wedge \eta$. Sous des hypothèses de décroissance (polynomiale) des quantités intrinsèques $|d_A u|$, $|1 - |u||$ et dA le flux normalisé du champ magnétique est un entier N

$$N = \frac{1}{2\pi} \int_{\mathfrak{r}^2} dA$$

qui correspond au degré de $u/|u|$ sur des cercles de rayons suffisamment grands. C'est ce que l'on appelle la classe d'homotopie du couple (u, A) . N étant donné, disons $N \geq 0$, il a été observé par E.B. Bogomol'nyi [Bog] que, pour la valeur particulière $\kappa = 1/\sqrt{2}$, la fonctionnelle \mathcal{F} se réécrit sous la forme suivante

$$\begin{aligned} \mathcal{F}_\kappa(u, A) \int_{\mathfrak{r}^2} & |\Re(d_{A_1} u) - \Im(d_{A_2} u)|^2 + |\Re(d_{A_2} u) - \Im(d_{A_1} u)|^2 \\ & + |*dA + \frac{1}{2}(|u|^2 - 1)|^2 + 2\pi N \end{aligned}$$

(\Re désigne la partie réelle et \Im la partie imaginaire d'un nombre complexe). Un des résultats principaux de [JT] est alors le suivant :

THÉORÈME 2.1. — [JT] *Tout point critique (u, A) de $\mathcal{F}_{1/\sqrt{2}}$ d'énergie finie a une classe d'homotopie N définie et vérifie*

$$(2) \quad \mathcal{F}_{1/\sqrt{2}}(u, A) = 2\pi |N|.$$

En particulier, il minimise $\mathcal{F}_{1/\sqrt{2}}$ dans sa classe d'homotopie.

La démonstration de ce résultat peut être comprise ainsi. Considérons un point critique (u, A) de \mathcal{F}_κ . Des équations d'Euler (1), on peut déduire les équations elliptiques suivantes vérifiées par les quantités intrinsèques $1 - |u|^2$ et $h = *dA$

$$(3) \quad \begin{cases} -\Delta \frac{(1 - |u|^2)}{2} + 2\kappa^2 |u|^2 \frac{(1 - |u|^2)}{2} = |d_A u|^2 \\ -\Delta h + |u|^2 h = (d_A u; i * d_A u) \end{cases}$$

où $-\Delta = d^*d$ et $(d_A u; i * d_A u) = - \langle \Re(d_A u), \Im(*d_A u) \rangle + \langle \Im(d_A u), \Re(*d_A u) \rangle$ ($\langle \cdot, \cdot \rangle$ est le produit scalaire sur les 1-formes). À partir de (3), le principe du maximum nous permet de déduire, d'une part que $|u| < 1$ (sauf si $|u| \equiv 1$ sur κ^2), d'autre part que les quantités intrinsèques $|h|$, $|d_A u|$ et $|1 - |u|^2|$ décroissent exponentiellement vite à l'infini, ce qui permet en particulier d'avoir une classe d'homotopie bien définie pour (u, A) . Ces décroissances exponentielles ont une interprétation physique importante liée à la masse de la particule de Higgs (voir [JT]).

Un autre ingrédient essentiel de la preuve du théorème 2.1 est l'utilisation de la loi de conservation du tenseur énergie-impulsion qui va revenir régulièrement tout au long de l'exposé. Le tenseur énergie-impulsion $(T_{ij})_{i,j \in \{1,2\}}$ est donné par

$$(4) \quad T_{i,j} = 2\delta_{ij}|dA|^2 + 2(d_{A_i}u, d_{A_j}u) - \delta_{ij}f_\kappa(u, A),$$

où $f_\kappa(u, A)$ est la densité d'énergie libre $f_\kappa(u, A) = |d_A u|^2 + \frac{\kappa^2}{2}|1 - |u|^2|^2 + |dA|^2$. La loi de conservation de ce tenseur résulte d'une part du fait que (u, A) est un point critique de \mathcal{F}_κ , d'autre part du fait qu'il est C^∞ , ce que l'on déduit des équations d'Euler (1) ; il est donc aussi point critique pour des variations du domaine. L'invariance par translations du domaine donne alors, d'après le théorème de Noether, des quantités à divergence nulle qui constituent cette loi de conservation

$$(5) \quad \forall j = 1, 2 \quad \frac{\partial}{\partial x_i} T_{ij} = 0.$$

De cette loi se déduit l'identité de Pohozaev suivante qui est conséquence du fait que (u, A) est point critique pour l'action infinitésimale des dilatations $r \frac{\partial}{\partial r}$ (i.e. on multiplie (5) par x_j , on somme sur j et on intègre sur tout κ^2)

$$(6) \quad \int_{\kappa^2} \frac{\kappa^2}{2} |1 - |u|^2|^2 = \int_{\kappa^2} |dA|^2.$$

Restreignons-nous maintenant au cas $\kappa = 1/\sqrt{2}$. Il est aisé de voir alors, pour cette valeur particulière du paramètre, qu'en sommant ou soustrayant les deux équations de (3), l'application du principe du maximum donne

$$(7) \quad |dA| = |h| \leq \frac{|1 - |u|^2|}{2}.$$

Par ailleurs, pour $\kappa = 1/\sqrt{2}$, l'identité de Pohozaev devient

$$(8) \quad \int_{\mathcal{R}^2} \left(\frac{1 - |u|^2}{2} - *dA \right) \left(\frac{1 - |u|^2}{2} + *dA \right) = 0.$$

Si on écarte le cas simple où $|u| \equiv 1$ ($N = 0$) et donc $dA \equiv 0$ d'après (6), on a $|u| < 1$ et la confrontation de (7) et (8) donne que ou bien $*dA = \frac{1 - |u|^2}{2}$ (cas $N > 0$), ou bien $*dA = -\frac{1 - |u|^2}{2}$ (cas $N < 0$). En "bootstrapant" par exemple $*dA = \frac{1 - |u|^2}{2}$ dans les équations, on obtient aisément que $\Re(d_{A_1}u) = \Im(d_{A_2}u)$ et que $\Re(d_{A_2}u) = \Im(d_{A_1}u)$, ce qui, au moyen de l'observation de Bogomol'nyi, démontre le théorème.

Le théorème 1 nous dit donc que la résolution des solutions des équations du deuxième ordre (1) se ramène, dans le cas $\kappa = 1/\sqrt{2}$ et dans la classe des configurations (u, A) d'énergie finie, à l'étude des solutions des équations du premier ordre (cas $N > 0$)

$$(9) \quad *dA = \frac{1 - |u|^2}{2}$$

$$(10) \quad \Re(d_{A_2}u) = \Im(d_{A_1}u)$$

$$(11) \quad \Re(d_{A_1}u) = \Im(d_{A_2}u),$$

ce qui justifie le nom d'intégrable au cas considéré. L'étude qualitative des solutions de ces équations donne sans trop de difficultés que l'ensemble des zéros de $|u|$ est constitué d'un nombre fini de points où l'indice de u est strictement positif. On peut donc le représenter par exactement N points, éventuellement confondus, $\{x_1 \dots x_N\}$ de multiplicité 1 chacun. On vérifie alors que $v = \log |u|^2$ est solution de

$$(12) \quad -\Delta v + e^v - 1 = -4\pi \sum_{k=1}^N \delta_{x_k} \quad \text{dans } \mathcal{D}'(\mathcal{R}^2).$$

Un argument de convexité donne l'unicité de v solution de (12) pour toute configuration de points $\{x_1 \dots x_N\}$. On "bootstrape" aisément cette unicité de $|u|$ dans les équations (9), (10) et (11) pour en déduire l'unicité (modulo l'action de jauge) de la configuration (u, A) . On peut alors effectuer le chemin inverse et, partant de N points quelconques dans le plan, produire une solution de (9)...(11). On a alors démontré le théorème suivant.

THÉORÈME 2.2. — [JT] *L'espace des solutions de Yang-Mills-Higgs abélien (1) dans le plan, d'énergie finie, se confond (modulo l'action du groupe de jauge), dans le cas intégrable $\kappa = 1/\sqrt{2}$, avec l'espace des configurations de points dans le plan munis de multiplicités entières toutes de même signe.*

Un des aspects marquants du cas intégrable $\kappa = 1/\sqrt{2}$ est que, non seulement les vortex (zéros de u) peuvent être n'importe où dans le plan, mais aussi l'énergie de la solution est indépendante de la position relative des vortex entre eux, égale à 2π fois le nombre de vortex. C'est pourquoi le cas intégrable est aussi appelé le cas non interactif. Des considérations heuristiques amènent A. Jaffe et C. Taubes à conjecturer que $\kappa = 1/\sqrt{2}$ est le cas frontière entre deux comportements opposés des vortex entre eux : dans le cas $\kappa < 1/\sqrt{2}$ les vortex, quels que soient leurs signes auraient tendance à s'attirer, ce qui justifierait en quelque sorte l'absence d'état mixte, alors que dans le cas $\kappa > 1/\sqrt{2}$ les vortex de même signe devraient se repousser, leur présence étant cependant imposée par l'apport d'énergie due à l'application du champ extérieur (voir la section 3), d'où l'existence possible d'une phase mixte (voir figure 1). Précisément la conjecture de Jaffe et Taubes dans le cas répulsif ($\kappa > 1/\sqrt{2}$) est la suivante.

CONJECTURE 2.3. — [JT] *Pour $\kappa > 1/\sqrt{2}$, il existe des solutions stables de YMH (1) sur tout κ^2 si et seulement si $|N| = \pm 1, 0$; d'autre part elles sont à symétrie axiale (modulo l'action de jauge).*

Dans les deux sous-sections suivantes, nous allons rendre compte rigoureusement de ces comportements attendus des vortex entre eux dans le cas fortement répulsif ($\kappa \rightarrow +\infty$) et nous apporterons une réponse partielle à la conjecture 2.3 (voir théorème 2.10).

2.2. Le cas fortement répulsif ou “limite de London” $\kappa \rightarrow +\infty$: l'analyse asymptotique BBH.

À partir de maintenant, nous étudions le comportement des vortex dans le cas fortement répulsif ($\kappa \rightarrow +\infty$). Afin d'empêcher les vortex de se séparer à l'infini, comme le prédit la conjecture 2.3, on étudie \mathcal{F} sur une variété compacte M de dimension 2 et sans bord, ce qui permet d'éviter des conditions aux limites artificielles. Pour la clarté de l'exposé, nous nous restreindrons au cas où M est un tore plat. En réalité la métrique sur M , ainsi que sa topologie, ne modifient pas l'aspect qualitatif des résultats. On se donne d'autre part un fibré hermitien en droite complexe E sur M dont la classe d'Euler $e(E)$ vérifie $\int_M e(E) = N > 0$, ce qui signifie que toute section u intersectant transversalement la section nulle le fait algébriquement N fois. Nous fixons donc la vorticit  totale qui devient un param tre du probl me. Dans la section 3 de cet expos , on prendra en compte l'action du champ ext rieur et la vorticit  totale redeviendra une variable du probl me de minimisation comme le veut le mod le d'origine. L'existence d'un couple section-connexion (u_κ, A_κ) minimisant \mathcal{F}_κ dans les espaces de Sobolev $W^{1,2}$ de sections et connexions (l'expression dans une trivialis tion donn e du fibr  de (u, A) donne des fonctions et 1-formes $W^{1,2}$) est un probl me d sormais classique qui fait appel   l'utilisation de jauges de Coulomb afin de rendre la fonctionnelle coercitive (cf [Uh]). On se propose alors d' tudier le comportement de tels couples (u_κ, A_κ) minimisants lorsque

κ tend vers l'infini. La difficulté d'une telle analyse asymptotique vient du fait que l'on ne dispose pas d'estimations *a priori* suffisantes, indépendantes de κ , dans un quelconque espace fonctionnel, pour prouver une quelconque convergence faible vers quelque chose. En particulier, on peut vérifier que $\mathcal{F}_\kappa(u_\kappa, A_\kappa) \rightarrow +\infty$ et plus précisément on a

$$(13) \quad \mathcal{F}_\kappa(u_\kappa, A_\kappa) \simeq 2\pi N \log \kappa.$$

F. Béthuel, H. Brézis et F. Hélein donnent une description complète de cette asymptotique dans [BBH] pour le cas $A = 0$ sur un domaine de κ^2 (elle est approfondie dans [St], adaptée au cas du modèle invariant de jauge dans [BR1] puis dans [Qin] et faisait suite à l'étude du cas symétrique sur tout κ^2 dans [BC]). Ils établissent alors le résultat suivant :

THÉORÈME 2.4. — [BBH] *Étant donnée une suite de configurations (u_κ, A_κ) minimisant \mathcal{F}_κ pour une suite de κ tendant vers l'infini, il existe une sous-suite $(u_{\kappa'}, A_{\kappa'})$ et N points distincts $\{p_1 \dots p_N\}$ de M tels que*

$$(14) \quad (u_{\kappa'}, A_{\kappa'}) \longrightarrow (u_\star, A_\star) \quad \text{dans } C_{loc}^k(\tilde{M}),$$

où $\tilde{M} = M \setminus \{p_1 \dots p_N\}$, (u_\star, A_\star) est un couple section unitaire-connexion de E au-dessus de \tilde{M} point critique de la fonctionnelle

$$(15) \quad \mathcal{F}_\star(u_\star, A_\star) = \int_{\tilde{M}} |d_{A_\star} u_\star|^2 + |dA_\star|^2,$$

d'autre part l'indice de la section A_\star -harmonique singulière u_\star à chaque p_j est $+1$, ce qui donne en particulier que la courbure limite $h_\star = *dA_\star$ vérifie l'équation de London

$$(16) \quad d^*dh_\star + h_\star = 2\pi \sum_{k=1}^N \delta_{p_k} \quad \text{dans } \mathcal{D}'(M).$$

REMARQUE 1. — La position des vortex limites $p_1 \dots p_N$ détermine de façon unique le couple (u_\star, A_\star) (modulo l'invariance de jauge).

Le déroulement de la preuve est le suivant – on omet l'indice κ .

À nouveau, le principe du maximum appliqué à la première équation de (3) donne $|u| \leq 1$. En combinant cette majoration L^∞ de $|u|$ et la majoration de l'énergie de la configuration (u, A) donnée par (13), on obtient, au moyen d'estimations elliptiques classiques (dans l'esprit des inégalités d'interpolations du type Gagliardo-Nirenberg [BBH0]), le contrôle suivant de la norme L^∞ de la dérivée covariante de u (qui est en fait optimal)

$$(17) \quad \|d|u|\|_{L^\infty(M)} \leq \|d_A u\|_{L^\infty(M)} = O(\kappa).$$

La stratégie générale va consister à identifier et recouvrir le mieux possible le lieu d'annulation de $|u|$ qui va coïncider avec le lieu de perte de compacité de la suite de configuration (u, A) dans $W^{1,2}$. Nous appelons “mauvais ensemble” ce lieu ; plus précisément il s'agit de l'ensemble suivant :

$$(18) \quad \mathcal{M} = \{x \in M \quad \text{tel que} \quad |u|(x) < 1/2\}$$

(1/2 est en fait une constante choisie arbitrairement entre 0 et 1). On a alors le résultat de “quantification” suivant.

Il existe $\delta > 0$ indépendant de κ tel que, pour tout x_0 dans M ,

$$(19) \quad \kappa^2 \int_{B_{\kappa^{-1}}(x_0)} |1 - |u|^2|^2 \leq \delta \quad \implies \quad x_0 \notin \mathcal{M}.$$

Ce résultat est une conséquence immédiate de (17). Il signifie qu’un point de \mathcal{M} contribue, sur une boule de rayon κ^{-1} , à une portion “finie” de l’énergie de Higgs $\kappa^2 \int_{B_{\kappa^{-1}}(x_0)} |1 - |u|^2|^2$ (i.e. supérieure à une valeur $\delta > 0$ indépendante de κ).

L’utilisation conjointe de la majoration de l’énergie (13) et de la loi de conservation du tenseur énergie-impulsion (5), de laquelle on déduit une identité de Pohozaev sur toute boule géodésique, nous donne en particulier (cf [BR1])

$$(20) \quad \forall 0 < \alpha < 1 \quad \forall x \in M \quad \kappa^2 \int_{B_{\kappa^{-\alpha}}(x)} |1 - |u|^2|^2 \leq C_\alpha,$$

où C_α est indépendant de κ . L’utilisation de ces échelles intermédiaires $\kappa^{-\alpha}$ dans [BR1] entre les échelles naturelles du problème (1 et κ^{-1}) permet de faire disparaître le champ magnétique dans les identités de Pohozaev et d’obtenir (20). Ces échelles (pas vraiment nécessaires dans le cas $A = 0$ en dimension 2 [BBH]) sont très utilisées dans le modèle invariant de jauge ([BR1], [Ri1], [Ri2]...). La confrontation de (19) et de (20) donne le recouvrement suivant de \mathcal{M} sur toute boule de rayon $\kappa^{-\alpha}$.

$$(21) \quad \forall x \in M \quad \mathcal{M} \cap B_{\kappa^{-\alpha}}(x) \subset \cup_{j=1}^{N_\alpha} B_{\kappa^{-1}}(x_j) \quad \text{et } N_\alpha = O(1),$$

où $(x_j)_{j \in \{1 \dots N_\alpha\}}$ est une famille de points dans $B_{\kappa^{-\alpha}}(x)$ et N_α est majoré indépendamment de κ .

À nouveau l’utilisation de la loi de conservation du tenseur énergie-impulsion (5) permet de démontrer le résultat de quantification suivant :

LEMME 2.5. — [eta-compacité] Il existe $\eta > 0$ indépendant de κ tel que pour tout rayon $1 \geq \rho \geq \kappa^{-1}$ et tout x dans M on ait

$$(22) \quad \int_{B_\rho(x)} f_\kappa(u, A) \leq \eta \log(\rho\kappa) \\ \implies B_{\rho/2}(x) \cap \mathcal{M} = \emptyset,$$

où $f_\kappa(u, A)$ est la densité d’énergie libre $f(u, A) = |d_A u|^2 + \frac{\kappa^2}{2} |1 - |u|^2|^2 + |dA|^2$.

Ce lemme nous dit que la contribution du mauvais ensemble \mathcal{M} à l’énergie totale sur une boule de rayon ρ est au moins supérieure à $\eta \log(\rho\kappa)$. Il est dans l’esprit des lemmes d’ ϵ -régularité de la théorie des applications harmoniques ou encore de lemmes équivalents en théorie des surfaces minimales. Sa démonstration en dimension 2 est assez immédiate (cf [Ri1]), alors qu’en dimension supérieure elle est bien plus technique (cf [Ri2] [LR]).

La confrontation de la majoration de l'énergie (13), de (21) et du lemme de éta-compactité nous permet de conclure aisément que l'ensemble \mathcal{M} est contenu dans un nombre uniformément borné de boules de rayon κ^{-1} :

$$(23) \quad \mathcal{M} \subset \cup_{j=1}^n B_{\kappa^{-1}}(x_j) \quad n = O(1).$$

On peut alors extraire une sous-suite de la suite d'origine $\kappa \rightarrow +\infty$ telle que la famille des $(x_j)_{j=1..n}$ converge dans M .

En dehors de \mathcal{M} , $h = *dA$ vérifie l'équation elliptique suivante déduite de la deuxième équation de Ginzburg-Landau (1) :

$$(24) \quad d^* \left[\frac{1}{|u|^2} dh \right] + h = 0 \quad \text{dans } M \setminus \cup_{j=1}^n B_{\kappa^{-1}}(x_j).$$

L'obstruction à ce que cette équation soit vérifiée sur tout M vient de l'indice de u autour de son lieu d'annulation. Si donc $d_j = \deg(\frac{u}{|u|}; \partial B_{\kappa^{-1}}(x_j))$, on vérifie que

$$(25) \quad \int_{\partial B_{\kappa^{-1}}(x_j)} \frac{dh}{|u|^2} + \int_{B_{\kappa^{-1}}(x_j)} h = 2\pi d_j.$$

(24) et (25) nous disent alors que h est une κ^{-1} -approximation du problème linéaire suivant

$$(26) \quad d^* dk + k = 2\pi \sum_{j=1}^n \delta_{x_j}.$$

De (17) on déduit que les $|d_j|$ sont uniformément bornés et donc que les normes $W^{1,p}$ ($p < 2$) de k sont majorées indépendamment de κ . Ces majorations se transmettent aisément à $h = *dA$ et en les "bootstrapant" dans les équations de Ginzburg-Landau (1) on démontre le théorème 2.4 (par ailleurs afin d'établir $n = N$, on a utilisé la minimalité de la solution).

Contrairement au cas non-interactif $\kappa = 1/\sqrt{2}$, les vortex x_j et leurs limites $p_1 \dots p_N$ ne peuvent être n'importe où dans le domaine M , la configuration $(p_1 \dots p_N)$ minimise une certaine énergie W de $M^N \setminus \Delta$ dans \mathfrak{r} dite énergie renormalisée (Δ désigne la diagonale de M^N).

THÉORÈME 2.6. — [BBH] *La configuration de vortex limite $(p_1 \dots p_N)$ donnée par le théorème 2.4 est un état fondamental de la fonction suivante définie sur $M^N \setminus \Delta$*

$$(27) \quad W(z_1 \dots z_N) = 2 \sum_{i \neq j} \int_M dG_i \cdot dG_j + 2 \sum_{j=1}^N \int_M dR \cdot dG_j \\ + \int_M |dR|^2 + \int_M |k|^2$$

où $d^* dk + k = 2\pi \sum_{j=1}^N \delta_{z_j}$, $G_j(z) = \log|z - z_j|$ et $k = R + \sum_{j=1}^N G_j$.

Le caractère fortement répulsif des vortex entre eux est bien apparent dans W . En effet on vérifie que si deux vortex z_i et z_j se rapprochent (les autres restant fixes), on a $W \simeq 2\pi \log |z_i - z_j|$.

La preuve du théorème 2.6 repose sur la décomposition suivante. On se donne $\delta > 0$ indépendant de κ . On divise M en deux parties disjointes $M = \cup_{k=1}^N B_\delta(p_k) \cup M_\delta$. On décompose l'énergie totale comme la somme des énergies sur chaque boule $B_\delta(p_k)$ et de l'énergie sur M_δ . D'après le théorème 2.4, $|u|$ converge uniformément vers 1 tandis que $*dA$ converge uniformément vers h_\star solution de (16) sur M_δ . En utilisant alors la deuxième équation de Ginzburg-Landau on obtient

$$(28) \quad \lim_{\kappa \rightarrow +\infty} \int_{M_\delta} f_\kappa(u, A) = \int_{M_\delta} |dh_\star|^2 + |h_\star|^2.$$

Un calcul explicite permet de vérifier par ailleurs que

$$(29) \quad \lim_{\delta \rightarrow 0} \int_{M_\delta} |dh_\star|^2 + |h_\star|^2 - 2\pi N \log \frac{1}{\delta} = W(p_1 \dots p_N) + C,$$

où C est indépendant de la position des vortex. En utilisant le recouvrement de M par un nombre fini de boules de rayons κ^{-1} , on démontre, à l'aide de la convergence donnée par le théorème 2.4, que la partie principale de l'énergie autour d'un vortex limite est indépendant de sa position et de l'existence des autres vortex

$$(30) \quad \int_{B_\delta(p_k)} f_\kappa(u, A) = 2\pi \log(\delta\kappa) + C_0 + o_\delta(1),$$

où C_0 est une constante universelle. En combinant (28), (29) et (30) on vérifie aisément que pour optimiser \mathcal{F}_κ , il faut optimiser la configuration de vortex $(p_1 \dots p_N)$ relativement à W . Ceci démontre le théorème 4.

On a par ailleurs démontré que $\mathcal{F}_\kappa(u_\kappa, A_\kappa)$ admet le développement asymptotique suivant.

$$(31) \quad \mathcal{F}_\kappa(u_\kappa, A_\kappa) = 2\pi N \log \kappa + W(p_1 \dots p_N) + C_0 N + o(1)$$

Ce développement asymptotique peut être interprété ainsi : chaque vortex interagit sur lui-même avec une énergie de partie principale $2\pi \log \kappa$, W est l'énergie d'interaction des vortex entre eux, enfin C_0 est l'énergie renormalisée d'une particule isolée.

Si maintenant on suit une suite de solutions (u_κ, A_κ) des équations de Ginzburg-Landau (1) dans la limite de London ($\kappa \rightarrow +\infty$), qui n'est cette fois pas nécessairement une suite de minimums, on démontre des résultats correspondant aux théorèmes 2.4 et 2.6 à condition de rester dans des niveaux d'énergie "raisonnables" $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$. Les vortex limites $p_1 \dots p_Q$ ne sont plus alors forcément au nombre de N . À chacun est associée une multiplicité entière d_j qui est l'indice de la section singulière limite u_\star autour de p_j .

L'équation vérifiée par le champ limite h_* devient

$$(32) \quad d^*dh_* + h_* = 2\pi \sum_{j=1}^Q d_j \delta_{p_j} \quad \text{dans } \mathcal{D}'(M)$$

tandis que la position des p_j est point critique de la fonction W donnée par (27) où G_j est remplacé par $d_j G_j$.

2.3. L'espace des solutions dans la limite de London

Étude du lieu d'annulation dans la limite de London. On a vu que, dans la limite de London ($\kappa \rightarrow +\infty$), l'ensemble d'annulation de la section u_κ minimisante est contraint de converger vers un lieu bien précis du domaine : $p_1 \dots p_N$ point critique de la fonction W . On cherche désormais à avoir une description plus complète de ce lieu d'annulation et, éventuellement, à voir dans quelle limite la propriété obtenue dans le cas non interactif, disant que deux solutions ont le même lieu d'annulation sont équivalentes de jauge, reste vraie dans le cas fortement répulsif. Ceci ramènerait alors l'étude de l'espace des solutions à celle du lieu d'annulation possible et justifierait le glissement désormais courant dans le langage de la physique où l'on parle de vortex de Ginzburg-Landau pour désigner les solutions des équations de Ginzburg-Landau elles-mêmes.

Considérons une suite de minimums (u_κ, A_κ) de \mathcal{F}_κ convergeant suivant le théorème 2.2 vers un couple section-connexions singulières (u_*, A_*) . Nous cherchons, dans un premier temps, à décrire le lieu d'annulation de $|u_\kappa|$ – on omettra de mentionner l'indice κ sauf lorsque cela sera nécessaire. La section limite étant d'indice 1 à p_1 la somme des indices de u sur les mauvaises boules $B_{\kappa^{-1}}(x_i)$ convergeant vers p_1 est aussi +1. On peut toujours supposer que $|u|(x_j) < 1/2$.

Si jamais deux centres de mauvaises boules x_l et x_k s'écartent plus vite que $O(\kappa^{-1})$, on obtient une contradiction pour la raison suivante. Supposons que ceci se produit, alors modulo l'extraction d'une sous-suite, on peut supposer que

$$(33) \quad \kappa|x_k - x_l| \longrightarrow +\infty.$$

Considérons alors la dilatation de la section autour du point x_k donnée par (dans une jauge de Coulomb locale) : $\hat{u}_\kappa(z) = u(\kappa^{-1}z + x_k)$. La convergence du champ dA établie dans la preuve du théorème 2.4 ainsi que la majoration L^∞ de la dérivée covariante (17) permettent de conclure que, modulo l'extraction d'une sous-suite, \hat{u}_κ converge dans $C_{loc}^1(\mathfrak{R}^2)$ vers une solution \hat{u} de

$$(34) \quad \left\{ \begin{array}{l} \Delta \hat{u} + \hat{u}(1 - |\hat{u}|^2) = 0 \quad \text{on } \mathfrak{R}^2 \\ |\hat{u}|(0) < 1 \\ \int_{\mathfrak{R}^2} |1 - |\hat{u}|^2|^2 < +\infty. \end{array} \right.$$

Il est démontré dans [BMR] que l'énergie de Higgs des solutions de (34) est quantifiée :

$$(35) \quad \int_{\mathcal{R}^2} |1 - |u|^2|^2 \in 2\pi\mathcal{N}^*.$$

Si donc on revient à l'échelle d'origine, pour un rayon R assez grand, indépendant de κ , et κ assez grand, la contribution à l'énergie de Higgs de la boule $B_{\kappa R}(x_k)$ est au moins 2π à la limite. Ceci est aussi valable pour x_l et donc, comme les deux points s'écartent plus vite que $O(\kappa)$, de (33) on déduit que pour tout $\delta > 0$ indépendant de κ

$$(36) \quad \liminf_{\kappa \rightarrow +\infty} \kappa^2 \int_{B_\delta(p_1)} |1 - |u_\kappa|^2|^2 \geq 4\pi.$$

Par ailleurs, l'identité de Pohozaev sur la boule $B_\delta(p_1)$ déduite de la loi de conservation du tenseur énergie-impulsion combinée avec les convergences du théorème 2.4 donne aisément

$$(37) \quad \lim_{\kappa \rightarrow +\infty} \kappa^2 \int_{B_\delta(p_1)} |1 - |u_\kappa|^2|^2 = 2\pi + o_\delta(1).$$

(36) et (37) entrent en contradiction, donc (33) ne peut pas se produire même pour une sous-suite. Ainsi il existe $\lambda > 0$ indépendant de κ tel que la partie du "mauvais ensemble" \mathcal{M} convergeant vers p_1 peut être recouvert par exactement une seule boule de rayon $\lambda\kappa$. Soit donc x_1 le centre de cette boule et plus généralement x_i le centre de la boule convergeant vers p_i et contenant \mathcal{M} au voisinage de p_i . On peut toujours supposer que $u(x_i) = 0$ car l'indice limite de u_\star est non nul à p_i , mais égal à 1, et donc u doit s'annuler dans la partie de \mathcal{M} qui converge vers p_i qui est contenue dans $B_{\lambda\kappa^{-1}}(x_i)$. L'argument de dilatation précédent appliqué à partir de x_1 nous donne alors que, dans une jauge de Coulomb locale, $\hat{u}_\kappa(z) = u_\kappa(\kappa z + x_1)$ converge dans $C_{loc}^1(\mathcal{R}^2)$, modulo l'extraction d'une sous-suite, vers une solution \hat{u} du problème suivant.

$$(38) \quad \left\{ \begin{array}{l} \Delta \hat{u} + \hat{u}(1 - |\hat{u}|^2) = 0 \quad \text{dans } \mathcal{R}^2 \\ |\hat{u}(0)| = 0 \\ \int_{\mathcal{R}^2} |1 - |\hat{u}|^2|^2 < +\infty \\ \text{ind}(\hat{u}, +\infty) = +1. \end{array} \right.$$

(Dans [BMR] il est démontré que les trois premières lignes de (38) permettent de déduire que $|\hat{u}|$ converge uniformément vers 1 à l'infini et donc que l'indice de \hat{u} , $\text{ind}(\hat{u}, +\infty)$ y est bien défini.) (38) est l'équation, dite du "profil", des vortex de Ginzburg-Landau. Le problème de la multiplicité des solutions de (38) n'est pas un problème standard en équations elliptiques non-linéaires ; en effet \hat{u} n'est pas à valeurs réelles mais complexes et les approches classiques utilisant le principe du maximum pour démontrer une éventuelle symétrie de la solution ne peuvent s'appliquer ici.

P. Mironescu dans [Mi] a donné la démonstration suivante de l'unicité du "profil" (modulo les rotations).

Il est assez standard de vérifier qu'il existe une unique solution de (38) de la forme $\rho(r)e^{i\theta}$ ((r, θ) sont les coordonnées polaires sur \mathbb{R}^2). Divisons \hat{u} , solution quelconque de (38), par cette solution axialement symétrique. Le rapport $w = \hat{u}/\rho(r)e^{i\theta}$ est point critique sur $\mathbb{R}^2 \setminus \{0\}$ de la fonctionnelle

$$\mathcal{E}(w) = \int \rho^2 |\nabla w|^2 + 2 \frac{\rho^2}{r} \left(iw, \frac{\partial w}{\partial \theta} \right) + \rho^4 |1 - |w|^2|^2.$$

w est donc point critique pour l'action infinitésimale des dilatations $r \frac{\partial}{\partial r}$ sur $\mathbb{R}^2 \setminus 0$. Ceci donne alors l'identité de Pohozaev suivante

$$(39) \quad 0 = \int_{\mathbb{R}^2} r \frac{\rho'}{\rho} \left| \frac{\partial w}{\partial r} \right|^2 + \frac{r \rho \rho' + \rho^2}{2} |1 - |w|^2|^2$$

(l'indice 1 à l'infini étant utilisé pour faire disparaître des termes de bords à l'infini). Une étude simple du module de la solution radiale ρ permet de conclure que $\rho' > 0$. L'identité (39) nous dit alors que $|w| = 1$, $\frac{\partial w}{\partial r} = 0$ et donc $\hat{u} = \rho(r)e^{i(\theta+\alpha)}$ où α est une constante. Ce qui prouve l'unicité du profil modulo les rotations.

Comme \hat{u}_κ converge (modulo l'extraction d'une sous-suite) dans $C_{loc}^1(\mathbb{R}^2)$ vers ce profil (unique modulo les rotations), que d'autre part \hat{u} s'annule exactement à 0 et y est de rang maximal, on vérifie alors que sur $B_\lambda(0)$ $\hat{u}_\kappa^{-1}(\{0\}) = \{0\}$. Donc, sur la boule $B_\delta(p_1)$ u s'annule en un point unique.

Nous avons jusqu'ici utilisé le caractère minimal du couple (u_κ, A_κ) afin d'avoir la convergence du théorème 2.4 et l'indice 1 de u_\star à chaque p_j . Au vu de l'extension des résultats 2.4 et 2.6 aux configurations critiques (u_κ, A_κ) quelconques sous les niveaux d'énergie $O(\log \kappa)$ décrite à la fin de la section 2.2, nous avons démontré la proposition suivante.

PROPOSITION 2.7. — *Soit (u_κ, A_κ) une suite de configurations $(\kappa \rightarrow +\infty)$ solutions des équations de Ginzburg-Landau (1), vérifiant la majoration d'énergie $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$ et convergeant vers un couple section-connexion (u_\star, A_\star) singulier à $p_1 \dots p_Q$. Si l'indice de u_\star à p_j est ± 1 pour tout j , alors il existe κ_0 tel que pour $\kappa > \kappa_0$ on ait*

$$|u_\kappa|^{-1}(\{0\}) = \{x_1 \dots x_Q\} \quad \text{et} \quad x_j - p_j \rightarrow 0 \quad \text{pour tout } j$$

Le rôle de l'énergie renormalisée W . Mettons-nous à nouveau dans les hypothèses de la proposition 2.7. La section u_κ pour κ suffisamment grand intersecte donc exactement Q fois la section nulle avec à chaque fois un indice d'intersection ± 1 . Chacun de ces points d'annulation de $|u_\kappa|$ converge vers un des vortex limite $p_1 \dots p_Q$. Au vu de l'importance du rôle joué par la position du zéro dans l'argument d'unicité de P. Mironescu décrit plus haut, il est légitime de se demander si, de façon générale, deux solutions de (1) dans la limite de London ayant le même lieu d'annulation sont égales (modulo l'invariance de

jauge). Cette affirmation généraliserait la même proposition démontrée plus haut par Jaffe et Taubes dans le cas non interactif au cas fortement répulsif. La réponse à cette question est affirmative sous les mêmes hypothèses que la proposition 2.7.

PROPOSITION 2.8. — [PR] Soient (u_κ, A_κ) et (v_κ, B_κ) deux suites de solutions des équations de Ginzburg-Landau (1) dans la limite de London ($\kappa \rightarrow +\infty$) vérifiant la majoration d'énergie commune $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$, $\mathcal{F}_\kappa(v_\kappa, B_\kappa) = O(\log \kappa)$ et convergeant, toutes deux, vers la même configuration singulière (u_\star, A_\star) . On suppose que l'indice de u_\star à chaque singularité est ± 1 . Alors, il existe κ_0 tel que pour $\kappa > \kappa_0$

$$|u_\kappa|^{-1}(\{0\}) = |v_\kappa|^{-1}(\{0\}) \quad \implies \quad (u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$$

où \simeq désigne l'équivalence de jauge.

La démonstration de la proposition précédente repose, entre autres, sur la généralisation suivante de l'argument de P. Mironescu. Dans une jauge de Coulomb au voisinage d'un vortex limite quelconque p_k , on considère, sous les hypothèses de la proposition, à nouveau le rapport des deux solutions $w = \frac{u}{v}$. Au lieu d'effectuer des dilatations suivant le champ conforme $X = r \frac{\partial}{\partial r}$ centré au zéro commun de u et v qui ne donne dans ce cas aucun résultat, on dilate suivant le champ suivant, construit à partir de la deuxième solution v ,

$$Y = |v|^2 \frac{(iv, *dv)}{|(iv, *dv)|^2}$$

où on identifie le champ avec la 1-forme duale donnée par le produit scalaire. On peut observer que Y coïncide avec le champ usuel de dilatation conforme $X = r \frac{\partial}{\partial r}$ dans le cas où v est à symétrie radiale. L'action de ce champ sur la fonctionnelle dont (w, A) est point critique donne une identité de Pohozaev au voisinage de p_k , qui toutes mises ensemble permettent de conclure, pour κ suffisamment grand que $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$.

L'utilisation de Y plutôt que le champ usuel $X = r \frac{\partial}{\partial r}$ est *a posteriori* naturelle au vu de cette fonctionnelle dont (w, A) est point critique et de la propriété de $|v|$ -conformité de Y (voir [PR]).

La proposition 2.8 est la première étape pour décrire l'espace des solutions sous les niveaux d'énergie $O(\log \kappa)$ dans la limite de London. La deuxième étape consiste à construire une suite (v_κ, B_κ) convergeant vers un (u_\star, A_\star) dont la configuration de vortex associée $p_1 \dots p_Q$ est un point critique quelconque de W (non-dégénéré modulo l'action des isométries de M) et dont les multiplicités d_j sont ± 1 . Une telle construction a été effectuée dans [LL] pour $d_j = +1$ au moyen d'arguments variationnels en établissant un lien entre la topologie des ensembles de niveau de la fonction W et de la fonctionnelle \mathcal{F}_κ . Dans [PR] cette construction est effectuée pour $d_j = \pm 1$ au moyen du théorème d'inversion locale qui a l'avantage d'apporter en plus l'unicité locale. L'argument de linéarisation autour d'une solution approchée est rendu complexe par l'existence d'un "noyau à l'infini" : l'inverse du linéarisé de (1) autour d'un recollement de vortex qui converge vers (u_\star, A_\star) explose dans les normes standard $W^{2,2} - L^2$ lorsque κ tend vers l'infini. Ceci est dû à

l'action du groupe d'invariance de l'équation du profil limite (38) (ici les isométries du plan). On est donc amené à développer l'argument de réduction de Lyapunov-Schmidt qui consiste à travailler dans un premier temps perpendiculairement à ce noyau à l'infini que l'on fait réapparaître dans l'argument non-linéaire final. Cette technique appliquée aux E.D.P. elliptiques a souvent été utilisée dans de nombreux problèmes constructifs de géométrie différentielle comme l'existence de surfaces minimales, l'existence de surfaces à courbures moyenne constantes, le problème de Yamabe, les Monopôles de Yang-Mills... par N. Kapouleas, R. Mazzeo, F. Pacard, R. Schoen, K. Uhlenbeck, S.T. Yau... avec à chaque fois des difficultés nouvelles. (v_κ, B_κ) étant ainsi construite bénéficie d'une unicité locale dans un espace fonctionnel adapté. Si (u_κ, A_κ) est une autre solution de (1), convergeant vers la même configuration limite dont les singularités sont $p_1 \dots p_Q$, on sait (proposition 2.7) que les zéros de $|u_\kappa|$ convergent vers ses singularités et donc sont proches de ceux de $|v_\kappa|$. L'argument de la preuve de la proposition 2.8 est alors converti de façon à obtenir, non plus directement que $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$, mais une estimation suffisante de "l'éloignement" de (u_κ, A_κ) et (v_κ, B_κ) qui, au moyen de l'unicité locale précédemment établie, permet de conclure que $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$. On a donc le théorème suivant.

THÉORÈME 2.9. — [PR] *Soit $(p_1 \dots p_Q)$ un point critique de l'énergie renormalisée W pour des multiplicités $d_1 \dots d_Q$ dans $\{-1, +1\}$. Supposons que ce point critique soit non dégénéré modulo l'action des isométries de M . Soit (u_\star, A_\star) le couple section-connexion singulier associé aux (p_j, d_j) . Soient alors (u_κ, A_κ) et (v_κ, B_κ) deux suites des équations de Ginzburg-Landau (1) vérifiant la majoration d'énergie commune $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$ et $\mathcal{F}_\kappa(v_\kappa, B_\kappa) = O(\log \kappa)$, si leur lieu d'adhérence singulier est la configuration (u_\star, A_\star) , alors pour κ suffisamment grand*

$$(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$$

modulo l'action des isométries de M (\simeq est l'équivalence de jauge).

Commentaires. Le théorème 2.9 nous dit que, sous certaines hypothèses, l'étude des points critiques de la fonctionnelle \mathcal{F}_κ dans la limite de London - problème de dimension infini - se ramène à celle de l'étude des points critiques de la fonction W - problème de dimension fini. Les limites de notre description des solutions des équations de Ginzburg-Landau dans le cas fortement répulsif sont les suivantes. Tout d'abord, afin d'être dans le cadre de l'asymptotique BBH, on s'est restreint aux points critiques sous les niveaux d'énergie $C \log \kappa$. Qu'en est-il d'éventuelles autres solutions de (1) ? Dans le cas $A = 0$ sur un domaine étoilé, il est démontré que toutes les solutions de Ginzburg-Landau sont sous un certain niveau d'énergie $C \log \kappa$. La deuxième restriction dans les hypothèses du théorème 2.9 est la contrainte sur les multiplicités limites d_j devant appartenir à $\{+1, -1\}$. Qu'en est-il des branches de solutions raccordant un point critique de W où certains $|d_j|$ sont plus grands que 1 ? La question reste complètement ouverte. Ni les méthodes variationnelles, dans l'esprit de [LL] où [AB1], [AB2], ni les méthodes de points

critiques, vues plus haut, n'ont permis jusqu'à présent de construire des solutions (autres que les solutions axialement symétriques) convergeant vers un point critique de W avec des multiplicités $d_j \neq \pm 1$. On n'a pas même d'idée précise de l'aspect de telles solutions près des vortex. La compréhension des "profils" limites possibles dans ce cas fait aussi défaut (voir des conjectures à ce sujet dans [OS]). Néanmoins, il semble que le cas $d_j = \pm 1$ soit "générique" au sens suivant : pour une perturbation "générique" de la métrique, un point critique de W ayant des indices de multiplicité plus grande que 1 devrait se transformer en point critique dont les indices sont ± 1 .

Pour finir l'étude de l'énergie libre, mentionnons que les techniques présentées plus haut permettent d'apporter une réponse partielle aux conjectures de Jaffe et Taubes sur tout κ^2 dans le cas fortement répulsif.

THÉORÈME 2.10. — [Ri3] *La conjecture 2.3 est vraie dans la limite de London (κ suffisamment grand) si on remplace "point critique stable" par "minimum dans la classe d'homotopie".*

3. L'INTERACTION AVEC UN CHAMP EXTÉRIEUR ; VERS LES RÉSEAUX D'ABRIKOSOV.

3.1. Introduction

Jusqu'à présent nous avons étudié l'énergie libre de Ginzburg-Landau \mathcal{F}_κ sans le terme d'interaction avec le champ extérieur $2h_e \int dA$. Par ailleurs, nous nous sommes placés sur un tore plat afin d'ignorer, dans une première approche, les effets de bord qui sont pourtant très importants dans les études phénoménologiques, dont il serait très intéressant de retrouver les conclusions par l'étude mathématique elle-même. Enfin, en l'absence de champ extérieur, afin de ne pas travailler avec un état fondamental trivial, nous nous sommes imposés une vortacité globale à travers le choix d'un certain fibré sur notre tore de classe d'Euler non nulle. L'analyse de notre problème modèle nous a permis d'isoler et de comprendre le mécanisme

$$\text{vortacité} \quad \implies \quad \text{formation de vortex}$$

puis de ramener l'étude de l'espace des solutions à celle de la position des vortex gouvernée par l'énergie renormalisée.

Dans sa thèse [Se1], puis dans ses travaux en collaboration avec E. Sandier [SS1] [SS2], S. Serfaty considère la fonctionnelle complète \mathcal{G}_κ incluant le terme d'interaction avec le champ extérieur, sur un domaine borné de κ^2 , Ω simplement connexe. La vortacité est désormais libre et devrait apparaître spontanément en accroissant le champ extérieur et donc l'énergie totale du système ainsi qu'il est observé pour les supraconducteurs de type II ($\kappa > 1/\sqrt{2}$).

Nous nous plaçons toujours dans la limite de London ($\kappa \rightarrow +\infty$) et le champ extérieur h_e sera supposé uniforme (on confondra la 2-forme et la constante correspondante).

3.2. La solution de Meissner

Lorsque le champ extérieur h_e est nul l'état fondamental de \mathcal{G}_κ est clairement atteint par la solution d'état pur $|u| = 0$, $h = *dA = 0$. Lorsque l'on accroît le champ h_e le caractère supraconducteur du matériau s'altère un peu sans pour autant former de vortex ($|u_\kappa| \geq 1/2$ pour κ suffisamment grand). On s'attend à ce que cette configuration sans vortex, stable et minimum absolu pour des champ extérieurs pas trop forts, soit unique. C'est ce que démontre S. Serfaty en combinant des estimations d'énergie et des arguments de convexité.

THÉORÈME 3.1. — [Se3] *Dans la limite de London (κ suffisamment grand), il existe un unique point critique stable de \mathcal{G}_κ sans vortex ($|u_\kappa| \geq 1/2$) qui minimise \mathcal{G}_κ parmi les points critiques sans vortex sous la condition que $h_e = O(\kappa^\alpha)$ où α est une constante positive. Cette solution dite de "Meissner" vérifie en particulier*

$$(40) \quad \begin{cases} -\operatorname{div} \left(\frac{\nabla h}{|u|^2} \right) + h = 0 & \text{dans } \Omega \\ h = h_e & \text{sur } \partial\Omega. \end{cases}$$

Bien évidemment $1/2$ dans l'énoncé du théorème est une constante arbitraire entre 0 et 1. En fait la solution de Meissner reste stable jusqu'à une intensité critique du champ extérieur $h_e = H_{sh} \simeq C\kappa$ dite "super-heating field" (littéralement *champ de super-chauffe*) à partir de laquelle elle cesse d'être stable. Cette bifurcation a été étudiée en détail par H. Berestycki, A. Bonnet et J. Chapman dans [BBC]. Par ailleurs, l'unicité de la solution de Meissner au voisinage du champ H_{sh} est prouvée par A. Bonnet, J. Chapman et R. Monneau dans [BCM].

Il est alors naturel de se demander pour quelle intensité H_{c_1} du champ extérieur h_e la solution de Meissner cesse d'être un minimum absolu de \mathcal{G}_κ . Les calculs d'Abrikosov [Ab] sur tout κ^2 prédisent $H_{c_1} \simeq \frac{\log \kappa}{2}$. Comme nous allons le voir, les travaux de S. Serfaty ont permis de montrer que, suivant une conjecture dans [BR2], H_{c_1} est en fait plus grand sur un domaine borné. Cette conséquence des effets de bords est précisément la suivante : dans le terme principal de H_{c_1} le facteur $1/2$ devant $\log \kappa$ doit être remplacé par $k_1 = 1/2 \max |\xi_0|$ où ξ_0 est la solution de

$$(41) \quad \begin{cases} \Delta \xi_0 = \xi_0 + 1 & \text{dans } \Omega \\ \xi_0 = 0 & \text{sur } \partial\Omega. \end{cases}$$

ξ_0 est aussi solution de $-\Delta^2 \xi_0 + \Delta \xi_0 = 0$. On vérifie aisément par le principe du maximum que $\Delta \xi_0 > 1$ et que $\xi_0 < 0$ dans l'intérieur de Ω et donc $|\xi_0| < 1$. E. Sandier et S. Serfaty démontrent alors le résultat suivant.

THÉORÈME 3.2. — [SS1] *Il existe une constante $C_1 > 0$ telle que si*

$$h_e \leq k_1 \log \kappa - C_1 \log \log \kappa$$

la solution de Meissner est un minimum absolu de \mathcal{G}_κ et il existe C_2 telle que si

$$h_e \geq k_1 \log \kappa + C_2$$

la solution de Meissner n'est plus un minimum absolu de \mathcal{G}_κ .

Il est alors tentant de se situer au voisinage de $k_1 \log \kappa$ et de chercher à observer l'apparition des vortex un par un au fur et à mesure que le champ extérieur s'accroît. Techniquement cela signifie que l'on travaillerait avec le minimum absolu de \mathcal{G}_κ pour lequel on chercherait à recouvrir précisément le "mauvais ensemble" \mathcal{M} , dans l'esprit de l'asymptotique BBH, au moyen d'un nombre uniformément borné de boules de rayon κ^{-1} dont on étudierait alors la valeur minimale (éventuellement 0) et les positions respectives les plus avantageuses du point de vue de l'énergie.

Ce rêve se heurte à une difficulté initiale majeure qui n'a pas été levée jusqu'à présent. L'analyse BBH était fondée sur l'estimation initiale de l'énergie libre $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$ qui se traduit par une majoration *a priori*, indépendante de κ , du nombre de vortex (sachant qu'un vortex isolé "coûte" à peu près $2\pi \log \kappa$). Dans le cas présent, où en plus la vorticit  totale n'est pas contr lee et est devenue une variable du probl me, la majoration *a priori* de l' nergie est

$$(42) \quad |\mathcal{G}_\kappa(u_\kappa, A_\kappa)| = O(\log^2 \kappa) ;$$

on ne peut en d duire pour la partie d' nergie libre une estimation meilleure que $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log^2 \kappa)$. Ceci nous dit donc qu'*a priori* le nombre de vortex   consid rer est $O(\log \kappa)$ (ce qui doit  tre vrai pour un champ $h_e = 2k_1 \log \kappa$, mais pas pour $h_e = k_1 \log \kappa + O(1)$ o  on s'attend plut t   un nombre uniform ment major  de vortex). La difficult  de travailler avec un nombre *a priori* aussi grand de vortex explique que l'on doit quitter le cadre de l'asymptotique BBH et que l'impr cision est alors forc ment plus grande comme l'illustre la diff rence entre $k_1 \log \kappa - C \log \log \kappa$ et $k_1 \log \kappa + C_2$ dans le th or me 3.2 qui est pourtant d j  une tr s jolie r alisation.

La d monstration du th or me 3.2 a n anmoins ses fondements dans les arguments de l'asymptotique de la section pr c dente. Il s'agit d' tablir la meilleure minoration possible de l' nergie totale d'une configuration critique ayant des vortex et montrer que, pour un champ ext rieur suffisamment faible, la solution de Meissner est pr f rable  nerg tiquement. Afin d'obtenir une telle minoration, on commence par renormaliser le champ induit en "extrayant" le champ ext rieur de la fa on suivante d crite dans [BR2]. Soit ξ la solution de $\Delta \xi = h$ sur Ω  gale   0 sur $\partial \Omega$ ($A = *d\xi$ est la jauge de Coulomb de h sur Ω), on note ζ la diff rence $\zeta = \xi - h_e \xi_0$ (o  ξ_0 est donn e par (41)). $*d\zeta$ est la connexion de Coulomb sur tout Ω   laquelle on a retir  l'influence du champ ext rieur. On d montre

sans trop de difficultés que l'énergie de la configuration minimale (u_κ, A_κ) se décompose ainsi :

$$(43) \quad \begin{aligned} \mathcal{G}_\kappa(u, A) &= \mathcal{G}_\kappa(1, h_e * d\xi_0) + \mathcal{F}_\kappa(u, *d\zeta) + \int_\Omega |\Delta\zeta|^2 \\ &\quad - 2h_e \int_\Omega (iu, du) \wedge d\xi_0 + o(1). \end{aligned}$$

On suppose alors $h_e = O(\log \kappa)$ et on procède à une décomposition du domaine dans l'esprit de celle utilisée pour démontrer le théorème 2.6 mais plus complexe dans le cas présent où le "mauvais ensemble" est *a priori* bien plus grand. On va avoir $\Omega = \Omega_s \cup \Omega_v$ où Ω_s et Ω_v sont disjoints et Ω_v va être une union de boules disjointes $\{B_{r_i}(a_i)\}_{i \in I}$ recouvrant \mathcal{M} , le "mauvais ensemble", en contenant le plus d'énergie propre possible sans prendre en compte l'énergie d'interaction des vortex entre eux comme dans la preuve du théorème 2.6. Précisément au moyen d'une technique introduite par R. Jerrard [J], on démontre (voir [SS1]) l'existence d'une famille de boules $\mathcal{B} = \{B_{r_i}(a_i)\}_{i \in I}$ telle que

- $\mathcal{M} \subset \cup_{i \in I} B_{r_i}(a_i)$
- $\int_{B_{r_i}(a_i)} f_\kappa(u, *d\zeta) \geq 2\pi |d_i| (\log \kappa - O(\log \log \kappa))$ où $d_i = \deg\left(\frac{u}{|u|}; \partial B_{r_i}(a_i)\right)$
- $r_i = O(\log^{-6} \kappa)$
- $\text{Card } I = O(\log^2 \kappa)$.

On choisit alors $\Omega_v = \cup_{i \in I} B_{r_i}(a_i)$. Les rayons en $O(\log^{-6} \kappa)$ remplacent les rayons $\delta = O(1)$ de la preuve du théorème 2.6 car le nombre de vortex tend ici vers l'infini *a priori* en $O(\log \kappa)$. Les a_i jouent un peu le rôle de vortex intermédiaire de degré d_i et ce choix de recouvrement donne en particulier

$$(44) \quad \int_\Omega (iu, du) \wedge d\xi_0 = 2\pi \sum_{i \in I} d_i \xi_0(a_i) + o(1).$$

En combinant (43), le choix de Ω_v , (44) et le caractère minimisant de (u, A) pour \mathcal{G}_κ on a

$$(45) \quad 0 \geq 2\pi \sum_{i \in I} |d_i| (\log \kappa + O(\log \log \kappa)) - 4\pi h_e \max |\xi_0| \left(\sum_{i \in I} |d_i| \right)$$

et donc, si $\sum_{i \in I} |d_i| \neq 0$, on doit avoir $h_e \geq k_1 \log \kappa + O(\log \log \kappa)$, ce qui établit la première partie du théorème 3.2. La deuxième partie sera une conséquence du théorème 3.3.

Cette technique d'extraction du champ extérieur puis de décomposition du domaine afin d'optimiser la minoration de l'énergie est un thème récurrent des travaux de S. Serfaty et E. Sandier et apparaît sous des formes différentes pour démontrer l'ensemble des théorèmes ci-dessous.

3.3. Familles de solutions stables à 1, 2, 3...vortex ; le bourgeonnement du réseau.

Abandonnons un temps l'idée de regarder le minimum absolu de \mathcal{G}_κ tout en gardant un champ extérieur au voisinage de $k_1 \log \kappa$.

Dans un son premier travail [Se1], S. Serfaty construit des familles de solutions stables ayant un nombre fini de vortex pour des champs $1 \ll h_e \ll \kappa^\alpha$ (α est une constante positive non explicitée). Devant la difficulté présentée plus haut de mettre en évidence un nombre fini de vortex pour les états fondamentaux eux-mêmes, l'ansatz que fait S. Serfaty est de regarder non plus le minimum absolu (u_κ, A_κ) de \mathcal{G}_κ dans $W^{1,2}(\Omega, \mathbf{c}) \times W^{1,2}(\Omega, \mathbf{r}^2)$ tout entier mais pour (u_κ, A_κ) contenu dans un ensemble suivant.

$$\mathcal{D}_n = \left\{ u \in W^{1,2}(\Omega, \mathbf{c}) \quad \text{tel que} \quad 2\pi n \log \kappa + B \leq F_\kappa(u_T) \leq (2\pi n + 1) \log \kappa \right\},$$

où

$$F_\kappa(u_T) = \int_{\Omega} |du_T|^2 + \frac{\kappa^2}{2} |1 - |u_T|^2|^2.$$

u_T est la fonction u projetée sur le disque unité ($u_T = u$ si $|u| \leq 1$ et $u_T = u/|u|$ sinon) et B est une constante négative qui est fixée au cours de la preuve. L'idée est que, dans la jauge de Coulomb sur Ω , un vortex coûte une énergie $F_\kappa(u_T) \simeq 2\pi \log \kappa$. Ainsi, en se restreignant à \mathcal{D}_n , on contraint *a priori* notre configuration (u, A) à ne pas avoir plus de n vortex et on espère donc se ramener à l'asymptotique BBH afin d'établir la description complète de (u, A) . La difficulté principale de cette approche vient du fait qu'il n'est pas évident du tout qu'un minimum de \mathcal{G}_κ dans \mathcal{D}_n , lorsque l'on a établi son existence, soit dans l'intérieur de \mathcal{D}_n et donc soit une solution des équations de Ginzburg-Landau (1). Admettons que l'on ait prouvé l'existence d'un minimum (u_κ, A_κ) de \mathcal{G}_κ dans \mathcal{D}_n . On est tenté de reprendre des arguments de la section 2 pour recouvrir le "mauvais ensemble" \mathcal{M} ... afin d'établir un développement asymptotique de $\mathcal{F}_\kappa(u_\kappa, A_\kappa)$ semblable à (31) et vérifier que (u_κ, A_κ) est bien dans l'intérieur de \mathcal{D}_n (i.e. $2\pi n \log \kappa + B < F(u_T) < (2\pi n + 1) \log \kappa$) et donc solution de (1). La difficulté de cette approche tient au fait que, tant que l'on ne sait pas que (u, A) est solution de (1), on n'a pas d'estimation de la forme $\|d|u|\|_{L^\infty} = O(\kappa)$, on ne dispose plus du résultat de quantification (19) et l'ensemble \mathcal{M} peut être très diffus, etc. On n'a plus non plus de contrôle de sa taille qui utilisait les identités de Pohozaev dans la section 2. L'idée est alors, non plus de travailler avec (u, A) lui-même, mais de travailler avec son régularisé parabolique v qui minimise

$$\min_{v \in H^1(\Omega, \mathbf{c})} \int_{\Omega} |\nabla v|^2 + \frac{\kappa^2}{2} |1 - |v|^2|^2 + \kappa^{2\gamma} \int_{\Omega} |u_\kappa - v|^2$$

$\gamma > 0$. L'avantage de remplacer u par son régularisé parabolique v est que (v, A) va à la fois être très proche énergétiquement de (u, A) tout en bénéficiant des propriétés dont jouissait le minimum de \mathcal{F}_κ dans la section 2 (comme $|v| \leq 1$, $\|d|v|\| = O(\kappa)$...) qui sont issues de l'équation elliptique vérifiée par v . Cette idée du régularisé parabolique a été

introduite dans le cadre des équations de Ginzburg-Landau par L. Almeida et F. Béthuel dans [AB2] afin de définir une configuration de vortex approchée pour toute fonction u dans la zone d'énergie $F(u) = O(\log \kappa)$. On établit alors un développement asymptotique pour v dans l'esprit de la section 2.2, ce qui permet de déduire que (v, A) (et donc aussi (u, A)) qui lui est énergétiquement très proche est dans l'intérieur de \mathcal{D}_n . L'existence d'un minimum de \mathcal{G}_k dans \mathcal{D}_n fait aussi appel à l'utilisation du régularisé parabolique. On établit ainsi le théorème suivant que l'on présente dans le cas où Ω est un disque centré en 0 : $\Omega = B_R(0)$

THÉORÈME 3.3. — [Se1] *Soit D une constante positive arbitraire, soit $h_e(\kappa)$ une fonction de κ tendant vers $+\infty$ à l'infini et vérifiant $h_e(\kappa) \leq \kappa^\alpha$, où α est une constante positive. Alors il existe une constante positive $\kappa_0(D)$ telle que pour $\kappa \geq \kappa_0(D)$ et tout entier $n \in \mathcal{N}^*$, $n < \frac{D}{\pi}$, il existe une famille de points critiques stables (u_κ, A_κ) de \mathcal{G}_κ , solution des équations de Ginzburg-Landau (1), vérifiant*

- $|u_\kappa|^{-1}(\{0\}) = \{a_1^\kappa \dots a_n^\kappa\}$ où $a_1^\kappa \dots a_n^\kappa$ sont des points isolés de $B_R(0)$.
- Pour tout $j = 1 \dots n$, $\text{ind}(u_\kappa, a_j^\kappa) = +1$.
- La configuration $\tilde{a}_j = a_j \sqrt{h_e(\kappa)}$ converge, modulo l'extraction d'une sous-suite, vers une configuration de n points de κ^2 minimisant

$$(46) \quad W(a_1 \dots a_n) = -2\pi \sum_{i \neq j} \log |a_i - a_j| + 2\pi \ddot{\xi}_0(0) \sum_{i=1}^n |x_i|^2.$$

- Le développement asymptotique de $\mathcal{G}_\kappa(u_\kappa, A_\kappa)$ est

$$(47) \quad \mathcal{G}_\kappa(u_\kappa, A_\kappa) = \mathcal{G}_\kappa(1, h_e * d\xi_0) + 2\pi n \left(\left| \log \kappa - \frac{h_e}{k_1} \right| \right) + \pi(n^2 - n) \log h_e \\ + W(\tilde{a}_1 \dots \tilde{a}_n) + Q_n + o(1),$$

où Q_n ne dépend que de n .

Ce théorème est riche en informations qui doivent être extraites une à une au regard des questions posées dans l'introduction de cet exposé.

Contrairement au cas sans bord de la section 2, les vortex tendent ici vers un même point qui coïncide avec le lieu où $|\xi_0|$ est maximum. S'il y avait eu plusieurs point où $|\xi_0|$ est maximal, les vortex se seraient séparés pour se répartir en groupes de n_j vortex de façon optimale entre ses points (on optimise $\sum_j (n_j^2 - n_j)$).

On peut calculer, en fonction de h_e , le nombre de vortex n qui optimise l'énergie et trouver sous l'ansatz $F(u_T) \leq 2\pi D$ une estimation très précise de H_{c1} . On voit clairement que le nombre optimal de vortex est une fonction croissante de h_e à κ fixé (κ grand). On vérifie aisément l'existence de $k_2 > 0$ tel que, pour $h_e > k_1 \log \kappa + k_2$, la solution de Meissner cesse d'être minimisante parmi ces solutions à un nombre fini de vortex, ce qui prouve la deuxième partie du théorème 2.2. Plus généralement, si on note H_n la valeur du

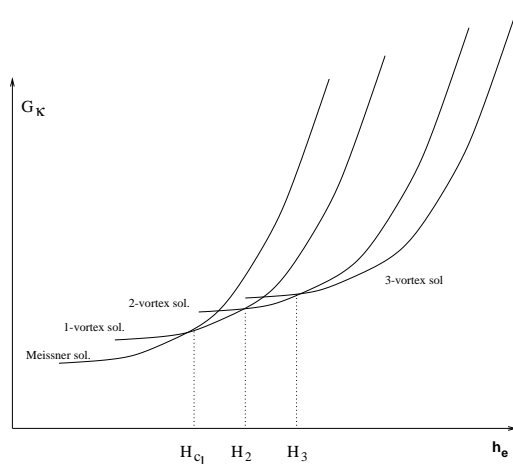


FIGURE 2. Solutions à 1,2,3... vortex

champ où la solution à n vortex devient préférable énergétiquement, on calcule ($n > 1$)

$$(48) \quad H_n - k_1 \log \kappa \simeq (n - 1)k_1 \log \log \kappa.$$

La figure 2 représente graphiquement ces résultats.

Savoir si les solutions à n vortex pour $H_n \leq h_e \leq H_{n+1}$ est un état fondamental sur l'espace $W^{1,2}(B_R; c) \times W^{1,2}(B_R; \kappa^2)$ tout entier est un problème ouvert semble-t-il difficile. Un point important est aussi le caractère stable de ses solutions démontré pour $1 \ll h_e \ll \kappa^\alpha$ et que l'on a vu plus haut pour la solution de Meissner et le champ de super-chauffe. Cette stabilité est responsable de phénomènes d'hystérésis observés dans la réalité. Le caractère stable des solutions ayant un nombre fini de vortex pour des champs inférieurs à H_{c1} avait été observé par Q. Du et F.H. Lin dans [DL].

Bien que l'expression de l'énergie renormalisée W soit relativement simple, les symétries éventuelles de ses états fondamentaux, qui constituent des bourgeonnements de réseaux d'Abrikosov, restent complètement à explorer. Dans [GS], S. Guéron et I. Shafrir ont fait des études de stabilité de configurations symétriques ainsi que des études numériques. Ils ont fait les observations suivantes sur les états fondamentaux probables (voir la figure 3)

- $n \leq 3$: ce sont des polygones réguliers centrés en 0.
- $7 \leq n \leq 10$: ce sont des étoiles régulières (polygone régulier + origine).
- $4 \leq n \leq 6$: les deux types de configurations précédents sont localement minimisants et sont observés.
- $n \geq 11$: ce sont des amas de polygones réguliers concentriques qui "convergent" vers un réseau triangulaire centré en 0 lorsque n devient grand.

3.4. La deuxième renormalisation ; le problème à frontière libre

Dans cette dernière section, nous revenons sur l'étude des états fondamentaux de \mathcal{G}_κ mais pour des champs bien plus forts que H_{c1} , à l'intérieur de la phase mixte (voir figure

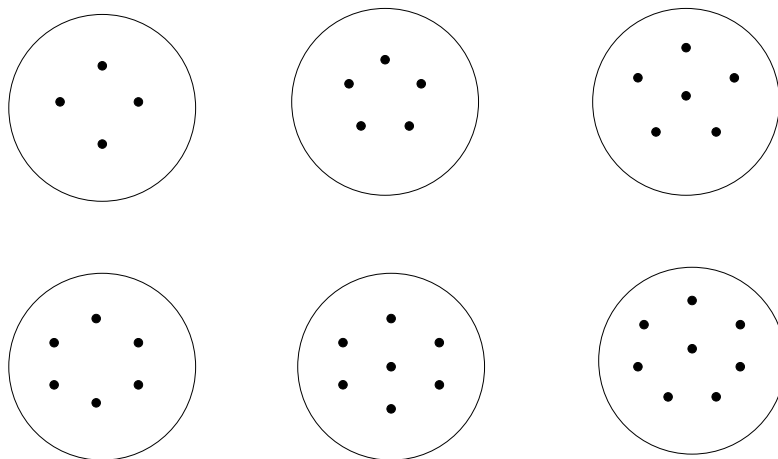


FIGURE 3. Bourgeoisements de réseaux d'Abrikosov

1) et toujours dans la limite de London ($\kappa \rightarrow +\infty$). Dans ce cas, *a priori*, le nombre de vortex N_κ va tendre vers l'infini avec κ . Il faut alors développer des outils adaptés afin de rendre compte de l'ensemble limite occupé par ces vortex dans Ω . La première renormalisation consistait à soustraire à l'énergie minimale l'interaction de chaque vortex sur lui-même ($2\pi \log \kappa$) c'est-à-dire la quantité $2\pi N \log \kappa$. Pour les raisons que nous avons vues plus haut, la précision sur N à l'ordre $O(1)$, à cause de l'insuffisance de l'estimation *a priori* (42), est très difficile à atteindre pour l'état fondamental sous des champs plus grands que $k_1 \log \kappa + O(\log \log \kappa)$. La deuxième renormalisation proposée par E. Sandier et S. Serfaty dans [SS2] consiste à diviser \mathcal{G}_κ par h_e^2 et à étudier la Γ -limite du rapport, au moins pour les états fondamentaux. Le résultat est alors le suivant.

THÉORÈME 3.4. — [SS2] *Soit $h_e(\kappa)$ une fonction positive telle que $\lambda = \lim_{\kappa \rightarrow +\infty} \frac{\log \kappa}{h_e}$ existe. Si $\lambda = 0$, on suppose $h_e(\kappa) = o(\kappa^2)$. Soit k_\star solution du problème suivant*

$$\min \left\{ \begin{array}{l} E(k) = \lambda \mu(\Omega) + \int_{\Omega} |\nabla k|^2 + |k - 1|^2 \\ k \in W^{1,2}(\Omega) \quad k = 1 \quad \text{on } \partial\Omega \\ \mu = -\Delta k + k \quad \text{est une mesure de Radon} \end{array} \right\}$$

Alors k_\star est unique, $\mu = -\Delta h_\star + h_\star$ est la fonction caractéristique du lieu où k_\star est minimal égal à $1 - \frac{\lambda}{2}$ et

$$(49) \quad \frac{h}{h_e} \rightharpoonup k_\star \quad \text{faiblement dans } W^{1,2}(\Omega).$$

Par ailleurs, l'absence de convergence forte est exactement donnée par la convergence de la mesure de défaut suivante

$$(50) \quad \lambda \mu = \lim_{\kappa \rightarrow +\infty} |\nabla(h/h_e)|^2 - |\nabla k|^2.$$

Enfin on a le développement suivant de l'énergie fondamentale renormalisée

$$(51) \quad \lim_{\kappa \rightarrow +\infty} \frac{\mathcal{G}_\kappa(u_\kappa, A_\kappa)}{h_e} = E(k_\star),$$

où (u_κ, A_κ) est une configuration minimale.

REMARQUE 2. — Lorsque λ est égal à 0 (i.e. $h_e \gg \log \kappa$) μ est la fonction caractéristique de Ω tout entier et h/h_e converge fortement vers 1.

Cette absence de convergence forte dans le cas $\lambda > 0$ peut être comprise au sens suivant. En dehors du mauvais ensemble que l'on recouvre par des boules $\mathcal{M} \subset \Omega_v = \cup_{i \in I} B_{\kappa^{-1}}(x_i)$ (supposées disjointes), $h = *dA$ vérifie

$$-\operatorname{div} \left(\frac{\nabla h}{|u|^2} \right) + h = 0 \quad \text{dans } \Omega \setminus \Omega_v.$$

h/h_e est alors très proche de la solution de

$$-\Delta k + k = 2\pi \sum_{i \in I} d_i \frac{\delta_{x_i}}{h_e},$$

où d_i est le degré de $u/|u|$ sur le bord de $B_{\kappa^{-1}}(x_i)$. On rappelle que la deuxième équation de Ginzburg-Landau (1) donne en particulier

$$-\int_{\partial B_{\kappa^{-1}}(x_i)} \frac{1}{|u|^2} \frac{\partial h}{\partial \nu} + \int_{B_{\kappa^{-1}}(x_i)} h = 2\pi d_i$$

et que, par ailleurs, toujours au moyen de cette équation,

$$\frac{\mathcal{G}_\kappa(u, A)}{h_e} \simeq \int_{\Omega \setminus \Omega_v} \frac{1}{|u|^2} \left| d \frac{h}{h_e} \right|^2 + \left| \frac{h}{h_e} - 1 \right|^2 \simeq \int_{\Omega \setminus \Omega_v} \left| d \frac{h}{h_e} \right|^2 + \left| \frac{h}{h_e} - 1 \right|^2.$$

Soit alors μ la limite de la densité de vortex par unité de champ extérieur

$$\mu = \lim_{\kappa \rightarrow +\infty} 2\pi \sum_{i \in I} d_i \frac{\delta_{x_i}}{h_e}.$$

Si on considère l'énergie de la limite k_\star solution de $-\Delta k_\star + k_\star = \mu$ au lieu de la limite de l'énergie

$$\int_{\Omega \setminus \Omega_v} |dk|^2 + |k - 1|^2,$$

on oublie que μ a été obtenu au moyen de sommes de Dirac et on manque alors la partie de l'énergie provenant de l'interaction de chaque Dirac sur lui-même divisée par h_e^2 qui est donc

$$\frac{2\pi N \log \kappa}{h_e^2} \longrightarrow \mu(\Omega) \lim_{\kappa \rightarrow +\infty} \frac{\log \kappa}{h_e} = \lambda \mu(\Omega)$$

La difficulté de l'analyse de cette deuxième renormalisation ne tient pas seulement dans la compréhension heuristique ci-dessus du mécanisme de répartition de l'énergie, mais

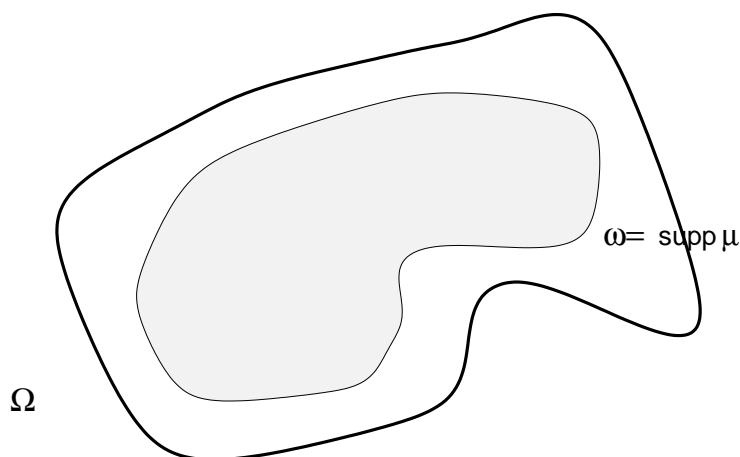


FIGURE 4. Le problème à frontière libre

surtout dans la mise en œuvre d'arguments rigoureux qui permettent d'établir le théorème 3.4. La preuve est assez dans l'esprit de celle du théorème 3.2 où, afin d'obtenir une minoration optimale de l'énergie, on procède à une décomposition du domaine qui sépare l'énergie propre des vortex du reste de l'énergie. Cette décomposition de \mathcal{M} repose sur un recouvrement du "mauvais ensemble" plus affiné que celui de la preuve du théorème 3.2 qui utilise la méthode de grossissement de boules introduites dans [Sa].

Ainsi que le décrit le petit raisonnement ci-dessus, μ est la densité limite de vortex par unité de champ extérieur appliqué. Elle est soit maximale égale à 1 dans le sous-domaine ω de Ω où $k_\star = 1 - \frac{\lambda}{2}$, soit égale à zéro dans son complémentaire (voir la figure 4). Le problème vérifié par k_\star est un problème à frontière libre au sens où la connaissance de ω détermine de façon unique k_\star . C'est un problème d'obstacle désormais assez classique abordé en particulier dans [R]. Il a été démontré par A. Bonnet et R. Monneau dans [BM] que $\partial\omega$ est régulier pour presque toutes les valeurs de λ et que, lorsque c'est le cas, ω est déterminé par l'existence de la solution k_\star du problème suivant

$$(52) \quad \begin{cases} -\Delta k_\star + k_\star = 0 & \text{dans } \Omega \setminus \omega \\ k_\star = 1 - \frac{\lambda}{2} & \text{dans } \omega \\ \frac{\partial k_\star}{\partial \nu} = 0 & \text{sur } \partial\omega \\ k_\star = 1 & \text{sur } \partial\Omega \end{cases}$$

3.5. Conclusion

Qu'en est-il des réseaux d'Abrikosov ? Il serait intéressant, dans l'approche précédente, de se placer à l'échelle $1/\sqrt{h_e}$ - distance minimale moyenne relative entre 2 vortex - et tenter alors de comprendre la limite $\kappa \rightarrow +\infty$. On peut espérer dans cette asymptotique

mettre en évidence une énergie renormalisée infinie sur des configurations dénombrables de points de κ^2 qui gouverne l'agencement des vortex entre eux et tenter de comprendre les états fondamentaux de cette énergie en se restreignant, dans une première approche, aux réseaux périodiques. La première difficulté de cette analyse repose sur la compréhension des termes d'ordre inférieurs dans le développement de l'énergie (51). Le premier terme ne donne que la densité moyenne de vortex et ne voit pas leurs positions relatives à $O(1/\sqrt{h_e})$ près.

Dans sa thèse, M. Dutour [D] adopte une approche différente que les limites de cet exposé nous empêchent de développer plus longuement. On se place sur un tore plat T quelconque vu comme une cellule élémentaire d'un réseau de κ^2 dont la taille et la géométrie sont des variables du problème. Sur cette cellule T que l'on considère occupée par un seul vortex (on fixe un fibré en droite complexe E sur T de classe d'Euler égale à 1), on étudie les minimums de la fonctionnelle de Ginzburg-Landau \mathcal{G}_κ sous l'action d'un champ extérieur supposé uniforme h_e . Lorsque le champ extérieur augmente, l'accroissement de la densité de vortex se traduit, dans ce modèle, par une diminution de la taille de la cellule T , etc. M. Dutour engage ainsi une description très complète du diagramme de phase (figure 1) notamment au voisinage de H_{c2} où il parvient à démontrer l'existence de deux solutions exactement ; une avec vortex (rapprochés de $\simeq \kappa$) et l'autre d'état normal $|u| = 0$, $dA = h_e$. Il établit alors le caractère minimisant ou non de chacune de ces deux solutions suivant h_e et κ et rend compte avec une très grande précision de la bifurcation d'Abrikosov depuis la valeur critique $\kappa = 1/\sqrt{2}$. L'optimalité de la géométrie de tel ou tel réseau n'y est que partiellement abordée, mais cette approche semble ouvrir une voie vers une compréhension rigoureuse de la disposition relative des vortex dans l'état fondamental (très loin du bord qui a disparu dans ce modèle) et éventuellement du caractère énergétiquement favorable des réseaux triangulaires d'Abrikosov.

BIBLIOGRAPHIE

- [Ab] A. Abrikosov – *On the magnetic properties of superconductors of the second type* Soviet Phys. JETP 5, 1174-1182, (1957).
- [Af] A. Aftalion – *On the minimizers of the Ginzburg-Landau energy for high kappa : the one-dimensional case*, European J. Appl. Math. 8, 331-345 (1997).
- [AT] A. Aftalion and W.C. Troy – *On the solutions of the one dimensional Ginzburg-Landau Equations*, Prépublication LMENS (1998).
- [AB1] L. Almeida and F. Béthuel – *Multiplicity results for the Ginzburg-Landau equation in presence of symmetries*, Houston J. Math. 23, 733-764 (1997).
- [AB2] L. Almeida and F. Béthuel – *Topological methods for the Ginzburg-Landau equations*, J. Math. Pures Appl. 77, 1-49 (1998).

- [BBC] H. Bérestycki, A. Bonnet and J. Chapman – *A semi-elliptic system arising in the theory of type-II superconductivity*, Comm. Appl. Non-linear Anal., **1**, 1-21 (1994).
- [BC] M. S. Berger and Y. Y. Chen, *Symmetric vortices for the Ginzburg-Landau equations of superconductivity and the nonlinear desingularization phenomenon*, Journal of Funct. Anal. **82**, 259-295 (1989).
- [BBH0] F. Béthuel, H. Brézis, and F. Hélein, *Asymptotics for the minimization of a Ginzburg-Landau functional*, Calc. Var. Partial Differ. Equ. **1**, 2, 123-148 (1993).
- [BBH] F. Béthuel, H. Brézis and F. Hélein, –*Ginzburg-Landau vortices*, Birkhäuser (1994).
- [BR1] F. Béthuel and T. Rivière, –*Vortices for a variational problem related to superconductivity*, Ann. Inst. Henri Poincaré, Anal. non linéaire **12**, 3, 243-303, (1995).
- [BR2] F. Béthuel et T. Rivière – *Vorticité dans les modèles de Ginzburg-Landau pour la supraconductivité*, Seminaire EDP de l'École Polytechnique, exposé XVI, (1994).
- [Bog] E.B. Bogomol'nyi – Soviet J. Nucl. Phys., **24**, 449 (1976).
- [BH1] C. Bolley and B. Helffer – *Rigorous results on Ginzburg-Landau models in a film submitted to an exterior parallel magnetic field I and II*, Non Linear Stud. **3**, 1-29 and 121-152, (1996).
- [BH2] C. Bolley and B. Helffer – *The ginzburg-Landau equations in a semi-infinite superconducting film in the large κ limit* European J. Appl. Math. **8**, 347-367 (1997).
- [BCM] A. Bonnet, S.J. Chapman and R. Monneau – *Convergence of Meissner minimizers of the Ginzburg-Landau energy of superconductivity as $\kappa \rightarrow +\infty$* , prépublication (1999).
- [BM] A. Bonnet and R. Monneau – *Existence of a smooth free-boundary in a superconductor with a Nash-Moser inverse function theorem argument*, prépublication (1999).
- [BBM] J. Bourgain, H. Brézis and P. Mironescu – *Lifting properties of Sobolev maps*, en préparation.
- [BMR] H. Brézis, F. Merle and T. Rivière – *Quantization effects for $-\Delta u = u(1 - |u|^2)$ in \mathbb{R}^2* , Arch. Rat. Mech. Analysis. **126**, 123-145 (1994).
- [DL] Q. Du and F.H. Lin – *Ginzburg-Landau vortices : dynamics, pinning and hysteresis*, Siam J. Math. Anal. **28**, 1265-1293 (1997).
- [D] M. Dutour – *Bifurcation vers l'état d'Abrikosov et diagramme de phase*, thèse de l'Université Paris-Sud, Orsay (1999).
- [dG] P.-G. De Gennes – *Superconductivity of Metal and Alloys*, Benjamin, New-York and Amsterdam (1966).
- [GS] S. Gueron and I. Shafrir – *On a discrete variational problem involving interacting particles*, à paraître dans SIAM J. Appl. Math. (1999).
- [JT] A. Jaffe and C. Taubes – *Vortices and Monopoles*, Birkhäuser (1980).

- [J] R. Jerrard – *Lower bounds for generalized Ginzburg-Landau functionals*, SIAM J. Math. Anal. **30**, 721-746 (1999).
- [LL] F.H. Lin and T.C. Lin – *Minimax solutions of the Ginzburg-Landau equations*, Selecta Mathematica, New Series **3**, 99-113 (1997).
- [LR] F.H. Lin and T. Rivière – *Complex Ginzburg-Landau equations and codimension 2 minimal surfaces*, J.E.M.S. **1**, 3, (1999).
- [Mi] P.Mironescu – *Les minimiseurs locaux pour l'équation de Ginzburg-Landau sont à symétrie radiale*, C.R. Acad. Sci. Paris, Ser 1, **323**, 593-598 (1996).
- [OS] Y.N. Ovchinnikov and I.M. Sigal – *The Ginzburg-Landau equation I. Static vortices*, Partial differential equations and their applications (Toronto, ON, 1995), 199-220, CRM Proc. Lecture Notes **12**, Amer. Math. Soc., Providence, RI (1997).
- [PR] F. Pacard and T. Rivière – *Linear and non-linear aspects of vortices à paraître chez Birkhäuser* (1999-2000).
- [Qin] J. Qing – *Renormalized energy for Ginzburg-Landau vortices on closed surfaces* Math. Z. **225**, 1-34 (1997).
- [Ri1] T. Rivière – *Asymptotic analysis for the Ginzburg-Landau equations*, ETH-Zürich course january 1997, Boll U.M.I. (1999).
- [Ri2] T. Rivière – *Line vortices in the $U(1)$ -Higgs Model*, COCV **1**, 77-167 (1995).
- [Ri3] T. Rivière – *Some progress towards Jaffe and Taubes Conjectures*, prépublication (1999).
- [R] J.F. Rodrigues – *Obstacle problems in mathematical physics*, Mathematical Studies, North Holland (1987).
- [SST] D. Saint-James, G. Sarma and E.J. Thomas – *Type-II Superconductivity*, Pergamon Press (1969).
- [Sa] E. Sandier – *Lower bounds for the energy of unit vector fields and applications*, J. Funct. Anal. **152**, 379-403 (1998).
- [SS1] E. Sandier and S. Serfaty – *Global minimizers for the Ginzburg-Landau functional below the first critical magnetic field*, à paraître dans les annales de l'Inst. Henri Poincaré, Analyse non-linéaire (1999).
- [SS2] E. Sandier and S. Serfaty – *A rigorous derivation of a free-boundary problem arising in superconductivity*, prépublication CMLA (1999).
- [Se1] S. Serfaty – *Etude mathématique de l'équation de Ginzburg-Landau de la supraconductivité* Thèse de l'Université Paris-Sud, Orsay (1999).
- [Se2] S. Serfaty – *Local minimizers for the Ginzburg-Landau energy near critical magnetic field I and II*, Comm Contemporary Mathematics, **1**, n 2 et 3, 213-254 (1999)
- [Se3] S. Serfaty – *Stable configurations in superconductivity : uniqueness, multiplicity and vortex-nucleation*, à paraître dans Arch. Rat. Mech. Anal. (1999).

- [St] M. Struwe, *On the asymptotic behavior of minimizers of the Ginzburg-Landau model in 2-dimensions*, J. Diff. Int. Equations **7** (1994), 1613-1624 and Erratum in J. Diff. Int.
- [Ti] M. Tinkham – *Introduction to Superconductivity*, 2nd edition, McGraw-Hill (1996)
- [Uh] K. Uhlenbeck – *Connections with L^p bounds on curvature*, Comm. Math. Phys. **83**, 31-42 (1982)

Tristan RIVIÈRE

Courant Institut of Mathematical Sciences
251, Mercer Street
New-York, NY 10012, USA

et

Centre National de la Recherche Scientifique
France

E-mail : `triviere@cims.nyu.edu`

**PROGRÈS RÉCENTS SUR LA CONJECTURE DE BAUM-CONNES.
CONTRIBUTION DE VINCENT LAFFORGUE**

par **Georges SKANDALIS**

Cet exposé fait de nouveau le point sur la conjecture de Baum-Connes après celui de Pierre Julg ([24]) qui rendait compte d'une avancée spectaculaire obtenue par Higson et Kasparov ([19]). Une nouvelle barrière vient juste d'être franchie par V. Lafforgue : celle de la propriété T de Kazhdan. Le but ici est d'exposer les diverses étapes du travail de Lafforgue. En fin d'exposé, nous donnerons quelques autres résultats en relation avec la conjecture de Baum-Connes et celle de Novikov obtenus récemment par G. Yu ([51]), N. Higson ([18]) et quelques autres.

La conjecture de Baum-Connes prédit la K -théorie de la C^* -algèbre (réduite) d'un groupe discret G et, plus généralement, d'un groupe localement compact, et même d'un groupoïde localement compact (par exemple de la C^* -algèbre associée à un feuilletage) *etc.* Cette conjecture a plusieurs conséquences : l'injectivité du morphisme de Baum-Connes a des implications géométriques importantes - la plus fameuse étant la conjecture de Novikov sur les « hautes signatures » ; sa surjectivité a des applications fondamentales dans la théorie des C^* -algèbres (*e.g.* une conjecture de Kadison, déjà ancienne, sur les idempotents).

Malgré les efforts de plusieurs mathématiciens, une barrière subsistait pour la preuve de cette conjecture : la propriété T de Kazhdan (*cf.* [32], [12]), qui empêche certaines idées de démonstration d'aboutir et rend fausses certaines généralisations : en effet, cette conjecture est alors fautive si la C^* -algèbre pleine remplace la réduite (*cf.* [25], voir aussi [10]) ; la généralisation de cette conjecture au bifoncteur de Kasparov (qui généralise la K -théorie) peut elle-même être fautive (*cf.* [44]). Au-delà de cette barrière, les seuls cas connus avant le travail de Lafforgue étaient les groupes de Lie réels réductifs (A. Wassermann [50]) ou p -adiques (Baum-Higson-Plymen [6]) utilisant la description complète de la C^* -algèbre de ces groupes, mais aucun groupe discret infini possédant la propriété T n'avait pu être obtenu.

Lafforgue établit la conjecture de Baum-Connes pour beaucoup de groupes localement compacts ayant la propriété T : tout groupe de Lie semi-simple réel ou réductif p -adique ET les sous-groupes discrets cocompacts des groupes de Lie de rang réel 1 ou de SL_3 d'un corps local.

Dans tous ces cas, l'injectivité du morphisme de Baum-Connes était connue (travaux de Miščenko, Solovjev, Kasparov...). En fait, il y a un élément noté γ par Kasparov dans l'anneau de représentations (défini par Kasparov) $KK_G(\mathbf{C}, \mathbf{C})$ dont la différence avec 1, mesure la non surjectivité du morphisme de Baum-Connes. On doit montrer que $\gamma = 1$ en tant qu'endomorphisme de la K -théorie de la C^* -algèbre réduite $C_r^*(G)$ de G . Remarquons que, par la propriété T , γ lui-même n'est pas égal à 1 dans l'anneau $KK_G(\mathbf{C}, \mathbf{C})$ et même dans l'anneau $KK(C_r^*(G), C_r^*(G))$ (du moins si G est un sous-groupe discret cocompact de $Sp(n, 1)$ - cf. [44]).

La démonstration de Lafforgue a trois étapes. La première est sa principale contribution conceptuelle ; la seconde est - de loin - la plus difficile techniquement ; dans la troisième, il utilise des travaux de Jolissaint dans le cas $Sp(n, 1)$ et de Ramagge-Robertson-Steger dans le cas de SL_3 d'un corps local non archimédien ; enfin, il adapte la méthode de Ramagge-Robertson-Steger pour obtenir aussi le cas des sous-groupes discrets cocompacts de $SL_3(\mathbf{R})$ et $SL_3(\mathbf{C})$.

Plus explicitement :

1. Dans la première étape, Lafforgue construit un bifoncteur $KK^{\text{ban}}(A, B)$ pour des algèbres de Banach, en remplacement de la théorie bivariante de Kasparov. Il s'agit là d'une construction nouvelle et très intéressante. Elle suit d'assez près celle de Kasparov. Cependant, les C^* -modules hilbertiens qui sont à la base du travail de Kasparov, sont ici remplacés par des paires de modules banachiques en dualité (les modules hilbertiens des C^* -algèbres sont en ce sens autoduaux). Malheureusement, le produit de Kasparov ne se généralise pas directement à ce cadre. Cependant, Lafforgue parvient à étendre plusieurs constructions de Kasparov dans cette «*KK*-théorie banachique» : il démontre qu'elle opère sur la K -théorie, construit une KK -théorie banachique équivariante par un groupe localement compact, ainsi qu'un analogue de l'homomorphisme de descente $j_G : KK_G(A, B) \rightarrow KK(A \rtimes G, B \rtimes G)$ de Kasparov.

2. Dans la deuxième étape il établit l'égalité $\gamma = 1$ dans sa KK -théorie banachique équivariante pour tout groupe localement compact agissant de façon continue, propre et isométrique sur :

- une variété riemannienne complète, connexe, simplement connexe à courbure sectionnelle négative ou nulle (cette classe contient tous les groupes de Lie réels et tous leurs sous-groupes fermés).
- sur un immeuble affine (cette classe contient tous les groupes de Lie p -adiques et tous leurs sous-groupes fermés).

C'est la partie la plus technique. Il construit une homotopie qui utilise :

- dans le cas réel des espaces de Sobolev de formes L^p sur des variétés de courbure sectionnelle négative ;
- dans le cas p -adique une combinatoire assez élaborée (ainsi que des espaces ℓ^p).

A l'aide des parties 1 et 2, Lafforgue établit la conjecture de Baum-Connes à valeurs dans la sous-algèbre $L^1(G)$ de $C_r^*(G)$ (ce qui prouve une conjecture de Bost),

ainsi que toutes les complétions de $C_c(G)$ pour des «normes inconditionnelles» (dans le langage de Lafforgue) *i.e.* une norme d'algèbre N satisfaisant $N(f) \leq N(g)$ si $f, g \in C_c(G)$ satisfont, pour tout $x \in G$, $|f(x)| \leq |g(x)|$ (notons cependant que la norme de C_r^* n'est malheureusement pas «inconditionnelle»).

3. Pour prouver la conjecture de Baum-Connes, il suffit de construire une «complétion inconditionnelle» de $C_c(G)$, qui soit une sous-algèbre A de $C_r^*(G)$ avec la même K -théorie que $C_r^*(G)$. Pour cela, il suffit que A soit stable par calcul fonctionnel holomorphe dans $C_r^*(G)$. Une telle sous-algèbre A existe :

- si G est un groupe de Lie semi-simple, Lafforgue montre qu'une variante de l'algèbre de Harish-Chandra convient (*cf.* [36]) ;
- si G est un groupe hyperbolique *e.g.* un sous-groupe discret cocompact dans $Sp(n, 1)$ (travaux de Haagerup pour les groupes libres, de Jolissaint pour les sous-groupes cocompacts des groupes de Lie de rang 1, et de de la Harpe pour un groupe hyperbolique quelconque [17]) ;
- si G est discret cocompact dans SL_3 d'un corps local non archimédien (ce résultat est dû à Ramagge, Robertson et Steger [41]) ;
- si G est discret cocompact dans $SL_3(\mathbf{R})$ ou $SL_3(\mathbf{C})$: ce résultat est dû à Lafforgue [34].

Dans ces trois derniers cas, l'algèbre A est donnée par des conditions de décroissance rapide au sens ℓ^2 .

Nous commencerons cet exposé par quelques rappels, notamment sur les C^* -algèbres de groupes et la théorie de Kasparov. Ensuite, nous rappellerons la conjecture de Baum-Connes et ses généralisations. Nous renvoyons à ce sujet à l'excellent exposé de Pierre Julg [24]. Les trois paragraphes suivants seront consacrés aux trois étapes du travail de Lafforgue citées ci-dessus. Dans le dernier paragraphe, nous décrivons de très récents résultats de Yu ([51]) et Higson ([18]) reliés à la conjecture de Baum-Connes.

Ajout - mars 2000. — Depuis que cet exposé a été écrit, des contre-exemples à certaines généralisations de la conjecture de Baum-Connes ont été donnés (*cf.* [20]). Ceux-ci sont basés sur des idées de M. Gromov (*cf.* [15]).

Je tiens à remercier Vincent Lafforgue pour de nombreuses explications sur ses travaux. Je le remercie aussi, ainsi que Saad Baaï, Jean Renault et Alain Valette, pour une lecture attentive de ce manuscrit. Un grand merci enfin à Alberto Arabia pour son aide amicale.

1. RAPPELS ET NOTATIONS : C^* -ALGÈBRES DE GROUPES, THÉORIE DE KASPAROV

1.1. Algèbres de Banach ; C^* -algèbres

— Toutes les algèbres de Banach que nous considérons dans cet exposé sont complexes. Si A est une algèbre de Banach, on supposera toujours que sa norme vérifie $\|ab\| \leq \|a\|\|b\|$, pour tout $a, b \in A$. On ne suppose pas en général que nos algèbres de Banach possèdent un élément unité. Si A est une algèbre de Banach, on note \tilde{A} l'algèbre de Banach unifère construite de la façon suivante : en tant qu'espace vectoriel \tilde{A} est isomorphe au produit $A \times \mathbf{C}$; sa norme est donnée par $\|(a, \lambda)\| = \|a\| + |\lambda|$ (pour $a \in A$ et $\lambda \in \mathbf{C}$) et son produit est donné par $(a, \lambda)(b, \mu) = (ab + \lambda b + \mu a, \lambda\mu)$ (pour $a, b \in A$ et $\lambda, \mu \in \mathbf{C}$). On identifie A avec son image dans \tilde{A} par l'homomorphisme $a \mapsto (a, 0)$; l'élément $(0, 1)$ est l'unité de \tilde{A} .

— Un A -module banachique à droite (*resp.* à gauche) est un espace de Banach E muni d'une action à droite (*resp.* à gauche) de A telle que, pour tout $x \in E$ et tout $a \in A$, on ait $\|xa\| \leq \|x\|\|a\|$ (*resp.* $\|ax\| \leq \|a\|\|x\|$). Un A -module banachique à droite (*resp.* à gauche) est un \tilde{A} -module banachique à droite (*resp.* à gauche) : il suffit de poser $x(a, \lambda) = xa + \lambda x$ (*resp.* $(a, \lambda)x = ax + \lambda x$) pour $a \in A$, $x \in E$, et $\lambda \in \mathbf{C}$.

Rappelons qu'une *involution* sur une algèbre de Banach est une application antilinéaire, antimultiplicative, involutive et isométrique notée $x \mapsto x^*$. Une *algèbre de Banach involutive* est une algèbre de Banach munie d'une involution.

Une C^* -algèbre est une algèbre de Banach involutive A telle que, pour tout $x \in A$, on ait $\|x^*x\| = \|x\|^2$. Si A est une C^* -algèbre, l'algèbre de Banach obtenue à partir de A en adjoignant une unité est involutive (on pose $(a, \lambda)^* = (a^*, \bar{\lambda})$) ; ce n'est cependant pas une C^* -algèbre. Il existe néanmoins une (unique) norme de C^* -algèbre sur \tilde{A} .

1.2. C^* -algèbres de groupes

Soit G un groupe localement compact. Notons dx une mesure de Haar à gauche sur G et $C_c(G)$ l'algèbre de convolution des fonctions continues à support compact sur G . Rappelons que la convolée $f \star g$ de $f, g \in C_c(G)$ est donnée par la formule

$$f \star g(x) = \int_G f(y)g(y^{-1}x) dy$$

pour tout $x \in G$.

Par ailleurs, on définit une involution $f \mapsto f^*$ sur $C_c(G)$ en posant $f^*(x) = \Delta(x^{-1})\overline{f(x^{-1})}$ où Δ désigne la fonction module sur G (telle que, pour tout $f \in C_c(G)$, on ait $\int f(x^{-1}) dx = \int f(x)\Delta(x) dx$).

Il y a plusieurs façons de compléter $C_c(G)$ en une algèbre de Banach, c'est-à-dire des normes sur $C_c(G)$ telles que, pour tout $f, g \in C_c(G)$ on ait $\|f \star g\| \leq \|f\|\|g\|$. Nous en utiliserons (au moins) 3 :

1. La plus simple est de prendre la norme $\| \cdot \|_1$. On obtient l'algèbre de Banach $L^1(G)$.
2. Une autre façon est de considérer $C_c(G)$ comme opérant sur $L^2(G)$ par convolution : pour $f \in C_c(G)$, on note $\lambda(f)$ l'application de $L^2(G)$ dans lui-même donnée par $g \mapsto \lambda(f)(g) = f \star g$. On note $\|f\|_r$ la norme d'opérateur de $\lambda(f)$ (dans $L^2(G)$). Le complété de $C_c(G)$ pour cette norme est une C^* -algèbre appelée la C^* -algèbre réduite de G et notée $C_r^*(G)$. De façon équivalente, $C_r^*(G)$ s'identifie avec l'adhérence normique de $\lambda(C_c(G))$ dans l'algèbre des opérateurs de l'espace hilbertien $L^2(G)$. On peut remarquer que :
 - pour $f \in C_c(G)$, on a $\|\lambda(f)\| \leq \|f\|_1$, donc la représentation λ s'étend à $L^1(G)$, et que
 - la représentation λ est *involutive* : pour tout $f \in C_c(G)$, on a $\lambda(f^*) = \lambda(f)^*$ où, pour un opérateur T d'un espace hilbertien, T^* désigne l'adjoint de T .
3. Considérons toutes les représentations involutives σ de $C_c(G)$ dans un espace hilbertien H_σ telles que, pour tout $f \in C_c(G)$, on ait $\|\sigma(f)\| \leq \|f\|_1$. Posons $\|f\|_p = \sup_\sigma \|\sigma(f)\|$. On obtient encore une norme ; le complété de $C_c(G)$ pour cette norme est une C^* -algèbre notée $C^*(G)$ et appelée la C^* -algèbre de G ou la C^* -algèbre pleine de G . L'identité de $C_c(G)$ se prolonge évidemment en un morphisme $C^*(G) \rightarrow C_r^*(G)$. Ce morphisme est toujours surjectif. Il est injectif si et seulement si le groupe G est *moyennable*.

1.3. Produits croisés

Soient G un groupe localement compact et A une algèbre de Banach. Une *action de G dans A* est un homomorphisme α de G dans le groupe des automorphismes isométriques de A , tel que, pour tout $a \in A$, l'application $x \mapsto (\alpha(x))(a)$ soit continue. Le plus souvent, pour $x \in G$ et $a \in A$ l'élément $(\alpha(x))(a) \in A$ s'écrira juste $x.a$.

Soit α une action d'un groupe G dans une algèbre de Banach A . L'espace de Banach $L^1(G; A)$ est muni du produit de convolution donné par la formule

$$f \star g(x) = \int_G f(y) y.(g(y^{-1}x)) dy$$

pour tout $f, g \in C_c(G; A) \subset L^1(G; A)$ et $x \in G$.

Si A est une C^* -algèbre et α respecte l'involution $(x.(a^*)) = (x.a)^*$ pour tout $a \in A$ et tout $x \in G$, alors l'algèbre de Banach $L^1(G; A)$ est naturellement munie d'une involution par la formule $(f^*)(x) = \Delta(x^{-1})x.(f(x^{-1}))^*$. Généralisant la construction de $C_r^*(G)$ et $C^*(G)$, on obtient par complétion de $L^1(G; A)$ deux C^* -algèbres $A \rtimes_r G$ et $A \rtimes G$ appelées respectivement produit croisé réduit et plein de A par G .

1.4. Théorie de Kasparov

Nous rappelons ici, très sommairement, certaines constructions et propriétés du bifoncteur de Kasparov (*cf.* [28], [29] ; voir aussi les exposés de T. Fack [13] et P. Julg [24]).

1.4.a) Modules hilbertiens

Les A -modules hilbertiens sont les espaces hilbertiens avec les scalaires remplacés par A .

Rappelons qu'un élément a d'une C^* -algèbre A est appelé *positif* s'il est de la forme x^*x ($x \in A$). L'ensemble A_+ des éléments positifs joue le rôle de \mathbf{R}_+ dans \mathbf{C} .

On appelle A -module préhilbertien sur A un A -module à droite E muni d'un « produit scalaire » *i.e.* d'une application sesquilinéaire $E \times E \ni (x, y) \mapsto \langle x, y \rangle \in A$, telle que, pour tout $x \in E$, l'application $y \mapsto \langle x, y \rangle$ est A -linéaire (à droite), satisfaisant $\langle x, y \rangle^* = \langle y, x \rangle$ et $\langle x, x \rangle \in A_+$ pour tout $x \in E$.

Si E est un A -module préhilbertien, l'application $x \mapsto \sqrt{\|\langle x, x \rangle\|}$ est une semi-norme sur E . Si E est séparé et complet pour cette norme on dit que c'est un A -module hilbertien.

1.4.b) Opérateurs ; opérateurs compacts

Soient E, F des A -modules hilbertiens. L'ensemble des opérateurs de E dans F que nous considérons sont les opérateurs $T : E \rightarrow F$ possédant un adjoint, *i.e.* une application $T^* : F \rightarrow E$ satisfaisant $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ pour tout $x \in E$ et $y \in F$. Un tel opérateur est automatiquement linéaire, A -linéaire et continu. L'ensemble des opérateurs de E dans F est noté $\mathcal{L}(E, F)$. L'ensemble $\mathcal{L}(E) = \mathcal{L}(E, E)$, muni de la norme des applications linéaires continues, est une C^* -algèbre.

En particulier, pour tout $x \in E$ et $y \in F$, posons $\theta_{y,x} : z \mapsto y\langle x, z \rangle$. C'est un élément de $\mathcal{L}(E, F)$ et $\theta_{y,x}^* = \theta_{x,y}$. On note $\mathcal{K}(E, F)$ le sous-espace vectoriel fermé de $\mathcal{L}(E, F)$ engendré par les $\theta_{y,x}$. L'ensemble $\mathcal{K}(E) = \mathcal{K}(E, E)$ est un idéal bilatère fermé dans $\mathcal{L}(E)$.

Il sera utile de considérer les modules hilbertiens $\mathbf{Z}/2\mathbf{Z}$ -gradués, c'est à dire les modules hilbertiens $E = E^{(0)} \oplus E^{(1)}$, où $E^{(0)}$ et $E^{(1)}$ sont orthogonaux pour \langle , \rangle . Pour $i \in \mathbf{Z}/2\mathbf{Z}$, on note $\mathcal{L}(E)^{(i)}$ l'ensemble des $T \in \mathcal{L}(E)$ tels que $T(x) \in E^{(i+j)}$ pour tout $j \in \mathbf{Z}/2\mathbf{Z}$ et tout $x \in E^{(j)}$.

1.4.c) Cycles en KK -théorie

Soient A et B des C^* -algèbres. Le groupe de Kasparov $KK(A, B)$ est donné par des cycles et une relation d'équivalence sur ces cycles. Un cycle est un triplet (E, π, F) où E est un B -module hilbertien $\mathbf{Z}/2\mathbf{Z}$ -gradué, $\pi : A \rightarrow \mathcal{L}(E)^{(0)}$ est un homomorphisme involutif (tel que $\pi(a^*) = \pi(a)^*$) et $F \in \mathcal{L}(E)^{(1)}$ est tel que $F = F^*$ et, pour tout $a \in A$, les opérateurs $F\pi(a) - \pi(a)F$ et $(F^2 - 1)\pi(a)$ sont dans $\mathcal{K}(E)$.

La somme des cycles (E, π, F) et (E', π', F') est, avec des notations évidentes, $(E \oplus E', \pi \oplus \pi', F \oplus F')$.

1.4.d) Homotopie

Notons $B[0, 1]$ la C^* -algèbre des applications continues de l'intervalle $[0, 1]$ dans B . Un cycle pour $KK(A, B[0, 1])$ est une famille (E_t, π_t, F_t) de cycles pour $KK(A, B)$: c'est par définition une homotopie entre (E_0, π_0, F_0) et (E_1, π_1, F_1) .

Le groupe de Kasparov $KK(A, B)$ est l'ensemble des classes d'homotopie de cycles. Ce groupe est fonctoriel en A et en B : pour tout A on a un foncteur covariant $B \rightarrow KK(A, B)$ et pour tout B on a un foncteur contravariant $A \rightarrow KK(A, B)$.

1.4.e) *Produit de Kasparov*

Kasparov a défini un produit $KK(A, B) \times KK(B, C) \rightarrow KK(A, C)$ (où A, B, C sont des C^* -algèbres) qui est bilinéaire, naturel et associatif. Le fait que ce produit soit hautement non trivial donne toute sa force à la théorie. Muni de ce produit, $KK(A, A)$ est un anneau. Son élément unité noté 1_A est représenté par $(A, \pi, 0)$ où A est gradué par $A^{(0)} = A$ et $\pi(a)b = ab$.

1.4.f) *Action sur la K -théorie*

Le groupe $KK(\mathbf{C}, A)$ a été identifié par Kasparov avec le groupe de K -théorie $K_0(A)$. Le produit de Kasparov donne en particulier un morphisme de groupes $KK(A, B) \rightarrow \text{Hom}(K_0(A), K_0(B))$.

1.4.g) *Le cas équivariant*

On suppose à présent qu'un groupe localement compact G opère sur les C^* -algèbres A et B . On définit un groupe $KK_G(A, B)$ en demandant que G opère aussi sur les cycles : un cycle pour $KK_G(A, B)$ est encore donné par un triple (E, π, F) , qui est un cycle pour $KK(A, B)$, tel que G opère dans E de façon continue (*i.e.* pour tout $x \in E$ l'application $g \mapsto g.x$ est continue) ; on suppose de plus que, pour tout $x, y \in E$, $a \in A$, $b \in B$ et $g \in G$, on a

$$g.(xb) = (g.x)(g.b), \quad \langle g.x, g.y \rangle = g.(\langle x, y \rangle), \quad g.(\pi(a)(x)) = \pi(g.a)(g.x);$$

enfin, pour tout $a \in A$ et tout $g \in G$ on a $\pi(a)(g.F - F) \in \mathcal{K}(E)$ et $g \mapsto \pi(a)(g.F)$ est continue en norme - où $g.F$ est donné par $(g.F)(x) = g.(F(g^{-1}.x))$. Le groupe $KK_G(A, B)$ est l'ensemble des classes d'homotopie de cycles. Toutes les propriétés se généralisent au cas équivariant ; en particulier, le produit de Kasparov se généralise à ce cadre.

Remarque. — Soient G un groupe discret et (H, π, F) un cycle pour $KK_G(A, \mathbf{C})$. L'action de G dans H et la représentation π de A donnent une représentation du produit croisé $A \rtimes_r G$ dans H . On en déduit que $KK_G(A, \mathbf{C}) = KK(A \rtimes_r G, \mathbf{C})$. Plus généralement si l'action de G dans B est triviale $KK_G(A, B) = KK(A \rtimes_r G, B)$.

1.4.h) *Le morphisme j_G*

Si (E, π, F) est un cycle pour $KK_G(A, B)$, on peut former le produit croisé $E \rtimes_r G$ qui est un $B \rtimes_r G$ -module hilbertien. On obtient aussi une représentation naturelle $\tilde{\pi} : A \rtimes_r G \rightarrow \mathcal{L}(E \rtimes_r G)$; de plus, on associe à F un opérateur $\tilde{F} \in \mathcal{L}(E \rtimes_r G)$. Le triplet $(E \rtimes_r G, \tilde{\pi}, \tilde{F})$ est un cycle pour $KK(A \rtimes_r G, B \rtimes_r G)$. On obtient ainsi un homomorphisme $j_G^r : KK_G(A, B) \rightarrow KK(A \rtimes_r G, B \rtimes_r G)$, compatible avec les produits de Kasparov.

On construit de même un homomorphisme pour les produits croisés pleins $j_G : KK_G(A, B) \rightarrow KK(A \rtimes_r G, B \rtimes_r G)$.

2. LA CONJECTURE DE BAUM-CONNES

La conjecture de Baum-Connes (*cf.* [5]) prédit la K -théorie de $C_r^*(G)$.

Si G est un groupe de Lie connexe semisimple, elle prédit que l'induction de Dirac est un isomorphisme de la K -théorie de $C^*(K)$, *i.e.* l'anneau des représentations de K ⁽¹⁾ - où K est un sous-groupe compact maximal de G , sur celle de $C_r^*(G)$.

Si G est un groupe discret sans torsion, la conjecture prédit que la K -théorie de $C_r^*(G)$ est décrite par les G -indices sur des revêtements \widetilde{M} de groupe G d'une variété compacte M (au sens de [3]).

Pour l'énoncer correctement dans le cas général, nous aurons besoin de revenir sur la notion d'espace classifiant pour un groupe.

2.1. Espaces classifiants

Soit G un groupe discret. L'espace classifiant BG est un espace muni d'un fibré principal ξ_0 de fibre G . De plus, pour tout espace topologique X ⁽²⁾ et tout fibré principal ξ de fibre G , il existe une application continue $f : X \rightarrow BG$, unique à homotopie près, telle que ξ soit isomorphe à $f^*(\xi_0)$. Ces propriétés déterminent BG à équivalence d'homotopie près.

Le groupe G opère sur l'espace total du fibré ξ_0 - traditionnellement noté EG et BG est le quotient EG/G par l'action de G . L'espace EG et l'espace BG sont en général construits comme complexes cellulaires infinis. On peut cependant les supposer localement compacts (à l'aide d'une construction de télescope infini).

L'action de G sur EG est *libre et propre* ⁽³⁾. De plus, supposons que G opère librement et proprement sur un espace localement compact \widetilde{X} . En considérant le fibré principal sur \widetilde{X}/G , on déduit qu'il existe une application continue G -équivariante $g : \widetilde{X} \rightarrow EG$, unique à homotopie G -équivariante près. En ce sens, EG est un G -espace libre et propre final dans cette catégorie.

La conjecture de Baum-Connes utilise un G -espace propre final dans la catégorie des G -espaces propres. Un tel espace existe et est unique à G -homotopie près (*cf.* [5], [31]). On le note \underline{EG} . La construction de cet espace est en fait plus simple que celle de EG :

- un modèle pour \underline{EG} est l'ensemble des mesures de probabilité sur G ;
- un modèle localement compact pour \underline{EG} est l'ensemble des mesures positives sur G de masse totale dans $]\frac{1}{2}, 1]$, muni de la topologie de dualité avec l'algèbre $C_0(G)$ des fonctions continues nulles à l'infini sur G ;
- si G est compact, on peut prendre pour \underline{EG} un espace à un point !

⁽¹⁾ à des problèmes mineurs de « K -orientation» près.

⁽²⁾ On suppose X suffisamment régulier *e.g.* localement compact σ -compact, ou un complexe cellulaire.

⁽³⁾ Autrement dit, l'application $(x, g) \mapsto (x, gx)$ est injective et propre de $X \times G$ dans $X \times X$.

- si G est un groupe de Lie réel connexe, on peut prendre $\underline{EG} = G/K$ où K est un sous-groupe compact maximal ;
- si G est un groupe de Lie réductif p -adique, on prendra pour \underline{EG} la réalisation géométrique de l'immeuble affine associé ;
- si G est discret, une présentation de \underline{EG} via un complexe simplicial localement fini rend possible l'utilisation d'arguments de type Mayer-Vietoris.

2.2. Le groupe de K -théorie topologique de Baum-Connes

Le groupe de K -théorie topologique $K_{0,top}(G)$, défini d'abord dans certains cas particuliers (groupes discrets, groupes de Lie, feuilletages... - cf. [4]), n'a été bien compris en général que récemment (cf. [5], [47]) :

Soit G un groupe localement compact. Le groupe de K -théorie topologique est

$$K_{0,top}(G) = \varinjlim KK_G(C_0(Y), \mathbf{C}),$$

où la limite inductive est prise sur les parties fermées G -invariantes Y de \underline{EG} telles que Y/G soit compact ⁽⁴⁾.

Remarque. — Supposons que le groupe G soit discret. Dans ce cas, $K_{0,top}(G) = \varinjlim KK(C_0(Y) \rtimes G, \mathbf{C})$ (voir remarque en 1.4.g)).

Si de plus G est sans torsion, toute action propre de G est libre. Alors, la C^* -algèbre $C_0(Y) \rtimes G$ est équivalente au sens de Morita à $C(Y/G)$ (en fait si G est infini, $C_0(Y) \rtimes G$ est isomorphe à $C(Y/G) \otimes \mathcal{K}$, où \mathcal{K} désigne l'algèbre des opérateurs compacts sur un espace hilbertien). Comme la KK -théorie est invariante par équivalence de Morita, on en déduit que $K_{0,top}(G) = \varinjlim KK(C(X), \mathbf{C})$ la limite inductive étant prise sur les parties compactes X de BG ; en d'autres termes, $K_{0,top}(G)$ est $K_{0,c}(BG)$, la K -homologie à support compact de BG .

2.3. L'homomorphisme de Baum-Connes

Soit Y un espace topologique localement compact muni d'une action propre de G ; on suppose de plus que Y/G est compact.

Si l'action de G est de plus libre, l'équivalence de Morita entre $C_0(Y) \rtimes G$ et $C(Y/G)$ est donnée par un $C_0(Y) \rtimes G$ -module hilbertien canonique E , tel que $\mathcal{K}(E)$ soit isomorphe à $C(Y/G)$ et qui est plein, c'est-à-dire que le sous-espace vectoriel fermé de $C_0(Y) \rtimes G$ engendré par les produits scalaires $\langle x, y \rangle$, $x, y \in E$ est $C_0(Y) \rtimes G$.

Si l'action est propre sans être libre, on fait exactement la même construction. La seule chose qui change est que le $C_0(Y) \rtimes G$ -module hilbertien canonique E n'est

⁽⁴⁾ A une action continue à gauche de G sur un espace localement compact Y , correspond une action de G sur la C^* -algèbre $C_0(Y)$ par la formule $(g.f)(y) = f(g^{-1}y)$, pour $g \in G$, $f \in C_0(Y)$, $y \in Y$.

plus plein. Cependant, $\mathcal{K}(E)$ est encore égal à $C(Y/G)$; en particulier, comme $\mathcal{K}(E)$ est unifié, $(E, \pi, 0)$ définit un élément Λ_Y de $KK(\mathbf{C}, C_0(Y) \rtimes G)$.

Une façon équivalente plus élémentaire de définir Λ_Y est la suivante : comme l'action de G dans Y est propre à quotient compact, il existe une fonction continue, positive, à support compact $c : Y \rightarrow \mathbf{R}$ telle que, pour tout $y \in Y$, $\int_G c(gy) dg = 1$. On note $p \in C_c(Y \times G) \subset C_c(G; C_0(Y)) \subset C_0(Y) \rtimes G$ la fonction $(y, g) \mapsto \Delta(g)^{-1/2} c(y)^{1/2} c(g^{-1}y)^{1/2}$. C'est un projecteur dans $C_0(Y) \rtimes G$ (i.e. $p = p^2 = p^*$) ; il définit donc un élément de $K_0(C_0(Y) \rtimes G) = KK(\mathbf{C}, C_0(Y) \rtimes G)$.

La classe de p ne dépend pas de la fonction c . En fait, le module $p(C_0(Y) \rtimes G)$ est isomorphe à E , ce qui fait le lien avec l'autre définition.

Soit maintenant $x \in K_{0, \text{top}}(G)$; il est représenté par une partie $Y \subset \underline{E}G$ fermée G -invariante telle que Y/G soit compact et un $y \in KK_G(C_0(Y), \mathbf{C})$. L'élément $\mu_G(x)$ est le produit de Kasparov de $\Lambda_Y \in KK(\mathbf{C}, C_0(Y) \rtimes G)$ et de $j_G(y) \in KK(C_0(Y) \rtimes G, C^*(G))$.

On vérifie que cette construction passe à la limite inductive et qu'on obtient ainsi un homomorphisme $\mu_G : K_{0, \text{top}}(G) \rightarrow K_0(C^*(G))$. L'homomorphisme de Baum-Connes, ou homomorphisme d'assemblage en K -théorie est l'homomorphisme $\mu_{G,r} : K_{0, \text{top}}(G) \rightarrow K_0(C_r^*(G))$ obtenu par composition de μ_G avec l'homomorphisme de K -théorie associé à l'homomorphisme de C^* -algèbres $\lambda : C^*(G) \rightarrow C_r^*(G)$.

CONJECTURE DE BAUM-CONNES. — *L'homomorphisme $\mu_{G,r}$ est un isomorphisme.*

2.4. Lien avec la conjecture de Novikov

Soit M une variété lisse compacte orientée. Notons Γ son groupe fondamental, \widetilde{M} son revêtement universel. L'opérateur de signature de M peut être considéré comme un opérateur Γ -invariant sur \widetilde{M} ; il définit ainsi un élément de K -homologie équivariante $KK_\Gamma(C_0(\widetilde{M}), \mathbf{C})$, donc de $K_{0, \text{top}}(\Gamma)$ (comme l'action de Γ sur \widetilde{M} est propre).

La conjecture de Novikov peut s'énoncer de la façon suivante : l'élément de $K_{0, \text{top}}(\Gamma)$ associé à l'opérateur de signature de M est rationnellement invariant par homotopie. Autrement dit, si $f : M \rightarrow N$ est une équivalence d'homotopie entre variétés lisses compactes orientées, la différence entre les éléments de $K_{0, \text{top}}(\Gamma)$ associés aux opérateurs de signature de M et de N est de torsion.

Miščenko a démontré que l'image par la flèche de Baum-Connes μ_Γ de cet élément est invariant par homotopie (cf. [39]). Donc l'injectivité rationnelle de l'homomorphisme de Baum-Connes implique la conjecture de Novikov.

Pour plus de détails sur ce sujet, on peut voir aussi les exposés [13], [45].

2.5. Généralisations et variantes

— Conjecture de Baum-Connes à coefficients.

Soient G un groupe localement compact et A une C^* -algèbre munie d'une action de G . On construit de même un groupe de K -théorie topologique à coefficients $K_{0,top}(G; A) = \varinjlim KK_G(C_0(Y), A)$ et des homomorphismes de Baum-Connes $\mu_G^A : K_{0,top}(G; A) \rightarrow \bar{K}_0(A \rtimes G)$ et $\mu_{G,r}^A : K_{0,top}(G; A) \rightarrow K_0(A \rtimes_r G)$. On conjecture alors que $\mu_{G,r}^A$ est un isomorphisme.

— Conjecture de Baum-Connes pour les groupoïdes.

Soit G un groupoïde localement compact. On peut encore (*cf.* [42]) lui associer une C^* -algèbre et une C^* -algèbre réduite (c'est de cette manière que Connes construit la C^* -algèbre d'un feuilletage - *cf.* [9]). A l'aide d'une généralisation aux groupoïdes de la K -théorie équivariante due à Le Gall (*cf.* [38]), on peut encore définir la K -théorie topologique et construire l'homomorphisme de Baum-Connes (*cf.* [4], [47]). On peut de plus définir un homomorphisme de Baum-Connes pour un groupoïde G à coefficients dans une G -algèbre.

— Conjecture de Baum-Connes à valeurs dans L^1 (due à J.-B. Bost).

En regardant d'un peu plus près la construction de l'homomorphisme de Baum-Connes, on se rend compte que l'on peut construire en fait un homomorphisme $\mu_{G,L^1} : K_{0,top}(G) \rightarrow K_0(L^1(G))$ et tel que μ_G est la composée de μ_{G,L^1} avec l'homomorphisme de K -théorie associé à l'homomorphisme $L^1(G) \rightarrow C^*(G)$. C'est en fait V. Lafforgue qui, grâce à sa KK -théorie banachique, a défini proprement l'homomorphisme μ_{G,L^1} .

2.6. Le point sur la conjecture

— La conjecture de Baum-Connes à coefficients dans n'importe quelle C^* -algèbre est connue pour tous les groupes (ou groupoïdes) localement compacts moyennables et, plus généralement, ceux opérant proprement par isométries affines sur un espace hilbertien ([19] - *cf.* [48] pour le cas des groupoïdes - voir à ce sujet l'exposé de P. Julg [24]).

Par ailleurs, Oyono ([40]) et Chabert ([8]) ont démontré des résultats de stabilité de la conjecture à coefficients par certaines opérations sur les groupes.

— L'injectivité de l'homomorphisme de Baum-Connes (à coefficients) est de plus connue

a) pour un groupe opérant proprement par isométries sur une variété riemannienne complète à courbure sectionnelle négative ou nulle ([29]), et en particulier pour un sous-groupe fermé d'un groupe de Lie réel ;

b) pour un groupe opérant proprement par isométries sur un immeuble affine et en particulier pour un sous-groupe fermé d'un groupe de Lie p -adique ([30]) ;

c) pour un groupe opérant proprement par isométries sur un espace métrique discret ayant un « bon » comportement à l'infini (faiblement géodésique, à géométrie bornée et « bolique»- [31]) ;

d) pour un groupe admettant une action moyennable sur un espace compact ([18]).

— Lafforgue démontre la conjecture de Baum-Connes :

e) Pour tout groupe de Lie semi-simple réel ou p -adique. Ce résultat était connu dans le cas des groupes de Lie réels linéaires connexes réductifs (*cf.* [50]) et les groupes GL_n p -adiques (*cf.* [6]) ;

f) pour les sous-groupes discrets cocompacts des groupes de Lie de rang réel 1 ou de $SL_3(K)$ - avec $K = \mathbf{R}, \mathbf{C}, \mathbf{Q}_p$. Dans ce cas on peut aussi admettre des coefficients dans une algèbre $C(X)$, où X est une variété riemannienne compacte munie d'une action isométrique de G . Dans le cas de rang 1, on n'a pas besoin de supposer que l'action de G dans X est isométrique.

C'est le résultat f) le plus spectaculaire de son travail, car il franchit la barrière de la propriété T pour un groupe discret, pour lequel il n'y a donc aucune description possible de $C_r^*(G)$.

2.7. L'élément γ de Kasparov

Dans tous les cas cités ci-dessus où l'on connaît l'injectivité de l'homomorphisme μ_G , on a en fait construit un homomorphisme inverse. Dans les cas a), b) et c) ci-dessus on sait décrire l'image de μ_G : on construit (*cf.* [29] dans le cas a), [26] dans le cas b), [31] dans le cas c)) un élément $\gamma \in KK_G(\mathbf{C}, \mathbf{C})$, qui est idempotent pour le produit de Kasparov et satisfait :

L'image de $\mu_{G,r}$ coïncide avec celle de l'endomorphisme idempotent de $K_0(C_r^*(G))$ associé à $j_G^r(\gamma) \in KK(C_r^*(G), C_r^*(G))$.

En d'autres termes, la conjecture de Baum-Connes se ramène à la formulation suivante : l'endomorphisme de $K_0(C_r^*(G))$ associé à $j_G^r(\gamma)$ est l'identité.

2.8. L'obstruction de la propriété T de Kazhdan

Rappelons (*cf.* [12]) qu'un groupe localement compact G possède la propriété T si la représentation triviale est isolée dans son dual. Du point de vue des C^* -algèbres, cela signifie qu'il existe un projecteur $p \in C^*(G)$ tel que, pour toute représentation π de $C^*(G)$ dans un espace hilbertien, $\pi(p)$ soit le projecteur orthogonal sur les vecteurs invariants par G .

Soit G un groupe localement compact non compact possédant la propriété T . Notons $\lambda : C^*(G) \rightarrow C_r^*(G)$ l'homomorphisme canonique. Comme la représentation régulière de G ne possède pas de vecteurs G -invariants (car G n'est pas compact), on a $\lambda(p) = 0$. Or la classe de p est un élément non trivial de la K -théorie de $C^*(G)$, puisque son image par la représentation triviale $\varepsilon : C^*(G) \rightarrow \mathbf{C}$ n'est pas nulle dans $K_0(\mathbf{C}) = \mathbf{Z}$. Il s'ensuit que λ n'induit pas un isomorphisme en K -théorie, donc que μ_G et $\mu_{G,r}$ ne sont pas tous deux bijectifs. Une démonstration éventuelle de la conjecture de Baum-Connes doit donc nécessairement distinguer entre $C^*(G)$ et $C_r^*(G)$.

Supposons de plus que G est un groupe comme dans a), b) ou c) ci-dessus. En particulier, il possède un élément γ . Par ce qui précède, l'endomorphisme de $K_0(C^*(G))$ associé à $j_G(\gamma)$ s'annule sur la classe de p ; il en résulte que $\gamma \neq 1$.

Par ailleurs, l'élément γ est représenté par un couple (H, F) où la représentation de G dans l'espace hilbertien H est un sous-multiple de la régulière; comme la représentation triviale est isolée, il est de toute façon impossible d'espérer une homotopie entre γ et 1.

3. KK -THÉORIE BANACHIQUE

On a vu qu'en présence de la propriété T , il n'est pas possible de faire une homotopie de γ à 1 dans $KK_G(\mathbf{C}, \mathbf{C})$.

Cependant, la représentation triviale est isolée parmi les représentations *unitaires* de G , c'est-à-dire des représentations dans des espaces hilbertiens; elle ne l'est pas dans les représentations banachiques. Par exemple la famille des représentations naturelles de G dans $L^p(G)$ contient faiblement, quand p tend vers $+\infty$, la représentation triviale.

Il est donc possible d'espérer construire une homotopie entre γ et 1 si on s'autorise des représentations banachiques.

Lafforgue est ainsi amené à construire une KK -théorie, où les espaces hilbertiens sont remplacés par des espaces de Banach - et donc les C^* -algèbres par des algèbres de Banach. C'est la KK -théorie banachique.

3.1. B -paires

Contrairement aux espaces hilbertiens, les espaces de Banach ne sont pas auto-duaux. Pour cette raison, Lafforgue remplace les modules hilbertiens par des paires de modules banachiques en dualité.

Soit B une algèbre de Banach. Une B -paire est un couple de modules banachiques en dualité. Autrement dit, une B -paire $(E^<, E^>)$ est la donnée d'un B -module banachique à gauche (*resp.* à droite) $E^<$ (*resp.* $E^>$) et d'une application $\langle \cdot, \cdot \rangle : E^< \times E^> \rightarrow B$ satisfaisant : $\forall x \in E^>, \xi \in E^<$, l'application $\eta \mapsto \langle \eta, x \rangle$ (*resp.* $y \mapsto \langle \xi, y \rangle$) est \mathbf{C} -linéaire, B -linéaire à gauche (*resp.* à droite) et $\|\langle \xi, x \rangle\| \leq \|\xi\| \|x\|$.

Soient $E = (E^<, E^>)$ et $F = (F^<, F^>)$ des B -paires. Un *morphisme de B -paires* de E dans F est un couple $f = (f^<, f^>)$ où $f^< : F^< \rightarrow E^<$ (*resp.* $f^> : E^> \rightarrow F^>$) est une application \mathbf{C} -linéaire, B -linéaire à gauche (*resp.* à droite), continue et satisfait : $\forall x \in E^>, \eta \in F^<$, on a $\langle \eta, f^>(x) \rangle = \langle f^<(\eta), x \rangle$. On note $\mathcal{L}(E, F)$ l'espace de Banach des morphismes de B -paires de E dans F (muni de la norme $(f^<, f^>) \mapsto \sup(\|f^<\|, \|f^>\|)$).

Si E, F, G sont des B -paires, $f \in \mathcal{L}(E, F)$ et $g \in \mathcal{L}(F, G)$ on définit $gf \in \mathcal{L}(E, F)$ en posant $(gf)^> = g^> \circ f^>$ et $(gf)^< = f^< \circ g^<$. Munie de ces opérations $\mathcal{L}(E) = \mathcal{L}(E, E)$ est une algèbre de Banach.

Soient $y \in F^>$ et $\xi \in E^<$. On note $\theta_{y, \xi} \in \mathcal{L}(E, F)$ (ou $|y\rangle\langle\xi|$) le morphisme donné par le couple d'applications $F^< \ni \eta \mapsto \langle \eta, y \rangle \xi \in E^<$ et $E^> \ni x \mapsto y \langle \xi, x \rangle \in F^>$. On

note $\mathcal{K}(E, F)$ le sous-espace vectoriel fermé de $\mathcal{L}(E, F)$ engendré par les morphismes $\theta_{y, \xi}$.

3.2. Définition de la KK -théorie banachique

Soient A et B des algèbres de Banach. Un cycle pour $KK^{\text{ban}}(A, B)$ est encore un triplet (E, π, f) où E est une B -paire $\mathbf{Z}/2\mathbf{Z}$ -graduée, $\pi : A \rightarrow \mathcal{L}(E)^{(0)}$ est un homomorphisme et $f \in \mathcal{L}(E)^{(1)}$ est tel que, pour tout $a \in A$, les morphismes $f\pi(a) - \pi(a)f$ et $(f^2 - 1)\pi(a)$ sont dans $\mathcal{K}(E)$.

La somme des cycles et l'homotopie sont définies comme dans le cas des modules hilbertiens. Il est facile de voir que l'ensemble des classes d'homotopie de cycles (muni de l'addition des cycles) est un groupe abélien.

Le groupe $KK^{\text{ban}}(A, B)$ est l'ensemble des classes d'homotopie de cycles.

La functorialité en A ne pose pas de problème. Soit $\varphi : B_1 \rightarrow B_2$ un homomorphisme d'algèbres de Banach. Notons encore $\tilde{\varphi} : \tilde{B}_1 \rightarrow \tilde{B}_2$ son extension unitale. Soit E_1 une B_1 -paire. Considérons-la comme une \tilde{B}_1 paire. On définit une B_2 -paire $E_2 = \varphi_*(E_1)$ qui est le quotient du produit tensoriel $E_1 \otimes_{\tilde{B}_1} \tilde{B}_2$ par le sous-espace fermé engendré par $xb_1 \otimes b_2 - x \otimes \varphi(b_1)b_2$ avec $x \in E_1$, $b_1 \in \tilde{B}_1$ et $b_2 \in \tilde{B}_2$. On construit de plus un homomorphisme $\varphi_* : f \mapsto (f \otimes 1)$ de $\mathcal{L}(E_1)$ dans $\mathcal{L}(E_2)$ qui envoie $\mathcal{K}(E_1)$ dans $\mathcal{K}(E_2)$. On en déduit un homomorphisme $\varphi_* : KK^{\text{ban}}(A, B_1) \rightarrow KK^{\text{ban}}(A, B_2)$.

3.3. Le cas équivariant

On suppose à présent qu'un groupe localement compact G opère sur les algèbres A et B . On définit le groupe $KK_G^{\text{ban}}(A, B)$ exactement de la même façon que le groupe de Kasparov correspondant.

Si A et B sont des C^* -algèbres, on a évidemment un homomorphisme naturel $KK_G(A, B) \rightarrow KK_G^{\text{ban}}(A, B)$.

Dans le paragraphe suivant, on esquissera le principal résultat technique de Lafforgue montrant (presque) que l'image de γ est égale à celle de 1 dans $KK_G^{\text{ban}}(\mathbf{C}, \mathbf{C})$. Pour en déduire la conjecture de Baum-Connes, il reste à construire les équivalents en KK -théorie banachique

- du morphisme $j_G : KK_G(\mathbf{C}, \mathbf{C}) \rightarrow KK(C_r^*(G), C_r^*(G))$
- de l'action de la KK -théorie sur la K -théorie, autrement dit un homomorphisme $KK^{\text{ban}}(A, B) \rightarrow \text{Hom}(K_0(A), K_0(B))$.

3.4. Action sur la K -théorie

Rappelons que si A est une algèbre de Banach unifère, le groupe $K_0(A)$ est le groupe de Grothendieck du monoïde des classes d'isomorphisme de A -modules projectifs de type fini. Un tel module est isomorphe à un pA^n où $p \in M_n(A)$ est un idempotent.

Remarquons que $KK^{\text{ban}}(M_n(A), B)$ est isomorphe à $KK^{\text{ban}}(A, B)$. Par ailleurs, tout idempotent $p \in A$, détermine un homomorphisme $i_p : \lambda \mapsto \lambda p$ de $\mathbf{C} \rightarrow A$. De ces faits, on déduit une application bilinéaire $\varphi : (p, x) \mapsto (i_p)^*(x)$ de $K_0(A) \times KK^{\text{ban}}(A, B)$ dans $KK^{\text{ban}}(\mathbf{C}, B)$.

Cet homomorphisme subsiste si A n'est plus supposée unifière.

Lafforgue démontre de plus que, comme dans le cas de la théorie de Kasparov, l'application naturelle $(x \mapsto \varphi(x, 1_A))$ de $K_0(A)$ dans $KK^{\text{ban}}(\mathbf{C}, A)$ est un isomorphisme.

On en déduit donc que la KK -théorie banachique opère encore sur la K -théorie.

3.5. Complétions «inconditionnelles» et le morphisme j_G

La construction de l'homomorphisme $j_G^r : KK_G(A, B) \rightarrow KK(A \rtimes_r G, B \rtimes_r G)$ n'est malheureusement pas possible telle quelle : pour cela, il faudrait un analogue du produit croisé pour les algèbres de Banach, qui pour une C^* -algèbre donnerait le produit croisé réduit.

Cependant, on peut construire dans le cadre des algèbres de Banach le produit croisé $L^1(G; A)$. On en déduit un homomorphisme $j_G^{L^1} : KK_G^{\text{ban}}(A, B) \rightarrow KK^{\text{ban}}(L^1(G; A), L^1(G; B))$.

A l'aide de cet homomorphisme $j_G^{L^1}$, l'égalité $\gamma = 1$ dans $KK_G^{\text{ban}}(\mathbf{C}, \mathbf{C})$ implique la conjecture de Baum-Connes dans $L^1(G)$. Pour conclure, il faudrait savoir que l'inclusion $L^1(G) \rightarrow C_r^*(G)$ induit un isomorphisme en K -théorie (la surjectivité suffit). Cependant, on n'a pas encore d'outil permettant de démontrer cet isomorphisme (cf. la discussion dans [7] où cet isomorphisme est établi dans quelques cas).

Ici intervient une très jolie et simple idée de Lafforgue : celle de «complétion inconditionnelle» : on remarque que pour certaines normes d'algèbre sur $C_c(G)$, en général plus petites que la norme L^1 , on peut quand même former le produit croisé banachique :

DÉFINITION 1. — Une norme N d'algèbre sur $C_c(G)$ sera dite une norme inconditionnelle si, pour tout $f, g \in C_c(G)$, on a $|f(x)| \leq |g(x)|, \forall x \in G \Rightarrow N(f) \leq N(g)$.

On peut construire des produits croisés relatifs à toute norme inconditionnelle. Soient en effet G un groupe localement compact, N une norme inconditionnelle sur $C_c(G)$ et A une algèbre de Banach munie d'une action isométrique de G . Munissons $C_c(G; A)$ du produit de convolution tordu par l'action de G dans A (voir 1.3).

Notons $p_N : C_c(G; A) \rightarrow \mathbf{R}_+$ l'application qui à f associe $N(\varphi)$ où φ est l'application $x \mapsto \|f(x)\|$ de $G \rightarrow \mathbf{R}$. L'application p_N est une norme d'algèbre sur $C_c(G; A)$.

Notons $A \rtimes_N G$ le complété de $C_c(G; A)$ pour la norme p_N . Si E est un espace de Banach muni d'une action isométrique de G , on peut de même construire l'espace de Banach $E \rtimes_N G$. Si $(E^<, E^>)$ est une B -paire, alors $(E^< \rtimes_N G, E^> \rtimes_N G)$ est une $B \rtimes_N G$ -paire.

On obtient un homomorphisme $j_G^N : KK_G^{\text{ban}}(A, B) \rightarrow KK^{\text{ban}}(A \rtimes_N G, B \rtimes_N G)$.

Soit G un groupe dans les cas a), b) ou c) de 2.6. Pour montrer la conjecture de Baum-Connes pour G , il suffit donc de

1. Démontrer que $\gamma = 1$ dans $KK_G^{\text{ban}}(\mathbf{C}, \mathbf{C})$.
2. Construire une complétion inconditionnelle de $C_c(G)$ qui a la même K -théorie que $C_r^*(G)$.

3.6. Représentations non isométriques de G

En fait, l'homotopie entre γ et 1 n'aura pas lieu dans le groupe $KK_G^{\text{ban}}(\mathbf{C}, \mathbf{C})$, mais dans une généralisation.

Soit ℓ une fonction longueur sur G , c'est-à-dire une application continue $G \rightarrow \mathbf{R}_+$ telle que $\ell(1) = 0$ et $\ell(xy) \leq \ell(x) + \ell(y)$ pour tout $x, y \in G$ (la plupart des fonctions longueur rencontrées, vérifient de plus $\ell(x^{-1}) = \ell(x)$).

On définit un groupe $KK_{G,\ell}^{\text{ban}}(A, B)$ ⁽⁵⁾ exactement comme $KK_G^{\text{ban}}(A, B)$ sauf qu'ici on suppose que l'action de G dans les B -paires n'est plus isométrique mais elle est *contrôlée par ℓ* c'est-à-dire qu'elle est continue et que la norme de l'action de $x \in G$ est $\leq \exp(\ell(x))$.

Soient ℓ une fonction longueur sur G , N une norme inconditionnelle sur $C_c(G)$ et A une G -algèbre de Banach. Notons N' la norme inconditionnelle $f \mapsto N(e^\ell f)$. Il est facile de construire une application $j_G^{N,\ell} : KK_{G,\ell}^{\text{ban}}(A, B) \rightarrow KK^{\text{ban}}(A \rtimes_{N'} G, B \rtimes_N G)$.

De plus soit ℓ une fonction longueur positive. Pour $s \in \mathbf{R}_+^*$, notons N_s la norme inconditionnelle $f \mapsto N(e^{s\ell} f)$. Pour toute G -algèbre de Banach A , la sous-algèbre $\bigcup_{s \in \mathbf{R}_+^*} A \rtimes_{N_s} G$ de $A \rtimes_N G$ a la même K -théorie que $A \rtimes_N G$.

Pour démontrer que γ agit comme l'identité dans $K_0(A \rtimes_N G)$, il suffit de démontrer l'égalité $\gamma = 1$ dans chacun des groupes $KK_{G,s\ell}^{\text{ban}}(\mathbf{C}, \mathbf{C})$.

4. HOMOTOPIE DE γ À 1

Il y a en fait ici deux cas :

- le cas «géométrique» (groupe opérant proprement et isométriquement sur une variété riemannienne complète de courbure sectionnelle négative ou nulle) ; ce cas contient les groupes de Lie semi-simples réels et leurs sous-groupes fermés.
- le cas «combinatoire» (groupe opérant proprement et isométriquement sur un espace métrique «fortement bolique») ; ce cas contient les groupes de Lie réductifs p -adiques et leurs sous-groupes fermés ainsi que les sous groupes discrets cocompacts des groupes de Lie semi-simples réels (et leurs sous-groupes).

⁽⁵⁾ Lafforgue note ce groupe $KK_{G,e^\ell}^{\text{ban}}(A, B)$

Nous nous contenterons ici de donner la construction dans un cas combinatoire : uniquement dans le cas des immeubles de type \widetilde{A}_2 .

4.1. Le cas des immeubles de type \widetilde{A}_2

Soient X un immeuble de type \widetilde{A}_2 (e.g. l'immeuble associé à $SL_3(\mathbf{Q}_p)$) et G un groupe localement compact opérant par isométries sur X . Dans ce cas l'élément γ a été construit par P. Julg et A. Valette ([26]).

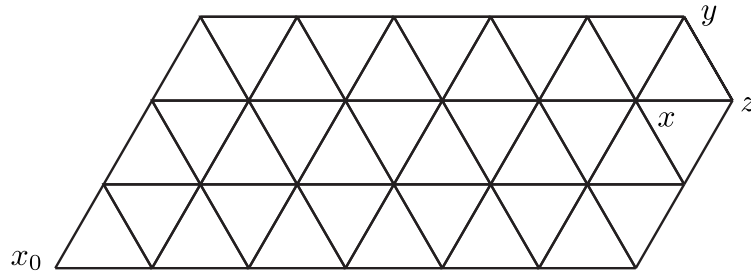
4.1.a) L'élément γ de Julg-Valette

On note $X^{(i)}$ ($i = 0, 1, 2$) l'ensemble des faces de dimension i de X . Notons $(e_x)_{x \in X^{(0)}}$ la base hilbertienne canonique de $\ell^2(X^{(0)})$. L'élément γ est représenté par un couple (H, F) :

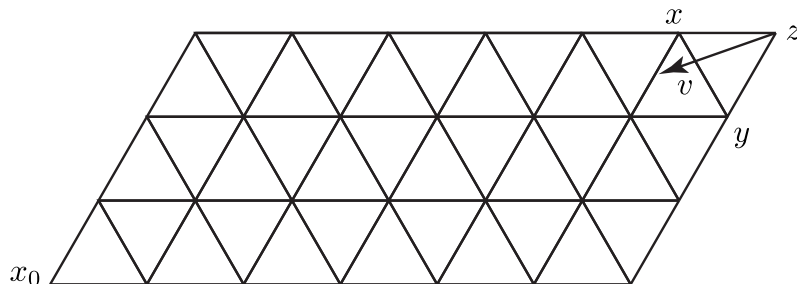
– L'espace hilbertien $H = \bigoplus_{i=0}^2 H_i$ où $H_0 = \ell^2(X^{(0)})$ et H_1 (resp. H_2) est le sous-espace de la deuxième (resp. troisième) puissance extérieure de $\ell^2(X^{(0)})$ engendré par les $e_x \wedge e_y$ (resp. $e_x \wedge e_y \wedge e_z$) où $\{x, y\} \in X^{(1)}$ (resp. $\{x, y, z\} \in X^{(2)}$). Il est gradué sur $\mathbf{Z}/2\mathbf{Z}$ par $H^{(0)} = H_0 \oplus H_2$ et $H^{(1)} = H_1$.

– L'opérateur F s'écrit $F_1 + F_1^*$, où F_1 est défini de la manière suivante : on choisit une origine $x_0 \in X^{(0)}$. Soit $\sigma = \{x, y, z\}$ une chambre et A un appartement contenant x_0 et A . On écrit x_0 dans l'espace affine A à l'aide des coordonnées barycentriques $x_0 = \lambda x + \mu y + \nu z$, ($\lambda, \mu, \nu \in \mathbf{R}$ de somme 1). Quitte à permuter x, y, z , on peut supposer que $\lambda \geq \mu \geq \nu$.

- Si $\lambda > 0 > \mu \geq \nu$, on prendra $F_1(e_y \wedge e_z) = e_x \wedge e_y \wedge e_z$.



- Supposons $\mu \geq 0 \geq \nu$; soit (λ', μ') dans l'espace euclidien \mathbf{R}^2 , positivement proportionnel à (λ, μ) et de norme 1. On pose $v = \lambda' e_x + \mu' e_y$ $F_1(e_z) = v \wedge e_z$, $F_1(e_y \wedge e_z) = v \wedge (e_y \wedge e_z) = \lambda'(e_x \wedge e_y \wedge e_z)$ et $F_1(e_x \wedge e_z) = \lambda' e_x + \mu' e_y \wedge (e_x \wedge e_z) = -\mu'(e_x \wedge e_y \wedge e_z)$.



On vérifie sans peine que F_1 est ainsi bien défini, qu'on a $F = F^*$, que $1 - F^2$ est le projecteur orthogonal sur e_{x_0} . L'action de G ne fait que changer l'origine x_0 . Mais si x_0 et y_0 sont deux points de $X^{(0)}$, la différence entre les coordonnées barycentriques de x_0 et de y_0 sont bornées sur tout l'immeuble. On en déduit que $F_{x_0} - F_{y_0}$ est compact, donc que (H, F) est un élément de $KK_G(\mathbf{C}, \mathbf{C})$.

Il résulte de [30] que sa classe est l'élément γ au sens de 2.7.

4.1.b) *L'homotopie*

Dans la construction qui suit, nous ne considérons pas des \mathbf{C} -paires mais des espaces de Banach. Cependant, à un espace de Banach E correspond une \mathbf{C} -paire (E^*, E) .

Notons $\varphi : X \rightarrow \mathbf{R}_+$ la fonction qui à une face f associe la distance à x_0 de son point le plus éloigné et $\ell : G \rightarrow \mathbf{R}_+$ la fonction $g \mapsto \varphi(g(x_0))$. C'est une fonction longueur puisque $\ell(gh) = d(gh(x_0), x_0) \leq d(gh(x_0), g(x_0)) + d(g(x_0), x_0)$; or, comme l'action de G est isométrique, $d(gh(x_0), g(x_0)) = d(h(x_0), x_0)$.

Le principal résultat technique de Lafforgue se lit dans notre cas :

THÉORÈME 1. — *Pour tout $s > 0$, les classes de γ et de 1 coïncident dans $KK_{G, s\ell}^{\text{ban}}(\mathbf{C}, \mathbf{C})$.*

Pour $p \in [1, +\infty]$ notons E_p l'espace de Banach, gradué par 0, 1, 2 dont la composante de degré i est $\ell^p(X^{(i)})$ (vu comme partie de la puissance extérieure $i^{\text{ème}}$ de $\ell^p(X^{(0)})$).

Considérons l'opérateur $\partial : E_p \rightarrow E_p$ donné par $\partial(e_x) = 0$, $\partial(e_x \wedge e_y) = e_y - e_x$ et $\partial(e_x \wedge e_y \wedge e_z) = (e_y \wedge e_z) - (e_x \wedge e_z) + (e_x \wedge e_y)$ (pour tout $(x, y, z) \in X^{(2)}$).

Soit enfin $t \in \mathbf{R}_+$; notons A_t l'opérateur non borné de multiplication par $e^{t\varphi}$ et posons $\partial_t = A_t \circ \partial \circ A_t^{-1}$. C'est encore un opérateur borné. De plus, pour tout $g \in G$, $\partial_t - g.\partial_t$ est compact dans tout E_p . En effet, $g.\partial_t$ est construit comme ∂_t , mais avec l'origine x_0 remplacée par gx_0 . Mais dans une chambre située loin de x_0 , les fonctions distance à x_0 et à gx_0 diffèrent d'une constante additive.

Le résultat crucial est le suivant :

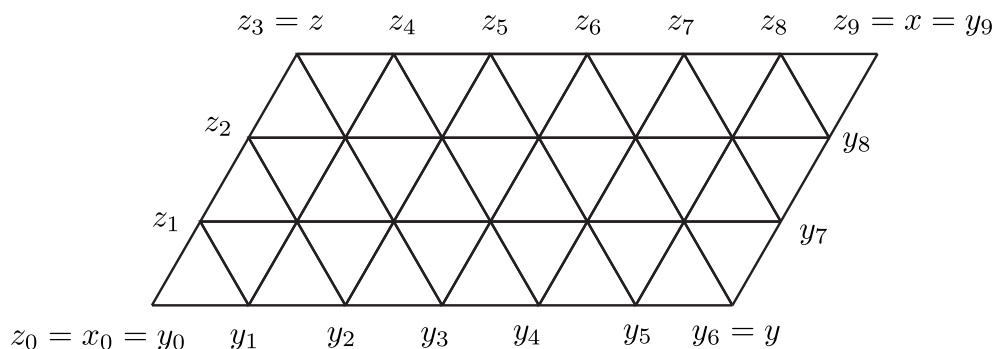
PROPOSITION 1. — *Soit $s > 0$. Il existe $T_s : E_1 \rightarrow E_1$ tel que $\text{id}_{E_1} - (T_s \partial_s + \partial_s T_s)$ soit un idempotent de rang 1 et tel que, pour tout $g \in G$, $T_s - g.T_s$ soit compact. De plus, il existe $t > 0$ tel que $T_t = A_{t-s} \circ T_s \circ A_{t-s}^{-1}$ soit continu de E_p dans E_p pour tout $p \in [1, 2]$, et vérifie :*

- a) $\text{id}_{E_p} - (T_t \partial_t + \partial_t T_t)$ est un idempotent de rang 1 ;
- b) pour tout $g \in G$, $T_t - g.T_t$ est compact dans tout E_p .

L'opérateur T_s est défini (par exemple) de la manière suivante :

Soit $x \in X^{(0)}$. Les points x et x_0 déterminent un parallélogramme x_0, y, x, z dans l'immeuble X . Notons j et $n - j$ les distances respectives de x à y et z . Notons

$z_0 = x_0, z_1, \dots, z_j = z, \dots, z_n = x$ et $y_0 = x_0, y_1, \dots, y_{n-j} = y, \dots, y_n = x$ les points sur les deux trajets de x_0 à x par z et y respectivement.



On pose

$$T_0(e_x) = \left(1 - \frac{j}{n}\right) \sum_{k=1}^n e_{z_{k-1}} \wedge e_{z_k} + \frac{j}{n} \sum_{k=1}^n e_{y_{k-1}} \wedge e_{y_k}$$

Il est clair que $\partial \circ T_0(e_x) = e_x - e_0$.

Pour définir $T_0(e_x \wedge e_y)$ on utilise le lemme suivant :

LEMME 1. — $e_x \wedge e_y - T_0 \partial(e_x \wedge e_y)$ est dans l'image de ∂ .

Il suffit pour démontrer ce lemme de se restreindre à un parallélogramme contenant $\{x_0, x, y\}$ et de remarquer que la restriction de ∂ à ce parallélogramme est exacte en dimension 1 et 2.

On définit $T_0(e_x \wedge e_y)$ comme l'élément ξ tel que $\partial(\xi) = e_x \wedge e_y - T_0 \partial(e_x \wedge e_y)$ décrit par le lemme 1 (ou plutôt sa démonstration).

Pour achever la démonstration de la proposition 1, il suffit de poser $T_s = A_s \circ T_0 \circ A_s^{-1}$. On montre que :

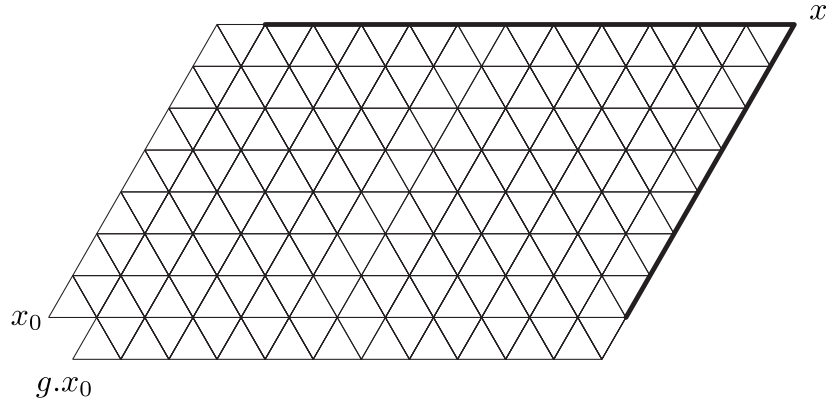
1. Pour tout $s > 0$, T_s est continu de E_1 dans E_1 .

Pour cela, il suffit de montrer que $\|T_s(e_x)\|$ et $\|T_s(e_x \wedge e_y)\|$ sont bornées indépendamment de $\{x, y\} \in X^{(1)}$: en effet, les faces que font intervenir $T_0(e_x)$ et $T_0(e_x \wedge e_y)$ sont situées dans le parallélogramme considéré ci-dessus ; leur nombre est polynomial en la distance de x à x_0 ; les coefficients apparaissant sont bornés. En conjuguant par A_s , on multiplie ces coefficients par un terme qui décroît exponentiellement, d'où le résultat.

2. Pour tout $s > 0$ et tout $g \in G$, $T_s - g.T_s \in \mathcal{K}(E_1)$.

Il suffit pour cela de vérifier que, quand $\{x, y\}$ tend vers l'infini, $\|(T_s - g.T_s)(e_x)\|$ et $\|(T_s - g.T_s)(e_x \wedge e_y)\|$ tendent vers 0. Pour cela, remarquons que $g.T_0$ est construit comme T_0 mais avec l'origine x_0 remplacée par gx_0 . Si x est loin de x_0 , les trajets de x à x_0 et de x à gx_0 utilisés dans la construction de T_0 coïncident près de x . Par

ailleurs, ce sont les points près de x qui « comptent » à cause de la conjugaison par A_s .



3. Il existe $t > 0$ tel que, pour tout $p \in [1, 2]$, T_t soit continu de E_p dans E_p et, pour tout $g \in G$, $T_t - g.T_t \in \mathcal{K}(E_p)$.

La difficulté ici vient de ce que $T_t^*(e_x \wedge e_y)$ fait intervenir tous les points z tels que x, y soit sur le trajet de z à x_0 . Le coefficient apparaissant décroît en $\exp(-t\varphi(z))$; le nombre de tels z croît exponentiellement. Cependant, en prenant t suffisamment grand, on peut contrôler la norme ℓ^1 de $T_t^*(e_x \wedge e_y)$ et celle de $T_t^*(e_x \wedge e_y \wedge e_z)$. Il en résulte que T_t^* est continu de E_1 dans E_1 , donc que T_t est continu de E_∞ dans E_∞ . Comme il est continu de E_1 dans E_1 , il est continu de E_p dans E_p pour tout p (par interpolation). Un raisonnement analogue montre que $T_t - g.T_t \in \mathcal{K}(E_p)$ pour tout $g \in G$.

En faisant varier $u \in [s, t]$ puis $p \in [1, 2]$ on dispose à présent d'une homotopie entre $(E_1, \partial_s + T_s)$ et $(E_2, \partial_t + T_t)$.

Maintenant, pour achever la démonstration du théorème, on va utiliser les deux faits suivants : pour s tendant vers 0, ∂_s devient de plus en plus G -invariant « donc » $(E_1, \partial_s + T_s)$ tend vers 1 ; pour t tendant vers $+\infty$, $(E_2, \partial_t + T_t)$ ressemble de plus en plus à γ .

Pour mettre en œuvre cette idée, Lafforgue procède de la manière suivante :

On peut construire une homotopie entre $(E_1, \partial_s + T_s)$ et $(E_1^s, \partial + T_0)$ où E_1^s est l'espace de Banach, gradué par $0, 1, 2$ dont la composante de degré i est l'espace ℓ^1 de $X^{(i)}$ à poids $e^{s\varphi}$. Notons que l'action de G dans E_1^s n'est plus isométrique ; c'est pour cela qu'on doit ici quitter $KK_G^{\text{ban}}(\mathbf{C}, \mathbf{C})$ et le remplacer par $KK_{G, s\ell}^{\text{ban}}(\mathbf{C}, \mathbf{C})$.

Pour montrer que la classe de $(E_1^s, \partial + T_0)$ est 1 et celle de $(E_2, \partial_t + T_t)$ est γ , on utilise quelques lemmes généraux de KK -théorie banachique.

Soient G un groupe localement compact muni d'une fonction longueur ℓ , A, B des algèbres de Banach, E une B paire $\mathbf{Z}/2\mathbf{Z}$ -graduée munie d'actions de A et de G « contrôlées par ℓ » (au sens de 3.6). Notons D l'ensemble des morphismes S de E tels que $[S, a]$ (pour tout $a \in A$) et $g.S - S$ (pour tout $g \in G$) soient dans $\mathcal{K}(E)$ et

$g \mapsto g.S$ soit normiquement continu. Remarquons que tout $F \in D$ de degré 1 tel que $\text{id}_E - F^2 \in \mathcal{K}(E)$ définit un élément de $KK_{G,\ell}^{\text{ban}}(A, B)$.

LEMME 2. — Soit $S \in D$ de degré 1 tel que $S^2 \in \mathcal{K}(E)$.

a) Supposons qu'il existe un morphisme $T \in D$ de degré 1 tel que $\text{id}_E - (TS + ST)$ soit compact. Alors, il existe un tel T avec en plus T^2 compact. Dans ce cas $(E, S+T)$ est un élément de $KK_{G,\ell}^{\text{ban}}(A, B)$.

b) Dans ce cas, la classe de $(E, S+T)$ dans $KK_{G,\ell}^{\text{ban}}(A, B)$ ne dépend pas de T .

En effet, dans a) il suffit de remplacer T par TST , en remarquant que, modulo $\mathcal{K}(E)$, $ST = (\text{id}_E - TS)$, donc TS et ST commutent, donc $STTS \in \mathcal{K}(E)$; on en déduit aussi que $ST - STST$ qui est égal, modulo $\mathcal{K}(E)$ à $STTS$ est dans $\mathcal{K}(E)$; de même $TS - TSTS \in \mathcal{K}(E)$.

Pour montrer b), il suffit de remarquer que l'ensemble des $T \in D$ tels que $ST + TS - \text{id}_E \in \mathcal{K}(E)$ est un sous-espace affine de D .

LEMME 3. — Soient $S, S' \in D$ de degré 1 tels que $S^2 \in \mathcal{K}(E)$ et $(S')^2 \in \mathcal{K}(E)$. Supposons que le spectre dans l'algèbre de Banach quotient $D/\mathcal{K}(E)$ de $SS' + S'S$ ne rencontre pas \mathbf{R}_- . Alors :

a) Il existe $T, T' \in D$ tels que $T^2, (T')^2, \text{id}_E - (ST + TS), \text{id}_E - (S'T' + T'S') \in \mathcal{K}(E)$.

b) Les classes de $(E, S+T)$ et $(E, S'+T')$ dans $KK_{G,\ell}^{\text{ban}}(A, B)$ coïncident.

Raisonnons dans l'algèbre de Banach quotient $D/\mathcal{K}(E)$. On a $(SS' + S'S)S = SS'S = S(SS' + S'S)$; de même, $S'(SS' + S'S) = (SS' + S'S)S'$. Donc $(SS' + S'S)^{-1}$ commute avec S et S' . On pose $T = S'(SS' + S'S)^{-1}$ et $T' = S(SS' + S'S)^{-1}$. De plus, comme le spectre de $SS' + S'S$ ne rencontre pas \mathbf{R}_- , on peut définir un logarithme de $SS' + S'S$. On obtient une homotopie $S(SS' + S'S)^{-t} + S'(SS' + S'S)^{t-1}$ entre $S+T$ et $S'+T'$.

LEMME 4. — Soient $S, T \in D$ de degré 1. On suppose $T^2 \in \mathcal{K}(E)$, que S commute exactement à A et à G , que $S^2 = 0$ et $ST + TS = \text{id}_E$. Alors la classe de $(E, S+T)$ dans $KK_{G,\ell}^{\text{ban}}(A, B)$ est nulle.

On peut remplacer T par TST . Dans ce cas on a $T^2 = 0$. Comme S commute avec les éléments de A et de G , le sous-espace $S(E)$ de E est invariant par ces éléments. Dans la décomposition $E = S(E) \oplus T(E)$, la matrice de ces éléments est de la forme $\begin{pmatrix} c_{1,1} & c_{1,2} \\ 0 & c_{2,2} \end{pmatrix}$. Notons que $c_{1,2}$ est dans $\mathcal{K}(E)$ car $T \in D$. Il suffit de changer ces actions en $\begin{pmatrix} c_{1,1} & 0 \\ 0 & c_{2,2} \end{pmatrix}$ à travers une homotopie $\begin{pmatrix} c_{1,1} & tc_{1,2} \\ 0 & c_{2,2} \end{pmatrix}$ ($t \in [0, 1]$). Au bout de cette homotopie, $S+T$ commute exactement avec A et G et $(S+T)^2 = \text{id}_E$. Le lemme en résulte facilement.

Pour finir la démonstration du théorème, notons u la classe dans $KK_{G,s\ell}^{\text{ban}}(\mathbf{C}, \mathbf{C})$ de $(E_1^s, \partial + T_0)$; posons $F = E_1^s \oplus \mathbf{C}$ où \mathbf{C} est muni de l'action triviale de G et est de degré 1 pour la graduation. Prolongeons ∂ de F dans F en posant $\partial(e_x) = 1 \in \mathbf{C}$ pour tout $x \in X^{(0)}$ et posons $T_0(1) = e_{x_0}$. Il est facile de voir que la classe de cet

élément est $u - 1$ (ou plutôt u - l'image de 1). Par le lemme 4 appliqué à $S = \partial$, on trouve donc que u et l'image de 1 dans $KK_{G,sl}^{\text{ban}}(\mathbf{C}, \mathbf{C})$ coïncident.

Pour montrer que la classe de $(E_2, \partial_t + T_t)$ est l'image de γ , on commence d'abord par modifier très légèrement F_1 en un opérateur F_2 : le vecteur $\lambda'e_y + \mu'e_z$ par lequel on prend le produit extérieur est normalisé en norme 1 au lieu de l'être en norme 2. Il suffit alors, par le lemme 3, de démontrer que le spectre de $\partial_t F_2 + F_2 \partial_t$ est disjoint de \mathbf{R}_- . Pour cela, on remarque que l'image par $(\partial_t F_2 + F_2 \partial_t) - \text{id}_{E_2}$ d'une face f est portée par des faces voisines, plus proches de x_0 que f . Donc, pour t assez grand, le rayon spectral de $(\partial_t F_2 + F_2 \partial_t) - \text{id}_{E_2}$ est < 1 .

4.2. Les autres cas

4.2.a) *Le cas « combinatoire »*

Dans ce cas, la construction est exactement la même. La seule chose qui devient réellement plus difficile ici est la démonstration du fait que le complexe ∂ est acyclique.

4.2.b) *Le cas « géométrique »*

Soient M une variété riemannienne complète, connexe, simplement connexe, de courbure sectionnelle négative ou nulle (bornée inférieurement) et G un groupe localement compact agissant sur M par isométries.

Soit x_0 un point de M et notons $\varphi : M \rightarrow \mathbf{R}$ l'application donnée par $\varphi(x) = \exp(\sqrt{d(x, x_0)^2 + 1})$.

Dans ce cas, on remplace le bord de la cohomologie simpliciale par la cohomologie de de Rham. On est amené à travailler avec des espaces $W(p, k, s - k, \varphi^t)$: espace de Sobolev L^p , des k formes $s - k$ fois différentiables à poids φ^t (s est, par exemple, la dimension de l'espace M).

Une difficulté qui apparaît ici est que le calcul pseudodifférentiel ne marche bien que pour $p \in]0, +\infty[$. Cependant, $t > 0$ étant choisi, le complexe de de Rham $(W(p, k, s - k, \varphi^t), d)$ est acyclique pour p assez proche de $+\infty$.

Le reste de la construction est essentiellement le même que dans le cas « combinatoire ».

Je dois cependant dire que, bien que je sois absolument convaincu que la construction marche aussi bien que dans le cas combinatoire, je n'ai pas encore vérifié tous les détails du cas géométrique.

5. FIN DE LA DÉMONSTRATION

Soit G un groupe localement compact opérant proprement et isométriquement sur une variété riemannienne complète de courbure sectionnelle négative ou nulle ou sur un immeuble affine. D'après ce qui précède, G possède un élément γ et $\gamma = 1$

dans la KK -théorie banachique. Donc, pour établir la conjecture de Baum-Connes, il suffit de construire une complétion inconditionnelle A de $C_c(G)$ telle que le morphisme $A \rightarrow C_r^*(G)$ induise une surjection au niveau des groupes K_0 .

5.1. Cas des groupes de Lie

Si G est un groupe de Lie semi-simple, réel ou p -adique, Lafforgue ([36]) montre que certaines variantes de l'espace de Schwarz (ou algèbre de Harish-Chandra) de G fournissent des complétions inconditionnelles de $C_c(G)$ qui permettent de terminer la démonstration. On obtient ainsi une preuve plus directe de résultats de Wassermann dans le cas réel ([50]) et une généralisation de Baum, Higson et Plymen dans le cas p -adique ([6]).

5.2. Propriété (RD) de Haagerup-Jolissaint

Soit G un groupe discret ℓ une fonction longueur sur G (i.e. une application $G \rightarrow \mathbf{R}_+$ telle que $\ell(e) = 0$, $\ell(xy) \leq \ell(x) + \ell(y)$ pour tout $x, y \in G$). Notons $H^\infty(G, \ell)$ l'espace vectoriel des fonctions $f : G \rightarrow \mathbf{C}$ telles que, pour tout $p \in \mathbf{R}_+$, $\sum_{x \in G} \ell(x)^p |f(x)|^2 < \infty$. Si G est finiment engendré, on note $H^\infty(G)$ l'espace $H^\infty(G, \ell)$ où ℓ est la fonction longueur associée à un système fini de générateurs (l'espace $H^\infty(G)$ n'en dépend pas).

Dans [16], Haagerup a démontré :

PROPOSITION 1. — *Si G est un groupe libre à un nombre fini de générateurs, $H^\infty(G) \subset C_r^*(G)$.*

En effet :

Pour $p \in \mathbf{N}$, notons χ_p la fonction caractéristique des mots de longueur p (pour un système libre de générateurs. Soient $f, g : G \rightarrow \mathbf{C}$ des fonctions. On utilise le lemme suivant :

LEMME 1. — *Pour tout $p, q, n \in \mathbf{N}$, on a $\|((f\chi_p) \star (g\chi_q))\chi_n\|_2 \leq \|f\chi_p\|_2 \|g\chi_q\|_2$. De plus si $((f\chi_p) \star (g\chi_q))\chi_n \neq 0$, alors n est compris entre $|p - q|$ et $p + q$ et a la même parité que $p + q$.*

On peut évidemment supposer que $f = f\chi_p$ et $g = g\chi_q$.

La contrainte sur n est claire. Ecrivons $p = p_1 + r$, $q = q_1 + r$ et $n = p_1 + q_1$. Notons C_k l'ensemble des mots de longueur k .

Soient $x, y, z \in G$ de longueur respective p, q, n tels que $xy = z$. Ecrivons $x = x_1x_2$ où $x_1 \in C_{p_1}$ et $x_2 \in C_r$ et $y = y_2y_1$ où $y_1 \in C_{q_1}$ et $y_2 \in C_r$ (ces éléments sont uniquement déterminés par x et y). Comme xy est de longueur n on voit que nécessairement $z = x_1y_1$ et $y_2 = x_2^{-1}$. Notons que x_1 et y_1 sont entièrement déterminés par z .

Il s'ensuit que

$$|f \star g(z)| = \left| \sum_{u \in C_r} f(x_1 u) g(u^{-1} y_1) \right| \leq \left(\sum_{u \in C_r} |f(x_1 u)|^2 \right)^{1/2} \left(\sum_{u \in C_r} |g(u y_1)|^2 \right)^{1/2}$$

par l'inégalité de Cauchy-Schwarz.

On a alors

$$\begin{aligned} \|(f \star g)\chi_n\|_2^2 &= \sum_{x_1 \in C_{p_1}, y_1 \in C_{q_1}, x_1 y_1 \in C_n} |f \star g(x_1 y_1)|^2 \\ &\leq \sum_{x_1 \in C_{p_1}, y_1 \in C_{q_1}, x_1 y_1 \in C_n} \left(\sum_{u \in C_r} |f(x_1 u)|^2 \right) \left(\sum_{u \in C_r} |g(u y_1)|^2 \right) \\ &\leq \sum_{x_1 \in C_{p_1}, y_1 \in C_{q_1}} \left(\sum_{u \in C_r} |f(x_1 u)|^2 \right) \left(\sum_{u \in C_r} |g(u y_1)|^2 \right) \\ &= \|f\|_2^2 \|g\|_2^2. \end{aligned}$$

La proposition 1 se déduit très facilement du lemme 1 :

On écrit $\ell^2(G)$ comme somme hilbertienne de $\ell^2(C_r)$.

Soient $f \in C_c(G)$ et $p \in \mathbf{N}$. L'opérateur de convolution à gauche par $(f\chi_p)$ peut être vu dans la décomposition ci-dessus comme une matrice $T_{r,s}$ où $T_{r,s} : \ell^2(C_s) \rightarrow \ell^2(C_r)$ est un opérateur de norme $\leq \|(f\chi_p)\|_2$, nul si $|r - s| \geq p$, ou si $r - s$ n'est pas de la même parité que p . On peut écrire T comme somme des $p + 1$ opérateurs S_k où S_k est l'opérateur de matrice $T_{r,s}$ avec $r - s = -p + 2k$ ($0 \leq k \leq p$). Il est clair que $\|\lambda(f\chi_p)\| \leq \sum_k \|S_k\| \leq (p + 1)\|(f\chi_p)\|_2$. Il vient $\|\lambda(f)\| \leq \sum (p + 1)\|f\chi_p\|_2 \leq (\sum (p + 1)^{-2})^{1/2} \sum_p ((p + 1)^4 \|f\chi_p\|_2^2)^{1/2}$ par l'inégalité de Cauchy-Schwartz, d'où la proposition.

On peut en fait, à partir du lemme 1, démontrer :

PROPOSITION 2. — *Si G est un groupe libre à un nombre fini de générateurs, $H^\infty(G)$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel holomorphe. En particulier, l'inclusion $H^\infty(G) \rightarrow C_r^*(G)$ induit un isomorphisme au niveau de la K -théorie.*

Cette idée est reprise par Jolissaint ([22]). Il donne la définition suivante :

DÉFINITION 2. — *On dit qu'un groupe finiment engendré G possède la propriété (RD) si $H^\infty(G) \subset C_r^*(G)$.*

Il démontre que si G a la propriété (RD), alors $H^\infty(G)$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel. En fait (cf. [23]), il existe $r \in \mathbf{N}$ tel que l'ensemble $A_r = \{f \in \mathbf{C}^G, \sum_{x \in G} \ell(x)^r |f(x)|^2 < \infty\}$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel holomorphe.

Remarquons que A_r est une algèbre de Banach et une complétion inconditionnelle.

5.3. Groupes possédant la propriété (RD)

- Jolissaint ([22]) étend le travail de Haagerup en démontrant que plusieurs groupes se comportant comme les groupes libres, en particulier les sous-groupes cocompacts de groupes de Lie réels simples de rang 1, ont la propriété (RD).
- Ce résultat est lui-même étendu par de la Harpe ([17]) aux groupes hyperboliques de Gromov.
- Très récemment, Ramagge, Robertson et Steger ([41]) ont établi la propriété (RD) pour les groupes discrets opérant proprement sur certains immeubles \tilde{A}_2 . Leur résultat contient en particulier les sous-groupes discrets de $SL_3(\mathbf{Q}_p)$.

Comme dans la démonstration de Haagerup, on doit analyser les couples x, y ayant un produit donné. On est amené à étudier les triangles dans l'immeuble. La clé de la démonstration est que tout triangle se contracte uniquement sur un triangle équilatéral dans un appartement.

- Lafforgue enfin, en remplaçant certaines égalités par des estimations, a adapté la démonstration de Ramagge, Robertson et Steger au cas des sous-groupes cocompacts de $SL_3(\mathbf{R})$ et $SL_3(\mathbf{C})$ (*cf.* [34]).

5.4. Limite de la méthode

D'après une conjecture de Valette ([49]), tout sous-groupe cocompact d'un groupe de Lie simple devrait posséder la propriété (RD).

Cependant, la méthode a ses limites : plusieurs groupes discrets de groupes de Lie (non cocompacts) ne possèdent pas la propriété (RD) du fait de l'existence d'éléments paraboliques. Par exemple, les seuls groupes moyennables possédant la propriété (RD) sont les groupes à croissance polynomiale (*cf.* [22]). De plus cette méthode marche très mal pour les produits croisés - et ne permet de démontrer la conjecture de Baum-Connes à coefficients que dans des cas très particuliers (*cf.* [37]).

6. QUELQUES AUTRES AVANCÉES RÉCENTES

En plus du résultat de Lafforgue que nous venons de discuter, il y a eu récemment quelques autres avancées significatives sur la conjecture de Baum-Connes et des sujets connexes.

6.1. Actions moyennables ([18])

N. Higson vient de faire une remarque très simple et astucieuse qui permet d'établir l'injectivité de l'homomorphisme de Baum-Connes (à coefficients) pour les groupes discrets possédant une action moyennable sur un espace compact.

La moyennabilité d'un groupe est une notion dégagée depuis plusieurs décennies. On dispose à présent de plusieurs propriétés équivalentes. De plus, cette notion a été

étendue aux actions de groupes et aux groupoïdes. Nous renvoyons à l'excellente monographie [2] pour une discussion très actuelle sur toutes les notions de moyennabilité.

Du point de vue des C^* -algèbres, un groupe localement compact G est moyennable si $C^*(G) = C_r^*(G)$. Une action de G sur un espace compact X est moyennable si $C(X) \rtimes G = C(X) \rtimes_r G$.

Rappelons que Higson et Kasparov ont démontré la conjecture de Baum-Connes pour tout groupe moyennable, et que Tu a étendu ce résultat au cas des groupoïdes moyennables.

Toute action d'un groupe moyennable est moyennable. Cependant, plusieurs groupes, bien que non moyennables, possèdent des actions moyennables sur des espaces compacts. Par exemple, $G = SL_n(\mathbf{R})$ n'est pas moyennable pour $n \geq 2$, mais son action dans l'espace compact des drapeaux G/P est moyennable où P désigne le sous-groupe de $SL_n(\mathbf{R})$ formé des matrices triangulaires supérieures (à coefficients diagonaux positifs).

La démonstration de Higson est (un peu schématiquement) la suivante :

Si l'action de G dans un espace compact X est moyennable, son action dans l'espace M des mesures de probabilité sur X , muni de la topologie de la convergence vague est moyennable.

Soit A une C^* -algèbre munie d'une action de G . Notons B la C^* -algèbre $C(M; A)$ des fonctions continues de M dans A . L'homomorphisme $A \rightarrow B$ qui à $a \in A$ associe la fonction constante égale à a étant équivariant, on obtient un homomorphisme $A \rtimes_r G \rightarrow B \rtimes_r G$, d'où un homomorphisme de groupes de K -théorie. De même, on a un homomorphisme $K_{0,top}(G; A) \rightarrow K_{0,top}(G; B)$. On obtient un diagramme commutatif

$$\begin{array}{ccc} K_{0,top}(G; A) & \xrightarrow{\mu_G^A} & A \rtimes_r G \\ \downarrow & & \downarrow \\ K_{0,top}(G; B) & \xrightarrow{\mu_G^B} & B \rtimes_r G \end{array}$$

Or, l'action de G dans M étant moyennable, par la généralisation par Tu ([48]) du théorème de Higson-Kasparov, l'homomorphisme de Baum-Connes μ_G^B est un isomorphisme. De plus, comme M est contractile, l'application $A \rightarrow B$ est une équivalence d'homotopie, d'où l'on déduit que l'homomorphisme $K_{0,top}(G; A) \rightarrow K_{0,top}(G; B)$ est aussi un isomorphisme. En fait pour avoir l'isomorphisme ici, Higson doit supposer que le groupe G est discret, il utilise un modèle simplicial pour $\underline{E}G$ et une suite exacte de Mayer-Vietoris. Il s'ensuit que l'homomorphisme μ_G^A est injectif (et scindé).

On peut facilement obtenir un résultat légèrement plus fort que celui de Higson en remplaçant la condition de moyennabilité de l'action de G par l'existence d'un plongement uniforme de G dans un espace hilbertien (au sens ci-dessous cf. [46]).

6.2. Conjecture de Baum-Connes et géométrie «à l'infini» ([43], [51])

Le résultat de [18] cité ci-dessus (ainsi que sa généralisation [46]) avait d'abord été démontré par Yu dans [51] sous l'hypothèse que G était sans torsion et BG un complexe cellulaire fini ⁽⁶⁾.

Le démonstration de Yu est tout à fait différente de celle de Higson : elle utilise le langage de la géométrie «grossière» (coarse), je dirais plutôt géométrie à l'infini.

Il s'agit là de tout un monde - et il n'est pas possible de citer ici toutes les contributions importantes dans le sujet. Nous renvoyons à [43] et [51] pour une discussion détaillée de toutes ces notions.

Il est apparu (cf. [14]) que la conjecture de Novikov pour un groupe discret G est en fait assez peu dépendante de la structure de groupe. La seule chose qui compte est la structure à l'infini de G .

Soit (X, d) un espace métrique. De même que pour une structure uniforme, une structure à l'infini de X est basée sur les ensembles de la forme $\Delta_r = \{(x, y) \in X \times X, d(x, y) < r\}$ (pour $r \in \mathbf{R}_+$). Cependant on s'intéresse à r grand.

On dit qu'une partie $U \subset X \times X$ est un *entourage* pour la structure à l'infini *s'il est contenu* dans un ensemble Δ_r .

Le rôle des applications uniformément continues est joué ici par les applications $f : X \rightarrow Y$ telles que l'image directe de tout entourage soit un entourage (autrement dit $\forall r \in \mathbf{R}_+, \exists R \in \mathbf{R}_+, d(x, x') \leq r \Rightarrow d(f(x), f(x')) \leq R$).

Une application $f : X \rightarrow Y$ est un *plongement uniforme* si, pour toute partie U de $X \times X$, U est un entourage si et seulement si $f(U)$ est un entourage.

A un espace métrique, on associe une C^* -algèbre construite de la manière suivante : on fixe un espace hilbertien H et une représentation $\pi : C_0(X) \rightarrow \mathcal{L}(H)$ telle que $\pi^{-1}(\mathcal{K}(H)) = \{0\}$ et $\pi(C_0(X))H = H$. Le support d'un élément $T \in \mathcal{L}(H)$ est le plus petit fermé F de $X \times X$ tel que si $f, g \in C_0(X)$ sont telles que $(x, y) \mapsto f(x)g(y)$ est nulle sur F , alors $\pi(f)T\pi(g)$. La C^* -algèbre de X est l'adhérence de l'ensemble des opérateurs $T \in \mathcal{L}(H)$ dont le support est un entourage et tels que, pour tout $f \in C_0(X)$, $T\pi(f)$ et $\pi(f)T$ sont compacts.

Il y a aussi une conjecture de Baum-Connes qui prédit la K -théorie de cette C^* -algèbre (cf. [43]), du moins si X est à géométrie à l'infini bornée. Rappelons qu'un espace métrique X est dit à *géométrie à l'infini bornée* si, pour tout $r \in \mathbf{R}_+$, il existe $N \in \mathbf{N}$ tel que le nombre d'éléments de toute partie de X de diamètre $\leq r$ est borné par N .

Le lien avec la conjecture de Baum-Connes est le suivant : soit G un groupe finiment engendré. Munissons-le de la distance invariante à gauche associée à la longueur de mots (par rapport à un système fini de générateurs). Si G n'a pas de

⁽⁶⁾ C'est Higson et Roe dans [21] qui ont reformulé le théorème de [51] en termes de moyennabilité.

torsion et BG est un complexe cellulaire fini, la conjecture de Baum-Connes à l'infini pour l'espace métrique G implique l'injectivité de $\mu_{G,r}$.

En utilisant une variante de la méthode de Higson-Kasparov, Yu ([51], voir aussi [46] pour une autre démonstration de ce résultat) montre :

THÉORÈME 1. — *Tout espace métrique à géométrie à l'infini bornée qui admet un plongement uniforme dans un espace hilbertien satisfait la « conjecture de Baum-Connes à l'infini ».*

BIBLIOGRAPHIE

- [1] C. ANANTHARAMAN-DELAROCHE, *Classification des C^* -algèbres purement infinies nucléaires (d'après E. Kirchberg)*, Séminaire Bourbaki, Vol. 1995/96, exp. n° 805, Astérisque n° **241** (1997), 7-27.
- [2] C. ANANTHARAMAN-DELAROCHE et J. RENAULT, *Amenable groupoids*, (preprint 1999) à paraître à Ens. Math.
- [3] M.F. ATIYAH, *Elliptic operators discrete groups and von Neumann algebras*, Astérisque **32-33** (1976), 43-72.
- [4] P. BAUM and A. CONNES, *K-theory for Lie groups and foliations*, (preprint 1982) à paraître à Ens. Math.
- [5] P. BAUM, A. CONNES and N. HIGSON, *Classifying space for proper actions and K-theory of group C^* -algebras*, Contemporary Math. **167** (1994), 241-291.
- [6] P. BAUM, N. HIGSON and R. PLYMEN, *A proof of the Baum-Connes conjecture for p -adic $GL(n)$* , C. R. Acad. Sci. Paris, Sér. I Math. **325** n° 2 (1997), 171-176.
- [7] J.B. BOST, *Principe d'Oka, K-théorie et systèmes dynamiques non commutatifs*, Inv. Math. **101** (1990), 261-333.
- [8] J. CHABERT, *Stabilité de la conjecture de Baum-Connes pour certains produits semi-directs de groupes*, C. R. Acad. Sci. Paris, Sér. I Math. **328** n° 12 (1999), 1129-1132.
- [9] A. CONNES, *Sur la théorie non commutative de l'intégration*, Lect. Notes in Math. **725** (1979), 19-143.
- [10] J. CUNTZ, *K-theoretic amenability for discrete groups*, J. Reine ang. Math. **344** (1983), 180-195.
- [11] J. CUNTZ, *Bivariante K-théorie für lokalconvexe Algebren und der Chern-Connes Character*, Doc. Math. **2** (1997), 139-182.
- [12] C. DELAROCHE et A.A. KIRILLOV, *Sur les relations entre l'espace dual d'un groupe et la structure de ses sous-groupes fermés (d'après D. A. Kazhdan)*, Sém. Bourbaki, 1967/68, exp. n° 343, Soc. Math. France vol. **10** (1995), 507-528.
- [13] T. FACK, *K-théorie bivariante de Kasparov*, Sém. Bourbaki, (1982/83), Astérisque **105-106** Soc. Math. France (1983), 149-166.
- [14] M. GROMOV, *Geometric reflections on the Novikov conjecture*, Novikov conjectures, index theorems and rigidity, Vol. 1 (Oberwolfach, 1993), London Math. Soc. Lecture Note Ser. **226**, Cambridge Univ. Press (1995), 164-173.
- [14] M. GROMOV, *Spaces and questions*, (preprint 1999).

- [16] U. HAAGERUP, *An example of a nonnuclear C^* -algebra which has the metric approximation property*, *Inv. Math.* **50** (1979), 279-293.
- [17] P. DE LA HARPE, *Groupes hyperboliques, algèbres d'opérateurs et un théorème de Jolissaint*, *C. R. Acad. Sci. Paris, Sér. I Math.* **307** (1988), 771-774.
- [18] N. HIGSON, *Bivariant K -theory and the Novikov conjecture*, (preprint 1999).
- [19] N. HIGSON and G. KASPAROV, *Operator K -theory for groups which act properly and isometrically on Hilbert space*, *Electronic Research Announcements, AMS* **3** (1997), 131-141.
- [20] N. HIGSON, V. LAFFORGUE and G. SKANDALIS, *Counterexamples to the Baum-Connes Conjectures*, (preprint en préparation).
- [21] N. HIGSON and J. ROE, *Amenable group actions and the Novikov conjecture*, (preprint 1999).
- [22] P. JOLISSAINT, *Rapidly decreasing functions in reduced C^* -algebras of groups*, *Trans. Amer. Math. Soc.* **317** (1990), 167-196.
- [23] P. JOLISSAINT, *K -theory of reduced C^* -algebras and rapidly decreasing functions on groups*, *K -theory* **2** (1989), 723-735.
- [25] P. JULG, *Remarks on the Baum-Connes conjecture and Kaszdan's property T* , *Fields Inst. Comm.* **13** (1997), 145-153.
- [24] P. JULG, *Travaux de N. Higson et G. Kasparov sur la conjecture de Baum-Connes*, *Sém. Bourbaki*, vol. 1997/98, exp. n° 841, *Astérisque* n° **252** (1998), 151-183.
- [26] P. JULG et A. VALETTE, *Fredholm modules associated to Bruhat-Tits Buildings*, *Proc. of the Center for Math. Analysis, Australian National University* **16** (1988), 143-155.
- [27] G.G. KASPAROV, *Hilbert C^* -modules : theorems of Stinespring and Voiculescu*, *J. Operator Theory* **4** (1980), 133-150.
- [28] G.G. KASPAROV, *The operator K -functor and extensions of C^* -algebras*, *Math. USSR Izv.* **16** (1981), n° 3, 513-572. *Translated from : Izv. Akad. Nauk. S.S.S.R. Ser. Mat.* **44**, (1980), 571-636.
- [29] G.G. KASPAROV, *Equivariant KK -theory and the Novikov conjecture*, *Inv. Math.* **91** (1988), 147-201.
- [30] G.G. KASPAROV and G. SKANDALIS, *Groups acting on buildings, Operator K -theory and Novikov's conjecture*, *K -theory* **4** (1991), 303-337.
- [31] G.G. KASPAROV and G. SKANDALIS, *Groupes «boliques» et conjecture de Novikov*, *Note C.R.A.S.* **319** Sér. I (1994), 815-820.
- [32] D. KAZHDAN, *Connection of the dual space of a group with the structure of its closed subgroups*, *Funct. Anal. and its Appl.* **1** (1967), 63-65.
- [33] V. LAFFORGUE, *Une démonstration de la conjecture de Baum-Connes pour les groupes réductifs sur un corps p -adique et pour certains groupes discrets possédant la propriété (T)* , *C. R. Acad. Sci. Paris, Sér. I Math.* **327**, n° 5 (1998), 439-444.
- [34] V. LAFFORGUE, *A proof of property (RD) for discrete cocompact subgroups of $SL_3(\mathbf{R})$* , (preprint 1998) à paraître à *Journal of Lie theory*.
- [35] V. LAFFORGUE, *K -théorie bivariante pour les algèbres de Banach et conjecture de Baum-Connes*, (preprint 1998).
- [36] V. LAFFORGUE, *Espaces de Schwartz*, (preprint 1998).

- [37] V. LAFFORGUE, *K-théorie bivariante pour les algèbres de Banach, groupoïdes et conjecture de Baum-Connes*, (preprint en préparation 1999).
- [38] P-Y LE GALL, *Théorie de Kasparov équivariante et groupoïdes*, C. R. Acad. Sci. Paris, Sér. I Math. **324**, n° 6 (1997), 695-698.
- [39] A.S. MIŠČENKO, *Homotopy invariance of non simply connected manifolds, I : Rational invariance*, Math. USSR Izv. **4** (1970), 509-519, transl from *Izv. Akad. Nauk. SSSR Ser. Math.* **34**, (1970), 501-514.
- [40] H. OYONO-OYONO, *La conjecture de Baum-Connes pour les groupes agissant sur les arbres*, C. R. Acad. Sci. Paris, Sér. I Math. **326**, n° 7 (1998), 799-904.
- [41] J. RAMAGGE, G. ROBERTSON and T. STEGER, *A Haagerup inequality for $\tilde{A}_1 \times \tilde{A}_1$ and \tilde{A}_2 buildings*, Geometric and Funct. Anal. **8** n° 4 (1998), 702-731.
- [42] J.N. RENAULT, *A groupoid approach to C^* -algebras*, Lecture Notes in Math. **793**, Springer-Verlag, New York, (1980).
- [43] J. ROE, *Index Theory, Coarse Geometry, and Topology of Manifolds*, CBMS Regional Conf. Series in Math. **90**, AMS, (1996).
- [44] G. SKANDALIS, *Une notion de nucléarité en K-théorie (d'après J.Cuntz)*, K-theory **1** (1988), 549-573.
- [45] G. SKANDALIS, *Approche de la conjecture de Novikov par la cohomologie cyclique. D'après Connes-Gromov-Moscovici*, Séminaire Bourbaki, vol. 1990/91, exposé n° 739, Astérisque **201-202-203** (1992), 299-316.
- [46] G. SKANDALIS, J.L. TU and G. YU, *Coarse Baum-Connes conjecture and Groupoids*, (preprint 1999).
- [47] J.L. TU, *La conjecture de Novikov pour les feuilletages hyperboliques*, K-theory **16**, n° 2 (1999), 129-184.
- [48] J.L. TU, *La conjecture de Novikov pour les feuilletages moyennables*, K-theory **17**, n° 3 (1999), 215-264.
- [49] A. VALETTE, *Questions, Novikov conjectures, index theorems and rigidity*, Vol. 1 (Oberwolfach, 1993), London Math. Soc. Lecture Note Ser. **226**, Cambridge Univ. Press (1995), p. 74.
- [50] A. WASSERMANN, *Une démonstration de la conjecture de Connes-Kasparov pour les groupes de Lie linéaires connexes réductifs*, C. R. Acad. Sci. Paris, Sér. I Math. **304** (1987), 559-562.
- [51] G. YU, *The coarse Baum-Connes conjecture for spaces which admit uniform embeddings into Hilbert space*, (preprint 1999) à paraître à Inv. Math.

Georges SKANDALIS

UMR 7586 du C.N.R.S.

Université de Paris VII

U.F.R. de Mathématiques

Case Postale **7012**

2, place Jussieu

F-75251 Paris cedex 05 PARIS

E-mail : skandal@math.jussieu.fr

THÉORIE DES MODÈLES ET CONJECTURE
DE MANIN-MUMFORD
[d'après Ehud Hrushovski]

par Elisabeth BOUSCAREN

1. INTRODUCTION

Je vais présenter ici l'une des applications récentes de la théorie des modèles à la Géométrie Diophantienne, la nouvelle démonstration, due à E. Hrushovski ([Hr2]), de la Conjecture de Manin-Mumford (généralisée) qui est basée sur l'étude, du point de vue de la théorie des modèles, des corps algébriquement clos munis d'un automorphisme.

THÉORÈME 1.1 (Conjecture de Manin-Mumford). — *Soit G un groupe algébrique commutatif connexe défini sur un corps de nombres k et X une sous-variété de G . Alors il y a un nombre fini de points de torsion de G , a_1, \dots, a_M , et de sous-groupes algébriques de G , G_1, \dots, G_M tels que pour chaque i , $a_i + G_i$ est contenu dans X et*

$$X(k^{alg}) \cap \text{Tor}(G) = \bigcup_{i=1}^M a_i + \text{Tor}(G_i).$$

De plus on peut borner de façon effective le nombre M de translatés: il existe c, e deux constantes ne dépendant que d'invariants liés à G (et non à X) et du choix de deux premiers de bonne réduction pour G , tels que $M \leq c \deg(X)^e$.

En utilisant des méthodes de calcul assez grossières, on obtient que les constantes c et e sont doublement exponentielles.

La démonstration de Hrushovski passe par une première étape où il prend un premier de bonne réduction pour G , de caractéristique résiduelle p , et montre le résultat pour la torsion première à p , c'est-à-dire pour le sous-groupe $\text{Tor}_{p'}(G)$ des éléments de torsion de G dont l'ordre n'est pas divisible par p . On obtient dans ce cas que

$$X(k^{alg}) \cap \text{Tor}_{p'}(G) = \bigcup_{i=1}^N a_i + \text{Tor}_{p'}(G_i),$$

où il est facile de montrer directement que $N \leq a \deg(X)^b$, avec a doublement exponentiel (en la dimension de A , le degré de l'addition dans A et la cardinalité du corps résiduel) et b exponentiel en la dimension de A .

Le but de cet exposé, en présentant la démonstration de Hrushovski, est d'essayer d'expliquer pourquoi la théorie des modèles a quelque chose à dire sur ce type de questions et aussi de donner une idée des résultats qu'on montre et des outils qu'on développe en théorie des modèles.

Nous passerons donc rapidement sur les parties purement "géométrie algébrique" de la démonstration.

De même, la très brève présentation historique et bibliographique qui suit n'a, cela sera visible tout de suite, aucune prétention d'exhaustivité.

1.1. Petit historique et conjecture de Lang

La conjecture de Manin-Mumford originelle porte sur l'intersection d'une courbe avec les points de torsion d'une variété abélienne. Elle a été tout d'abord démontrée par M. Raynaud ([Ra1]), puis R. Coleman ([Col]) en a donné une autre démonstration par des méthodes très différentes; Raynaud a ensuite ([Ra2]) démontré le résultat pour une sous-variété quelconque, puis M. Hindry ([Hi1]), en s'inspirant d'une idée de S. Lang ([La1]) pour le cas des courbes, a démontré le résultat général pour un groupe algébrique commutatif.

On trouve une autre approche de ces questions dans les travaux sur la Conjecture de Bogomolov, on se contentera ici de renvoyer à l'exposé de A. Abbes sur le sujet dans le cadre de ce séminaire ([Ab]) et aux travaux récents de S. David et P. Philippon ([DaPh]).

En revanche, nous allons dire quelques mots de la Conjecture de Lang (dite aussi Conjecture de Mordell-Lang), énoncée par Lang dans [La1] et qui regroupe dans un cadre général la conjecture de Manin-Mumford et la Conjecture de Mordell (voir aussi par exemple [La2]).

Nous l'énonçons ici pour la caractéristique zéro:

Conjecture de Lang (absolue): *Soient K un corps algébriquement clos de caractéristique zéro, A une variété abélienne définie sur K , X une sous-variété de A définie sur K et Γ un sous-groupe de $A(K)$ de rang fini. Alors il existe $\gamma_1, \dots, \gamma_n \in \Gamma$ et B_1, \dots, B_n sous-variétés abéliennes de A tels que $\gamma_i + B_i \subset X$ et*

$$X(K) \cap \Gamma = \cup_{i=1}^n \gamma_i + (B_i(K) \cap \Gamma).$$

Cette conjecture a finalement été démontrée par G. Faltings ([Fa]) après de nombreux travaux entre autres de Manin, Vojta, Laurent, Hindry et Faltings lui-même. On peut trouver des bibliographies récentes, commentées et complètes, dans [Hi2] ou

[Maz]. L'extension au cas des variétés semi-abéliennes est due à Vojta([Voj]) et McQuillan ([McQ]).

Si on veut étendre cette conjecture au cas de caractéristique non nulle, il faut se contenter d'une version relative, dite aussi, Conjecture de Lang pour les corps de fonctions. En voici un énoncé un peu simplifié:

Conjecture de Lang pour les corps de fonctions: *Soient $K_0 < K$ deux corps algébriquement clos, A une variété abélienne définie sur K de K/K_0 -trace zéro, et X une sous-variété de A définie sur K . Soit Γ un sous-groupe de rang fini de $A(K)$. Alors il existe $\gamma_1, \dots, \gamma_n \in \Gamma$ et B_1, \dots, B_n sous-variétés abéliennes de A tels que $\gamma_i + B_i \subseteq X$ et*

$$X(K) \cap \Gamma = \cup_{i=1}^n \gamma_i + (B_i(K) \cap \Gamma).$$

On rappelle que Γ est de *rang fini* si il existe un groupe Γ_0 finiment engendré tel que, pour chaque $\gamma \in \Gamma$, il y a un entier $n \geq 1$ tel que $n\gamma \in \Gamma_0$. Si on est en caractéristique $p > 0$, on demande que l'entier n ne soit pas divisible par p . Dire que A est de K/K_0 -trace zéro est équivalent à dire que A n'a aucune sous-variété abélienne homomorphe à une variété abélienne définie sur K_0 .

Dans [Hr1], Hrushovski a donné une démonstration modèle-théorique de la conjecture de Lang pour les corps de fonctions en toute caractéristique (on peut trouver une présentation de ces résultats dans l'exposé [Go] de ce séminaire, ou dans [Pi3], et une exposition plus détaillée dans le livre [Bo]). Pour la caractéristique $p > 0$, il s'agissait de la première démonstration du cas général, les résultats précédents (voir [AbrVo]) nécessitant des hypothèses supplémentaires. Un peu auparavant, A. Buium ([Bu]) avait donné une nouvelle démonstration du cas de caractéristique zéro, dont l'idée de départ était de passer à un corps différentiellement clos et de remplacer le groupe Γ par un groupe Δ -fermé (ou différentiellement algébrique) de dimension finie. C'est cette idée qui est l'inspiration des démonstrations de Hrushovski, pour les deux résultats, Lang pour les corps de fonctions et Manin-Mumford.

1.2. Le rapport avec la théorie des modèles

L'idée qui est à la base de ces deux démonstrations est donc la suivante: on veut remplacer le sous-groupe Γ ($Tor(G)$ dans notre cas) par un sous-groupe "fermé" ou "algébrique", pour lequel on montrera le résultat. On ne peut pas le faire sans, d'une manière ou d'une autre, rajouter des "fermés" supplémentaires (c'est-à-dire de la structure supplémentaire), puisque, quand G par exemple est une variété abélienne, la torsion est dense dans G . C'est exactement ce que fait Buium dans [Bu], quand il passe dans un corps différentiellement clos, remplace le groupe Γ par un groupe Δ -fermé de dimension finie et utilise ensuite l'arsenal des outils développés dans ce cadre. Or c'est là une méthode qui rentre tout à fait dans la problématique développée en théorie des modèles et pour

laquelle nous disposons d'un cadre systématique, d'outils et de résultats déjà existants. En effet, puisant son inspiration dans la géométrie algébrique et les géométries combinatoires, la théorie des modèles développe dans un cadre abstrait des notions d'indépendance, de dimension et de géométrie. Cela a permis d'isoler une notion qui va se retrouver au centre de toutes ces applications récentes: la notion de *modularité* qui caractérise, à partir du comportement de la relation d'indépendance, les groupes dans lesquels les seuls sous-ensembles qu'on peut définir sont les (combinaisons booléennes finies de) translatés de sous-groupes. Montrer que le groupe $Tor(G)$ satisfait l'énoncé de Manin-Mumford, c'est exactement, une fois qu'on s'est placé dans le bon cadre, montrer que c'est un groupe modulaire, nous le verrons dans les sections 2.2 et 2.3; et l'on a des outils pour cela.

1.3. Le cadre dans lequel on travaille

On a donc un groupe algébrique, G , défini sur un corps de nombres k . On va considérer un automorphisme σ dans $Gal(k^{alg}/k)$, judicieusement choisi, de manière à ce que la torsion soit contenue dans le sous-groupe S_σ des solutions, dans $G(k^{alg})$, d'une équation faisant intervenir σ (ce qu'on appelle traditionnellement une équation de différence) qu'on peut explicitement calculer. On plonge alors (k^{alg}, σ) dans un gros corps K sur lequel σ se prolonge et qui est clos en ce sens qu'il contient déjà des solutions pour toute équation de différence qui a une solution dans une extension. C'est ce qu'on appellera un corps de différence générique. Et, de même qu'on a la topologie de Zariski sur un corps algébriquement clos ou la Δ -topologie sur un corps différentiellement clos, on définit une σ -topologie sur K . La théorie des modèles permet d'analyser la structure de K avec ces nouveaux ensembles géométriques. Cette analyse a été faite par Z. Chatzidakis et E. Hrushovski dans [ChHr] qui montrent qu'on peut y définir une bonne notion d'indépendance et de dimension. En particulier, ils démontrent en caractéristique zéro un théorème de Dichotomie, qui dit à peu près que les seuls ensembles de dimension finie qui ne sont pas "modulaires" se trouvent dans le corps fixé par σ . À partir de cette analyse et de ce théorème de dichotomie, Hrushovski trouve un critère "effectif" pour distinguer, parmi les sous-groupes de $G(K)$ définis par des équations de différence, ceux qui sont modulaires. Il ne reste plus qu'à vérifier que le sous-groupe S_σ judicieusement choisi plus haut est du bon type, puis à borner le nombre de translatés qui vont intervenir par des calculs simples à partir de l'équation qui définit S_σ .

C'est exactement comme cela que cela se passe dans le cas de la torsion première à p . Pour la torsion totale, c'est un peu plus compliqué, car, s'il est facile, en travaillant à partir de deux premiers de bonne réduction distincts, de montrer qu'il existe une σ -équation qui s'annule sur la torsion toute entière et est du bon type, on ne sait pas le faire de manière effective. Donc, si on veut garder des bornes sur le nombre de translatés, il faut travailler un peu plus, en gardant deux équations distinctes, l'une pour la torsion première à p , l'autre pour la p -torsion puis combiner les deux.

Dans une première partie, nous allons présenter le cadre modèle théorique, c'est-à-dire les corps de différence génériques et les groupes que l'on peut y définir avec des équations de différence (section 2). Ensuite nous expliquerons comment utiliser les résultats ainsi obtenus pour démontrer la conjecture de Manin-Mumford (section 3). En fait, nous traiterons le cas de la torsion première à p et ne dirons que quelques mots sur le passage à la torsion totale. Enfin nous terminerons en mentionnant d'autres applications de la même méthode (section 4).

Merci à celles et ceux qui ont bien voulu relire tout ou partie de ce texte, ou répondre aux nombreuses questions auxquelles je ne pouvais manquer d'être confrontée, puisqu'il me faut parler ici non seulement de théorie des modèles mais aussi de géométrie algébrique, domaine qui n'est pas le mien. En particulier, merci à Z. Chatzidakis, F. Delon, D. Bertrand, M. Hindry et aussi à E. Hrushovski. Je suis évidemment seule responsable des erreurs qui se seraient ici glissées dans un énoncé de théorème ou dans une esquisse de preuve, à la suite d'une tentative indue de simplification.

2. LES CORPS DE DIFFÉRENCE GÉNÉRIQUES

2.1. Définition et existence

Nous appellerons **corps de différence** un corps K muni d'un automorphisme distingué σ . On peut par exemple considérer $K = k^{alg}$, la clôture algébrique d'un corps k , muni d'un automorphisme $\sigma \in Gal(k^{alg}/k)$ ou encore, si K est un corps parfait de caractéristique $p > 0$ et $n \geq 1$, K muni de l'automorphisme de Frobenius $x \mapsto x^{p^n}$.

Les corps de différence ont été à l'origine étudiés par Ritt dans les années 30; on peut trouver tous les résultats algébriques de base dans le livre de R. Cohn [Coh] (dans ce livre, un corps de différence est un corps avec un monomorphisme distingué; quand il s'agit d'un automorphisme, le corps est appelé corps de différence inversif).

Il était naturel pour des théoriciens des modèles de s'intéresser à la classe des corps de différence **existentiellement clos** c'est-à-dire tels que tout système fini d'équations de différence ayant une solution dans une extension du corps en a déjà une dans le corps. C'est l'analogue, pour les corps de différence, des corps algébriquement clos pour les corps, des corps différentiellement clos pour les corps différentiels, ou bien encore des corps réels clos pour les corps ordonnés. Les premières propriétés (axiomatisation, décidabilité etc) ont été étudiées par A. Macintyre, L. van den Dries et C. Wood (voir [Ma1]). Ensuite, une étude plus complète de la structure de ces corps du point de vue de la théorie des modèles a été faite par Z. Chatzidakis et E. Hrushovski dans un premier temps ([ChHr]), puis par les mêmes avec un troisième auteur, K. Peterzil ([ChHrPe]). C'est dans le premier de ces deux articles que se trouvent les résultats qui sont utilisés pour la démonstration de Manin-Mumford, notamment le résultat de dichotomie en caractéristique zéro qui est crucial.

Enfin, dans l'article où il donne sa démonstration de la conjecture de Manin-Mumford ([Hr2]), Hrushovski commence par pousser plus loin l'étude des groupes commutatifs définissables dans ces corps de différence. Dès la prochaine section, pour simplifier les énoncés, nous nous limiterons à la caractéristique zéro et aux résultats ayant un rapport direct avec les applications dont nous voulons parler. Pour en savoir plus sans se plonger dans les articles eux-mêmes, on pourra consulter [Ch2].

Commençons par décrire ces corps, que nous appellerons ici pour simplifier **corps de différence génériques** (au départ, dans [ChHr], ils apparaissent comme modèles d'une théorie appelée ACFA, "algebraically closed fields with an automorphism"; on les appelle aussi souvent corps algébriquement clos avec un automorphisme générique).

Convention et notation: nous utiliserons le terme de variété uniquement dans le cas irréductible. Si X est une variété affine définie sur K , X^σ est la variété affine définie en appliquant σ aux coefficients des équations qui définissent X .

DÉFINITION 2.1. — Soit (K, σ) un corps de différence. On dit que (K, σ) est un **corps de différence générique** si K est un corps algébriquement clos vérifiant la propriété:

(*) Soit X une variété affine définie sur K et Y une sous-variété de $X \times X^\sigma$ définie sur K , dont les projections sur les facteurs X et X^σ sont denses. Alors il existe $a \in X(K)$ tel que $(a, \sigma(a)) \in Y(K)$.

Les corps algébriquement clos satisfaisant (*) sont bien exactement les corps de différence existentiellement clos et tout corps de différence se plonge dans un corps de différence générique.

Ces corps, comme on va le voir dans la suite, sont munis d'une structure plus riche que celle induite uniquement par la topologie de Zariski mais que l'on sait quand même analyser!

Commençons par quelques propriétés élémentaires:

PREMIÈRES PROPRIÉTÉS: Soit (K, σ) un corps de différence générique. Le corps K est de degré de transcendance infini sur le corps premier. Le corps fixé par σ dans K , $Fix(\sigma)$, est un corps pseudo-fini. Le seul invariant nécessaire pour obtenir les complétions (au sens de la théorie des modèles) de la théorie des corps algébriquement clos est la caractéristique; pour les corps de différence génériques, il faut préciser en plus la classe de conjugaison de σ dans le groupe des automorphismes de la clôture algébrique du corps premier.

Enfin, mentionnons un résultat frappant qui vient confirmer l'importance de cette classe de corps. Cela faisait plusieurs années que se posait la question de ce que pouvait bien être la théorie d'un "Frobenius non-standard", par exemple la limite des corps de différence $(\mathbb{F}_p^{alg}, x \mapsto x^p)$, pour p tendant vers l'infini. Il s'agit de l'analogue, pour les automorphismes de Frobenius, de la question résolue par Ax quand il a montré que la théorie élémentaire des corps finis était exactement la théorie des corps pseudo-finis [Ax]. La question précise ici était de savoir s'il s'agissait des corps de différence génériques. Une réponse

positive a été indépendamment donnée par Hrushovski ([Hr3]) et Macintyre ([Ma2]): la théorie des corps de différence génériques est exactement la théorie de tous les ultraproducts non principaux de $(\mathbb{F}_q^{alg}, x \mapsto x^q)$, quand q varie sur l'ensemble des puissances de nombre premiers.

2.2. Ensembles définissables, indépendance

À partir de maintenant, (K, σ) sera un corps de différence générique de caractéristique zéro. Les résultats des sections 2.2 et 2.3 ont été démontrés dans [ChHr].

Comme d'habitude en Théorie des modèles (voir par exemple le chapitre d'introduction de [Bo]), on considère un corps de différence générique (K, σ) comme une structure du premier ordre. Cela veut dire que l'on étudie les ensembles que l'on peut définir à partir des opérations et des fonctions de base, dans ce cas $\{+, -, \cdot, 0, 1, \sigma\}$, en prenant la clôture par intersection et réunion finies, par complémentaire et par projection.

Dans un "pur" corps algébriquement clos K , il est naturel de commencer par considérer les sous-ensembles de K^n définis par des équations polynômiales, les fermés de Zariski. Ici il est naturel de considérer les σ -polynômes: $K[X_1, \dots, X_n]_\sigma$, **l'anneau des polynômes de différence** (ou σ -polynômes) est l'anneau des polynômes sur K en une infinité de variables, $X_1, \dots, X_n, \sigma(X_1), \dots, \sigma(X_n), \sigma^2(X_1), \dots, \sigma^2(X_n) \dots$. Un σ -fermé de K^n est l'ensemble des zéros d'un nombre fini de σ -polynômes de $K[X_1, \dots, X_n]_\sigma$. Un σ -idéal I de $K[X_1, \dots, X_n]_\sigma$ est un idéal clos par σ et on dit que I est **parfait** si chaque fois que $a^j \sigma^i(a) \in I$, $i, j \in \mathbb{N}$, alors a appartient à I . Les σ -fermés correspondent exactement aux σ -idéaux parfaits et comme on a la condition de chaîne ascendante sur les σ -idéaux parfaits (voir [Coh]), on définit ainsi une topologie noetherienne contenant strictement la topologie de Zariski.

On peut alors considérer les σ -**constructibles**, les combinaisons booléennes finies de σ -fermés. Mais là, contrairement à ce qui se passe avec la topologie de Zariski sur un corps algébriquement clos ou bien avec la Δ -topologie sur un corps différentiellement clos, la classe des σ -constructibles n'est pas close par projection. On clôt donc aussi par projection et on obtient ce qu'on appelle la classe des ensembles σ -**définissables**. Si V est une variété affine définie sur K , l'ensemble $V(K)$ des points K -rationnels de V est un σ -fermé dans un produit K^n . Pour une variété V définie sur K et donnée par un recouvrement fini de cartes affines, $V(K)$ est un ensemble σ -définissable. Le corps $Fix(\sigma)$ (défini par l'équation $\sigma(x) = x$) est un exemple de σ -fermé qui n'est pas de la forme $V(K)$. Nous dirons qu'une application est σ -définissable si son graphe est σ -définissable.

Bien que cela ne soit pas évident a priori, la classe des σ -définissables est également close, à bijection σ -définissable près, par quotient. Cela veut dire précisément que si $F \subseteq K^n$ est σ -définissable et que $R \subseteq K^n \times K^n$ est une relation d'équivalence σ -définissable sur K^n , alors il existe une bijection σ -définissable entre l'ensemble quotient F/R et un

sous-ensemble σ -définissable de K^m pour un certain m . On dit des structures qui ont cette propriété qu'elles **éliminent les imaginaires**.

On sera amené à s'intéresser à la **structure induite** par (K, σ) sur les sous-ensembles σ -définissables: si $D \subset K^n$ est σ -définissable, la structure induite sur D est D , muni, pour tout $m > 0$, de la trace sur D^m de tous les sous-ensembles σ -définissables de K^{nm} .

On montre que la structure induite sur le corps $Fix(\sigma)$ se réduit (une fois qu'on a fixé un certain nombre de constantes dans $Fix(\sigma)$) à la pure structure de corps: si D est un sous-ensemble σ -définissable de K^n , alors $D \cap Fix(\sigma)^n$ est un sous-ensemble de $Fix(\sigma)^n$ définissable dans le pur langage des corps (c'est-à-dire constructible au sens habituel du terme, combinaison booléenne de fermés de Zariski).

Il est classique maintenant en théorie des modèles de regarder s'il est possible de définir une bonne notion d'indépendance dans les structures qu'on étudie, du type de l'indépendance algébrique dans les corps algébriquement clos. C'est bien le cas ici:

DÉFINITION 2.2. — 1. Si A est un sous-ensemble de K , la **σ -clôture algébrique** de A , $acl_\sigma(A)$, est égale à la clôture algébrique (au sens habituel des corps) du sous-corps de différence de (K, σ) engendré par A , c'est-à-dire à la clôture algébrique de $\mathbb{Q}(\{\sigma^i(a); a \in A, i \in \mathbb{Z}\})$.

2. Si $A, B, C \subset K$, on dit que A et B sont **indépendants** au-dessus de C si $acl_\sigma(AC)$ et $acl_\sigma(BC)$ sont linéairement disjoints au-dessus de $acl_\sigma(C)$.

3. Si E est un sous-corps de différence de K et si a est une suite finie d'éléments de K , on définit le **σ -degré** de a au-dessus de E , $d_\sigma(a/E)$ comme étant le degré de transcendance de $E(a)_\sigma$, le corps de différence engendré par $E(a)$, au-dessus de E . Si $D \subset K^n$ est un sous-ensemble σ -définissable, on définit le σ -degré de D comme étant le maximum des degrés des éléments de D .

Le corps $Fix(\sigma)$ est de σ -degré égal à un; en revanche si V est une variété (de dimension positive) définie sur K , $V(K)$ est de σ -degré infini. Quand il est fini, le σ -degré est une bonne notion de dimension, en particulier il est clair que si $E = acl_\sigma(E)$ et si $d_\sigma(a/E)$ est fini, alors a et b sont σ -indépendants au-dessus de E si et seulement si $d_\sigma(a/E) = d_\sigma(a/acl_\sigma(E(b)))$. Il est possible (et indispensable) de définir d'autres notions de dimension ou de rang dans (K, σ) pouvant prendre des valeurs ordinales non finies, mais nous n'en parlerons pas ici. Ce qui est important pour nous, c'est que pour un a donné, ces différentes dimensions sont simultanément finies.

2.3. Modularité, le théorème de dichotomie

Comme nous l'avons dit dans l'introduction, cela fait plusieurs années que la pertinence de la notion abstraite d'ensemble "modulaire" (cette notion apparaît hélas dans la littérature sous plusieurs noms: localement modulaire, un-basé (one-based) etc.) s'est

imposée à travers les conclusions très précises que l'on peut déduire sur les propriétés algébriques des structures modulaires.

Nous allons donner les définitions et les résultats sous une forme adaptée à notre contexte précis. Mais il s'agit de cas particuliers ou d'adaptations de résultats valables dans un contexte beaucoup plus général.

Si $E = \text{acl}_\sigma(E) \subset K$ et si a, a' sont deux éléments de K^n , on dit que a et a' **ont même type (d'isomorphisme)** au-dessus de E si il existe un isomorphisme $\phi : \text{acl}_\sigma(E, a) \mapsto \text{acl}_\sigma(E, a')$ qui fixe E , envoie a sur a' et commute avec σ .

DÉFINITION 2.3. — *Un sous-ensemble σ -définissable D de K^n est*

- **stable** si, pour tout $E = \text{acl}_\sigma(E) \subset K$, pour toutes suites finies a, a' d'éléments de D ayant même type au-dessus de E , pour tout $F = \text{acl}_\sigma(F), E \subset F \subset K$, si a et F sont indépendants au-dessus de E et si a' et F sont indépendants au-dessus de E , alors a et a' ont aussi même type au-dessus de F ;

- **modulaire** si pour tout $E = \text{acl}_\sigma(E) \subset K$, pour toutes suites finies a, b d'éléments de D , $\text{acl}_\sigma(Ea)$ et $\text{acl}_\sigma(Eb)$ sont indépendants au dessus de $\text{acl}_\sigma(Ea) \cap \text{acl}_\sigma(Eb)$.

La stabilité veut donc dire que, si $E \subset K$ est σ -algébriquement clos, et si $a, b \in K$ alors il n'y a (à isomorphisme près) qu'une seule façon pour a et b d'être indépendants au-dessus de E .

Il y a des corps stables: les (purs) corps algébriquement clos, les corps séparablement clos ou encore les corps différentiellement clos. Mais les corps pseudo-finis ([Du]), donc en particulier $\text{Fix}(\sigma)$, ne sont pas stables. On le voit ici facilement: on peut trouver $E = \text{acl}_\sigma(E) \subset K$ et $a, b, c \in \text{Fix}(\sigma) \setminus E$, tels que a et c d'une part, b et c d'autre part sont indépendants au-dessus de E , mais tels que $\sqrt{a-c} \in \text{Fix}(\sigma)$ et $\sqrt{b-c} \notin \text{Fix}(\sigma)$.

Le corps $\text{Fix}(\sigma)$ n'est pas modulaire: on prend trois éléments de $\text{Fix}(\sigma)$, a, b, c transcendants sur \mathbb{Q} et algébriquement indépendants. Alors $\text{acl}_\sigma(a, b) = \mathbb{Q}(a, b)^{\text{alg}}$ et $\text{acl}_\sigma(c, ac + b) = \mathbb{Q}(c, ac + b)^{\text{alg}}$ s'intersectent en \mathbb{Q}^{alg} , et pourtant ils ne sont pas algébriquement indépendants au-dessus de \mathbb{Q}^{alg} . En fait, la non-modularité est la traduction "abstraite", dans le cas d'un ensemble de dimension un, de l'existence d'une famille de courbes planes (ici la famille des courbes $y = ax + b$) de dimension deux. En particulier, si un ensemble D est modulaire, alors cela entraîne que dans la structure induite par (K, σ) sur D , on ne peut pas définir de corps infini. On va voir un peu plus loin (2.6) que avec l'hypothèse de stabilité, cela entraîne beaucoup plus.

Mais ce qui est particulièrement intéressant dans le cas des corps de différence génériques, c'est que toute la non-modularité et la non-stabilité sont concentrées dans le corps fixé. C'est ce que nous dit le théorème de Dichotomie en caractéristique zéro.

On dit que deux ensembles σ -définissables D et F de (K, σ) sont **orthogonaux** si pour toute suite finie d d'éléments de D , pour toute suite finie b d'éléments de F , pour tout sous-corps $E = \text{acl}_\sigma(E)$ de K , d et b sont indépendantes au-dessus de E .

THÉORÈME 2.4. — Théorème de dichotomie: *Soit $D \subseteq K^n$ un sous-ensemble σ -définissable de σ -degré fini, alors D est stable et modulaire si et seulement si D et le corps fixé $\text{Fix}(\sigma)$ sont orthogonaux.*

La démonstration de ce théorème passe par une analyse des ensembles de dimension finie en termes de sous-ensembles de dimension un. Le même type de résultat avait été montré précédemment pour les corps différentiellement clos (avec le sous-corps des constantes, [HrSo]) et pour les corps séparablement clos de caractéristique $p > 0$ (avec le sous-corps des éléments infiniment p -divisibles, [Hr1]) mais par des méthodes différentes. Dans ces deux autres cas, la preuve utilise la théorie des Géométries de Zariski, version abstraite de la topologie de Zariski introduite par Hrushovski et Zil'ber ([HrZi], voir aussi [Mr]).

De façon générale, la dichotomie modulaire/non-modulaire est particulièrement utile dans le cas des groupes. En effet la modularité caractérise les structures de groupes stables de manière très précise. À nouveau nous énonçons dans 2.6 un résultat qui est vrai dans un contexte beaucoup plus général.

DÉFINITION 2.5. — *Soit $G \subseteq K^n$ un groupe σ -définissable. On dit que G est **de type abélien** si pour tout m et pour tout sous-ensemble σ -définissable X de K^{nm} , $X \cap G^m$ est une combinaison booléenne finie de translatés de sous-groupes connexes (σ -définissables) de G^m .*

Il s'ensuit que le groupe G a un sous-groupe (σ -définissable) abélien d'indice fini et donc en fait que la structure induite par (K, σ) sur G se réduit à une structure de type "module généralisé".

PROPOSITION 2.6 ([HrPi1]). — *Soit $G \subset K^n$ un groupe σ -définissable stable. Alors G est modulaire si et seulement si G est de type abélien*

On voit maintenant le rapport entre la modularité et les questions de type conjecture de Manin-Mumford ou plus généralement conjecture de Lang. On a d'ailleurs bien une équivalence formelle de la conjecture de Manin-Mumford avec l'énoncé de théorie des modèles suivant: pour tout groupe algébrique commutatif connexe G défini sur \mathbb{Q}^{alg} , la structure $(\mathbb{Q}^{alg}, \text{Tor}(G))$, dans le langage des anneaux avec un prédicat pour le groupe $\text{Tor}(G)$, est stable et le groupe $\text{Tor}(G)$ (qui est maintenant définissable dans cette structure) est modulaire (voir [Pi1]). Mais savoir cela ne nous donne pas pour autant une méthode pour montrer que cela est vrai. L'intérêt de passer par les corps de différence génériques c'est que là, on a un vrai critère utilisable pour reconnaître quand un groupe est modulaire grâce à la dichotomie.

2.4. Les sous-groupes σ -définissables des variétés abéliennes

Nous résumons ici les conséquences de l'étude des groupes commutatifs σ -définissables et en particulier des sous-groupes σ -définissables des points K -rationnels des variétés

abéliennes faite par Hrushovski dans [Hr2]. On peut aussi en trouver une exposition détaillée dans [Ch1].

L'une des étapes essentielles de cette étude est l'analyse de l'anneau des *endomorphismes σ -définissables* de $A(K)$ pour A une variété abélienne définie sur K .

DÉFINITION 2.7. — *Soit G un groupe algébrique défini sur K . Un **endomorphisme σ -définissable** de $G(K)$ est une application σ -définissable de $G(K)$ dans $G(K)$ qui est un homomorphisme du groupe $G(K)$.*

Les endomorphismes du groupe algébrique G qui sont définis sur K induisent des endomorphismes de $G(K)$ au sens de la définition ci-dessus. Il y en a en général beaucoup d'autres mais Hrushovski montre par exemple:

PROPOSITION 2.8. — *Soient A une variété abélienne simple définie sur K , $End_\sigma(A(K))$ l'anneau des endomorphismes σ -définissables de $A(K)$ et $End_{alg}(A(K))$ le sous-anneau des endomorphismes qui sont induits par les endomorphismes (algébriques) de la variété abélienne A .*

(1) *Si A n'est pas isogène à A^{σ^n} , alors*

$$\mathbb{Q} \otimes End_\sigma(A(K)) = \mathbb{Q} \otimes End_{alg}(A(K)).$$

(2) *Si non, $\mathbb{Q} \otimes End_\sigma(A(K))$ est isomorphe à un anneau de polynômes tordu au-dessus de $\mathbb{Q} \otimes End_{alg}(A(K))$ (que nous ne décrirons pas ici précisément). Il s'ensuit en particulier que $End_\sigma(A(K))$ est dénombrable, comme $End_{alg}(A(K))$.*

(3) *Pour chaque sous-groupe σ -définissable G de $A(K)$, il existe $f \in End_\sigma(A(K))$ tel que G est un sous-groupe d'indice fini de $Ker(f)$. Il s'ensuit qu'il n'y a qu'un nombre dénombrable de sous-groupes σ -définissables de $A(K)$.*

Des résultats similaires sont montrés pour $A = \mathbb{G}_m$, le groupe multiplicatif. On a alors $\mathbb{Q} \otimes End_\sigma(\mathbb{G}_m(K)) \simeq \mathbb{Q}[\sigma, \sigma^{-1}]$.

Soit A une variété semi-abélienne définie sur $Fix(\sigma)$ et $F[T] \in \mathbb{Z}[T]$ un polynôme à coefficients entiers. Alors $F(\sigma)$ induit naturellement un endomorphisme (σ -définissable) de $A(K)$: si $F[T] = \sum_{i=0}^r m_i T^i$, alors $F(\sigma)(a) = m_0 a + m_1 \sigma(a) + \dots + m_r \sigma^r(a)$, où $+$ est l'addition dans A et $ma = [m]a$ la multiplication par l'entier m dans A .

DÉFINITION 2.9. — *Les groupes B et C sont **commensurables** si $B \cap C$ est d'indice fini dans B et dans C . Un groupe σ -définissable B est **c-minimal** si tout sous-groupe σ -définissable infini de B est d'indice fini dans B .*

PROPOSITION 2.10. — *Supposons que A est une variété abélienne simple définie sur $Fix(\sigma)$ ou bien que A est le groupe multiplicatif \mathbb{G}_m . Soit B un sous-groupe σ -définissable*

de $A(K)$ de σ -degré fini et c -minimal. Alors B n'est pas de type abélien si et seulement si il existe n tel que $B \subseteq \text{Ker}(\sigma^n - 1)$.

ESQUISSE DE DÉMONSTRATION DE 2.10:

C'est bien sûr le Théorème de dichotomie (2.4) et l'équivalence (2.6) entre groupe stable modulaire et groupe de type abélien que nous allons utiliser.

Si B est inclus dans $\text{Ker}(\sigma^n - 1)$, c'est-à-dire $B \subseteq A(\text{Fix}(\sigma^n))$, alors comme $\text{Fix}(\sigma^n)$ est une extension finie de $\text{Fix}(\sigma)$, il existe une application σ -définissable à fibres finies de $\text{Fix}(\sigma)^n$ sur B , qui n'est donc pas orthogonal à $\text{Fix}(\sigma)$ et n'est donc pas de type abélien.

Réciproquement, supposons que B n'est pas de type abélien. Par 2.6 et 2.4, B et $\text{Fix}(\sigma)$ ne sont pas orthogonaux.

FAIT 2.11. — Soit B un sous-groupe σ -définissable de σ -degré fini de $H(K)$, où H est un groupe algébrique défini sur K . Si B n'est pas orthogonal à $\text{Fix}(\sigma)$, alors il existe un sous-groupe normal D σ -définissable d'indice infini dans B , un entier $m \geq 1$ et une surjection σ -définissable (du produit cartésien) $(\text{Fix}(\sigma))^m$ sur B/D .

Dans notre cas, la c -minimalité de B entraîne que D est fini. On en déduit l'existence d'une application injective σ -définissable h de B/D dans un produit de $\text{Fix}(\sigma)$. On transporte la loi de groupe de B/D par h sur $C = h(B/D)$. Le groupe C est alors σ -définissable, mais on sait que tout sous-ensemble σ -définissable dans $\text{Fix}(\sigma)$ est définissable dans le pur langage des corps. On peut donc utiliser des résultats antérieurs sur les groupes définissables dans les corps pseudo-finis ([HrPi2]) qui permettent de conclure qu'il existe un homomorphisme σ -définissable, de noyau fini, de C sur un sous-groupe d'indice fini de $G(\text{Fix}(\sigma))$, pour G un groupe algébrique défini sur $\text{Fix}(\sigma)$. On en déduit l'existence d'un homomorphisme σ -définissable g de B sur un sous-groupe d'indice fini de $G(\text{Fix}(\sigma))$, g de noyau fini contenant D . On peut supposer (en remplaçant B par un sous-groupe d'indice fini) que G est connexe, que $g(B) = G(\text{Fix}(\sigma))$ et donc que G est un groupe algébrique commutatif simple. Le graphe de g , qui est un sous-groupe σ -définissable de $(H \times G)(K)$ est commensurable avec le noyau d'un endomorphisme σ -définissable (par 2.8). Il n'y a qu'un nombre dénombrable de tels endomorphismes, et il suit que g est défini sur une extension finie k_1 de $\text{Fix}(\sigma)$. On peut donc bien trouver un n tel que σ^n fixe B point par point.

On peut maintenant en déduire le corollaire qui va être au centre de la démonstration de la conjecture de Manin-Mumford, en utilisant les lemmes suivants qui permettent en particulier de se ramener au cas d'une variété abélienne simple ou du groupe multiplicatif:

LEMME 2.12. — a) Si on a une suite exacte d'homomorphismes de groupe σ -définissables

$$0 \rightarrow A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow 0,$$

où A_1, A_2, A_3 sont des groupes σ -définissables, alors A_2 est stable modulaire si et seulement si A_1 et A_3 sont stables modulaires.

b) Si A est une variété abélienne simple définie sur K et si H est un sous-groupe σ -définissable de $A(K)$, alors H est c -minimal si et seulement si il existe $f \in \text{End}_\sigma A(K)$, f irréductible dans $\mathbb{Q} \otimes \text{End}_\sigma A(K)$, tel que H et $\text{Ker}(f)$ sont commensurables.

c) Si B est un sous-groupe σ -définissable de $A(K)$, pour A une variété abélienne définie sur K , et si B est commensurable avec $\text{Ker}(f_1 \cdots f_m)$ pour $f_1, \dots, f_m \in \text{End}_\sigma(A(K))$, alors B est stable modulaire si et seulement si tous les $\text{Ker}(f_i)$ sont stables modulaires.

Si $F(T) \in \mathbb{Z}[T]$ est un polynôme dont aucune racine n'est racine de l'unité, nous dirons que $F(T)$ est un polynôme sans facteur cyclotomique.

COROLLAIRE 2.13. — *Soit A une variété semi-abélienne définie sur $\text{Fix}(\sigma)$ et $F(T) \in \mathbb{Z}[T]$. Soit $H_\sigma = \text{Ker}(F(\sigma)) = \{a \in A(K); \sum_{i=0}^r m_i \sigma^i(a) = 0\}$. Alors le groupe H_σ est de type abélien si et seulement si $F(T)$ est sans facteur cyclotomique.*

ESQUISSE DE DÉMONSTRATION DE 2.13:

Tout d'abord H_σ est bien de σ -degré fini. Ensuite, en utilisant 2.12, on peut se ramener aux deux cas où A est une variété abélienne simple ou bien où $A = \mathbb{G}_m$, le groupe multiplicatif.

Si pour un $N > 0$ ($T^N - 1$) et $F(T)$ ont un facteur commun $P(T)$, alors $B = \text{Ker}(P(\sigma)) \subseteq \text{Ker}(\sigma^N - 1)$ n'est pas modulaire. Mais $B \subseteq H_\sigma$ qui n'est donc pas modulaire non plus.

Supposons maintenant que H_σ n'est pas de type abélien et, pour simplifier (en fait il faut utiliser b) et c) du Lemme 2.12), qu'il est c -minimal et que $F(\sigma) = f$ est un élément irréductible de $\mathbb{Q} \otimes \text{End}_\sigma(A)$. Par 2.10 à nouveau, il existe un N tel que $\text{Ker}(f) \subseteq C = \text{Ker}(\sigma^N - 1)$. Mais en choisissant N suffisamment grand on peut supposer que f agit sur C et $F(T)$ étant sans facteur cyclotomique, f devrait être inversible sur C .

3. APPLICATION À LA CONJECTURE DE MANIN-MUMFORD

Nous sommes maintenant en mesure d'expliquer comment appliquer le théorème de Dichotomie et son corollaire sur les groupes définis à partir d'un polynôme sans facteur cyclotomique (2.13) pour obtenir une nouvelle démonstration de la conjecture de Manin-Mumford. En fait, comme nous l'avons déjà dit dans l'introduction, nous n'allons vraiment traiter précisément que le cas de la torsion première à p , cas dans lequel l'application de 2.13 donne immédiatement le résultat et qui permet de comprendre pourquoi ce type de preuve fournit naturellement des bornes.

3.1. Préliminaires sans théorie des modèles

Comme on l'a déjà expliqué, la théorie des modèles va nous permettre de montrer que les ensembles qui nous intéressent sont des combinaisons booléennes finies de translatés de sous-groupes et l'on voudra en conclure que en fait ce sont des réunions finies de translatés. Remarquons donc tout de suite une fois pour toute l'équivalence des différentes versions que nous pourrons rencontrer:

LEMME 3.1. — Soit G un groupe algébrique commutatif connexe défini sur un corps algébriquement clos L , X une sous-variété de G également définie sur L et Γ un sous-groupe de $G(L)$. Les énoncés suivants sont équivalents:

- (i) $X \cap \Gamma$ est une combinaison booléenne finie de translatés de sous-groupes de Γ ,
- (ii) la clôture de Zariski de $X \cap \Gamma$ est une réunion finie de translatés de sous-groupes algébriques connexes de G ,
- (iii) $X \cap \Gamma$ est contenu dans une réunion finie de translatés C_1, \dots, C_n de sous-groupes algébriques connexes de G , chaque C_i étant contenu dans X ,
- (iv) $X \cap \Gamma$ est une réunion finie de translatés de sous-groupes de Γ .

Maintenant voici le principal fait "algébrique" que l'on va utiliser et sur lesquels nous ne nous étendrons pas ici.

Soit A une variété abélienne de dimension d définie sur un corps de nombres k et \mathfrak{p} un premier de l'anneau \mathfrak{K} des entiers de k , de corps résiduel $k_{\mathfrak{p}}$ de cardinalité q et de caractéristique p . On suppose que \mathfrak{p} est un **premier de bonne réduction pour A** . On a donc que $A_{\mathfrak{p}}$, la réduction de A modulo \mathfrak{p} , est une variété abélienne (définie sur $k_{\mathfrak{p}}$) de même dimension que A . On rappelle que $Tor_{p'}(A) = \{a \in Tor(A); p \text{ ne divise pas l'ordre de } a\}$ et que $Tor_p(A) = \{a \in Tor(A); p^n a = 0 \text{ pour un } n \geq 1\}$.

FAIT 3.2. — *Il existe un automorphisme $\sigma \in Gal(\mathbb{Q}^{alg}/k)$ et un polynôme à coefficients entiers, $F[T] \in \mathbb{Z}[T]$, sans facteur cyclotomique, tel que l'endomorphisme $F(\sigma)$ de $A(\mathbb{Q}^{alg})$ s'annule sur $Tor_{p'}(A)$. De plus le degré de $F(T)$ est inférieur ou égal à $2d$ et la somme des valeurs absolues de ses coefficients est bornée par $(1 + q^{1/2})^{2d}$.*

Des résultats classiques de Weil (voir [We]) sur les endomorphismes des variétés abéliennes définies sur les corps finis et le polynôme caractéristique de l'automorphisme de Frobenius, entraînent l'existence d'un tel polynôme $F[T] \in \mathbb{Z}[T]$ tel que $F(\Phi_q)$ s'annule sur $A_{\mathfrak{p}}(k^{alg})$, où Φ_q est l'automorphisme de Frobenius $\Phi_q : x \mapsto x^q$.

On peut alors relever cette équation fonctionnelle de la façon suivante: on prend pour σ un relèvement de Φ_q et, \mathfrak{p} étant un premier de bonne réduction, on déduit du lemme de Hensel que la réduction modulo \mathfrak{p} induit un isomorphisme de $Tor_{p'}(A)$ sur $Tor_{p'}(A_{\mathfrak{p}})$ et donc que $F(\sigma)$ s'annule sur $Tor_{p'}(A)$.

REMARQUE: Il est possible de faire un peu mieux en travaillant à partir des facteurs simples de A et en ne comptant qu'une seule fois les facteurs isogènes. On peut ainsi

borner le degré du polynôme $F[T]$ et la valeur absolue de ses coefficients en remplaçant d , la dimension de A , par un invariant de A , $d_r(A)$ qui est inférieur ou égal à la dimension de A et tel que, pour tout $n \geq 1$, $d_r(A) = d_r(A^n)$.

3.2. La torsion première à p

Nous allons donner la démonstration dans le cas des variétés abéliennes. Le cas des variétés semi-abéliennes est identique, il suffit de vérifier qu'on peut généraliser sans peine l'existence du polynôme adéquat sans facteur cyclotomique (3.2). En revanche, il faut encore travailler et utiliser de la théorie des modèles pour passer au cas d'un groupe commutatif arbitraire, nous l'expliquerons dans la section 3.2.2

3.2.1. Les variétés abéliennes. Soient A une variété abélienne de dimension d définie sur un corps de nombres k et X une sous-variété de A . On fixe un plongement de A dans un espace projectif \mathbb{P}_n et on peut ainsi considérer le degré de A et de ses sous-ensembles algébriques.

On fixe un premier \mathfrak{p} de bonne réduction pour A dont le corps résiduel est de cardinalité q et de caractéristique p .

PROPOSITION 3.3. — *Alors*

$$X \cap \text{Tor}_{\mathfrak{p}'}(A) = \bigcup_{i=1}^M a_i + \text{Tor}_{\mathfrak{p}'}(B_i),$$

où chaque B_i est une sous-variété abélienne de A et

$$M \leq c (\text{deg}(X))^{(2d+1)(2^d \dim(X))}$$

où la constante c ne dépend que d'invariants liés à A (et non à X et à son corps de définition) et de q , et est deux fois exponentielle en d .

DÉMONSTRATION:

1. POUR OBTENIR LA RÉUNION FINIE:

On considère l'automorphisme σ de $\text{Gal}(\mathbb{Q}^{\text{alg}}/k)$ et le polynôme $F(T) \in \mathbb{Z}[T]$, $F(T) = \sum_{i=0}^{2d} c_i T^i$, donnés par 3.2. Soit (K, σ) un corps de différence générique extension du corps de différence $(\mathbb{Q}^{\text{alg}}, \sigma)$. Alors

$$H_\sigma := \text{Ker}(F(\sigma) \upharpoonright_{A(K)}) = \{a \in A(K); \sum_{i=0}^{2d} c_i \sigma^i(a) = 0\}$$

est un sous-groupe de $A(K)$ défini par une σ -équation.

Puisque le polynôme $F(T)$ est sans facteur cyclotomique, 2.13 nous assure que H_σ est un sous-groupe de type abélien et donc que tous ses sous-ensembles σ -définissables sont des combinaisons booléennes finies de translatés de sous-groupes σ -définissables. C'est donc le cas en particulier pour $X \cap H_\sigma$ qui lui même contient $X \cap \text{Tor}_{\mathfrak{p}'}(A)$. Comme nous l'avons remarqué plus haut, cela entraîne que la clôture de Zariski de $X \cap H_\sigma$, que nous

appellerons Z , est réunion finie de translatées de sous-variétés abéliennes de A , chacune contenue dans X , $Z = \bigcup_{i=1}^M b_i + B_i$.

On en déduit que $X \cap \text{Tor}_{p'}(A)$ est réunion de au plus M translatés de la forme $a + \text{Tor}_{p'}(B)$ avec B sous-variété abélienne de A .

2. POUR BORNER LE NOMBRE M DE TRANSLATÉES:

On va en fait borner le nombre de composantes irréductibles de Z , la clôture de Zariski de $X \cap H_\sigma$, en utilisant la définition explicite de H_σ à partir du polynôme $F[T]$.

On considère

$$S = \{(a_0, \dots, a_{2d}) \in A^{2d+1}; \sum_{i=0}^{2d} c_i a_i = 0\}.$$

Donc $H_\sigma = \{a \in A(K); (a, \sigma(a), \dots, \sigma^{2d}(a)) \in S\}$. Soit $U = S \cap (X \times X^\sigma \times \dots \times X^{\sigma^{2d}})$ (chaque X^{σ^i} est une sous-variété de $A^{\sigma^i} = A$), alors

$$X \cap H_\sigma = \{a \in A(L); (a, \sigma(a), \dots, \sigma^{2d}(a)) \in U\}.$$

Pour faire le calcul, on va utiliser une définition un peu modifiée du degré pour les sous-variétés de $(\mathbb{P}_n)^m$, pour $m > 1$, (voir [Fu], exemple 8.4.4). On définit aussi le degré d'un fermé non irréductible comme la somme des degrés de (toutes) ses composantes irréductibles. Ce qui est important pour nous ici, c'est que le degré borne bien le nombre de composantes irréductibles.

Les propriétés du degré entraînent que

$$\text{deg}(U) \leq \text{deg}(S)(\text{deg}(X)^{2d+1}) \leq \text{deg}(S)(\text{deg}(A)^{2d+1})$$

et on a que $\dim(U) \leq \min\{\dim(S), (2d+1)\dim(X)\} \leq d(2d+1)$.

En utilisant le fait que le corps (K, σ) est un corps de différence générique, et donc satisfait la condition (*) de 2.1, on montre que:

LEMME 3.4. — Soit $r > 0$, E un sous-ensemble algébrique de \mathbb{P}_n^{r+1} ,

$$E_\sigma := \{a \in \mathbb{P}_n(K); (a, \sigma(a), \dots, \sigma^r(a)) \in E\}$$

et V la clôture de Zariski de E_σ . Alors $\text{deg}(V) \leq (\text{deg}(E))^{2\dim(E)}$.

Dans notre cas on obtient donc que $M \leq \text{deg}(Z) \leq \text{deg}(U)^{2\dim(U)}$, ce qui donne bien une borne du type annoncé, une fois qu'on a calculé le degré de S .

3.2.2. Groupes algébriques commutatifs. Il n'est pas très difficile de généraliser à tous les groupes algébriques commutatifs le Fait 3.2 et donc l'existence du polynôme $F(T)$ sans facteur cyclotomique et de l'automorphisme σ de \mathbb{Q}^{alg} adéquats. On considère comme plus haut un corps de différence générique (K, σ) étendant $(\mathbb{Q}^{alg}, \sigma)$.

La généralisation de la Proposition 3.3 aux variétés semi-abéliennes est alors immédiate puisque le corollaire central de la dichotomie (2.13) est vrai pour les variétés semi-abéliennes.

Mais dans le cas d'un groupe commutatif arbitraire 2.13 n'est plus vrai. En effet, si V est un sous-groupe vectoriel du groupe algébrique G , alors $Ker F(\sigma) \cap V(K)$ ne peut pas être de type abélien: tout sous-groupe σ -définissable de degré fini de $V(K)$ est un espace vectoriel σ -définissable de dimension finie sur le corps $Fix(\sigma)$ et ne lui est donc pas orthogonal.

Mais on peut séparer la partie de type abélien et la partie espace vectoriel sur $Fix(\sigma)$ et cette dernière n'intervient pas dans le cas de la torsion. Plus précisément:

DÉFINITION 3.5. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et V le sous-groupe vectoriel maximal de G . Un sous-ensemble σ -définissable D de $G(K)$ est **spécial** s'il est de la forme $D = C + W$, avec C un translaté d'un sous-groupe σ -définissable de $G(K)$ et W un sous-ensemble σ -définissable de $V(K)$.*

PROPOSITION 3.6. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et $H_\sigma = \{g \in G(K); F(\sigma)(g) = 0\}$ pour un polynôme sans facteur cyclotomique $F(T) \in \mathbb{Z}[T]$. Alors tout sous-ensemble σ -définissable de H_σ est une combinaison booléenne finie de sous-ensembles spéciaux de $G(K)$.*

LES INGRÉDIENTS DE LA DÉMONSTRATION DE 3.6:

On considère la suite exacte $0 \rightarrow V \rightarrow G \rightarrow B \rightarrow 0$, où V est le sous-groupe vectoriel maximal de G et B est donc une variété semi-abélienne. Maintenant par 2.13, $H_B := H_\sigma/V(K) \cap H_\sigma$ est de type abélien mais $H_V := V(K) \cap H_\sigma$ est, lui, un espace vectoriel sur $Fix(\sigma)$. Le théorème de Dichotomie (2.4) entraîne que, dans la structure (K, σ) , un groupe de type abélien de σ -degré fini et un groupe vectoriel σ -définissable ne peuvent qu'être orthogonaux. Il n'y a donc aucune relation possible entre éléments de H_B et de H_V . On peut alors montrer que tout sous-ensemble σ -définissable de $H_B \times H_V$ est une réunion finie de rectangles $B_i \times V_i$ avec $B_i \subseteq H_B$ et $V_i \subseteq H_V$ et dans notre situation particulière, en déduire que tout sous-ensemble σ -définissable de H_σ est une combinaison booléenne finie de sous-ensembles σ -définissables spéciaux de la forme: $B + W$, où B est un translaté d'un sous-groupe σ -définissable de $G(K)$ et W est un sous-ensemble σ -définissable de $V(K)$.

COROLLAIRE 3.7. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et X une sous-variété de G . Alors la clôture de Zariski de $X \cap T_p(G)$ est une réunion finie de translatés de sous-groupes algébriques de G .*

DÉMONSTRATION DE 3.7:

On considère $H_\sigma = \{g \in G(K); F(\sigma)(g) = 0\}$ pour le polynôme $F(T) \in \mathbb{Z}[T]$ donné par 3.2 (généralisé aux groupes commutatifs arbitraires). Par 3.6 $X \cap H_\sigma$ est combinaison booléenne de sous-ensembles spéciaux. Par passage à la clôture de Zariski, on obtient une réunion finie de sous-variétés spéciales de G , c'est-à-dire de sous-variétés de la forme $B+W$ où B est un translaté d'un sous-groupe algébrique connexe de G et W est une sous-variété de V , le sous-groupe vectoriel maximal de G . Il suffit maintenant de montrer que, si $C = B + W$ est une sous-variété spéciale de G d'intersection non vide avec $Tor_{p'}(G)$, alors la clôture de Zariski de $Tor_{p'}(G) \cap C$ est un translaté d'un sous-groupe algébrique de G . Par hypothèse, $B = g + E$ où E est un sous-groupe algébrique connexe de G . On peut se ramener au cas où $E \cap V = 0$ et on vérifie alors que $Tor_{p'}(G) \cap C$ est un translaté de $Tor_{p'}(E)$.

3.3. La torsion totale

3.3.1. Le cas des variétés semi-abéliennes. On a vu dans la section précédente comment, pour une variété semi-abélienne, utiliser un premier de bonne réduction \mathfrak{p} et puis un corps de différence générique $(L_{\mathfrak{p}}, \sigma_{\mathfrak{p}})$ pour “enfermer” le groupe des éléments de torsion première à p dans un groupe de type abélien en obtenant des bornes raisonnables. Puisque $Tor(A) = Tor_{p'}(A) \oplus Tor_p(A)$, l'idée est d'utiliser un second premier de bonne réduction \mathfrak{t} , de caractéristique résiduelle t différente de p , pour “enfermer” la torsion première à t et donc en particulier la p -torsion. C'est effectivement ce que fait Hrushovski, mais le problème va être de garder des bornes effectives. Si on cherche seulement le résultat qualitatif, il y a plusieurs preuves rapides utilisant cette méthode que l'on peut trouver par exemple dans [Pi2], [Ch2] ou [Hr2]. Nous allons en indiquer une dans la section suivante où l'on se trouve confronté à la question du calcul de $[L : k]$ où $L = k(Tor_{p'}(A)) \cap k(Tor_p(A))$.

Pour maîtriser les bornes, Hrushovski met en évidence des familles algébriques de composantes irréductibles, pour lesquelles il faut montrer une version uniforme de 3.4. Nous n'allons pas le faire ici. On peut trouver des énoncés et des calculs précis dans [Hr2] bien sûr mais aussi dans [Ch2]. On peut également en trouver une version “uniforme mais sans les calculs de bornes” dans [Pi2].

Il est important de signaler que, même si c'est moins frappant dans ce cas que dans le cas de la torsion première à p , l'existence de ces bornes effectives découle encore naturellement de la nature même de la démonstration. L'ordre de grandeur indiqué est ici aussi obtenu en effectuant les calculs avec des méthodes assez grossières.

3.3.2. Le cas des groupes commutatifs. Pour ce qui est du passage à un groupe commutatif arbitraire, on commence par remarquer que, une fois le résultat “qualitatif” connu, pour ce qui concerne le calcul des bornes, le cas général se réduit au cas des variétés semi-abéliennes: soit G un groupe algébrique commutatif défini sur un corps de nombres k et X une sous-variété de G . Soit V le sous-groupe vectoriel maximal de V , $A := G/V$

la variété semi-abélienne correspondante et Y l'image de X dans A . Alors le nombre de composantes irréductibles de la clôture de Zariski de $X \cap \text{Tor}(G)$ est borné par le nombre de composantes irréductibles de la clôture de Zariski de $Y \cap \text{Tor}(A)$. En effet, la projection de Y sur A est injective sur la torsion de G et induit donc une bijection entre les composantes irréductibles de la clôture de Zariski de $X \cap \text{Tor}(G)$ (dont on sait que ce sont des translatés de sous-groupes) et celles de $Y \cap \text{Tor}(A)$, (dont on sait aussi que ce sont des translatés de sous-groupes).

Il suffit donc de montrer, sans avoir à se préoccuper des bornes, que pour un groupe commutatif arbitraire il est bien vrai que l'intersection de X avec la torsion est réunion finie de translatés de sous-groupes algébriques. Comme nous l'avons dit plus haut, il y a plusieurs démonstrations possibles utilisant deux premiers de bonne réduction distincts.

Nous allons en donner une ici, si l'on admet un résultat de Jean-Pierre Serre sur l'indépendance des groupes de Galois l -adiques des variétés semi-abéliennes (voir [Se], pages 33-34 et 56-59 pour le cas des variétés abéliennes, merci à Jean-Pierre Serre de m'avoir confirmé la généralisation au cas des variétés semi-abéliennes), permet de conclure très rapidement. On a donc un groupe algébrique connexe commutatif défini sur un corps de nombres k et X une sous-variété de G . On se donne deux premiers de bonne réduction \mathfrak{p} et \mathfrak{t} , de caractéristiques résiduelles distinctes p et t , les polynômes $F_p(T)$ et $F_t(T)$ et les automorphismes $\sigma_p, \sigma_t \in \text{Gal}(\mathbb{Q}^{\text{alg}}/k)$ correspondants donnés par 3.2. Le résultat de Serre entraîne que $L = k(\text{Tor}_{\mathfrak{p}'}(G)) \cap k(\text{Tor}_{\mathfrak{p}}(G))$ est une extension galoisienne finie de k , au-dessus de laquelle $k(\text{Tor}_{\mathfrak{p}'}(G))$ et $k(\text{Tor}_{\mathfrak{p}}(G))$ sont linéairement disjoints. Donc si $m = [L : k]$, il existe $\tau \in \text{Gal}(\mathbb{Q}^{\text{alg}}/L)$ qui étend σ_p^m sur la torsion première à p et qui étend σ_t^m sur la p -torsion. Si f_1, \dots, f_{2d} sont les racines de $F_p(T)$ et g_1, \dots, g_{2d} sont les racines de $F_t(T)$, on pose $J(T) = \prod_{i=1}^{2d} (T - f_i^m)(T - g_i^m)$. Alors $J(\tau)$ s'annule sur la torsion de G . On se place dans un corps de différence générique (K, τ) qui étend $(\mathbb{Q}^{\text{alg}}, \tau)$; par 3.6, $X \cap \text{Ker}(J(\tau))$ est une combinaison booléenne finie de sous-ensembles spéciaux de $G(K)$. La même démonstration que pour 3.7 montre que alors la clôture de Zariski de $X \cap \text{Tor}(G)$ est bien une réunion finie de translatés de sous-groupes algébriques de G . Mais, pour ce qui concerne le calcul du nombre de translatés, la définition explicite de $J(T)$ en fonction des deux polynômes F_p et F_t dépend de $[L : k]$.

4. AUTRES APPLICATIONS

Nous terminons en mentionnant deux autres applications des critères de modularité dans les corps de différence génériques. La première est due à Hrushovski pour un premier cas, puis à T. Scanlon, et la seconde, qui elle utilise des résultats en caractéristique p démontrés dans [ChHrPe], à T. Scanlon.

4.1. La conjecture de Tate-Voloch

J. Tate et J.F. Voloch ont conjecturé l'existence d'une borne pour la distance des points de torsion d'une variété semi-abélienne à une sous-variété ([TaVo]):

CONJECTURE 4.1. — *Soit A une variété semi-abélienne définie sur \mathbb{C}_p et soit X une sous-variété de A . Il existe une constante c telle que si $a \in \text{Tor}(G)$, et que la distance p -adique de a à X est inférieure à c , alors $a \in X(\mathbb{C}_p)$.*

\mathbb{C}_p est la complétion de $\mathbb{Q}_p^{\text{alg}}$ pour la norme p -adique, qu'on note $|\cdot|_p$. Si X est une sous-variété de l'espace affine \mathbb{A}^n définie sur \mathbb{C}_p , on choisit un système générateur $\{f_1, \dots, f_n\}$ de l'idéal de définition de X et on définit alors la distance p -adique d'un point a de $\mathbb{A}^n(\mathbb{C}_p)$ à X , $d_p(a, X)$ comme étant le maximum des $|f_i(a)|_p$.

Dans [TaVo] la conjecture est montrée dans le cas des tores.

E. Hrushovski, dans un premier temps, a utilisé la même méthode que pour Manin-Mumford pour établir le résultat pour A définie sur $\mathbb{Q}_p^{\text{alg}}$ et pour la torsion première à p , avec l'hypothèse que p est de bonne réduction pour A (communication à Voloch). Cela a ensuite été généralisé par T. Scanlon, toujours par les mêmes méthodes, qui montre la conjecture pour toutes les variétés semi-abéliennes définies sur $\mathbb{Q}_p^{\text{alg}}$ dans [Sc1] et [Sc2].

Nous allons indiquer ici brièvement comment on peut déduire du corollaire 2.13 le résultat dans le cas d'une variété semi-abélienne A définie sur \mathbb{Q}_p , avec la condition que p est de bonne réduction pour A .

Par 3.2 on a un automorphisme σ tel que la torsion première à p , $\text{Tor}_{p'}(A)$ est contenu dans $\{a \in A(\mathbb{Q}_p^{\text{alg}}); F(\sigma(a)) = 0\}$, pour $F[T] \in \mathbb{Z}[T]$ sans facteur cyclotomique.

Considérons une suite $(a_i)_{i \in \mathbb{N}}$ d'éléments de $\text{Tor}_{p'}(A)$ tels que $d_p(a_i, X)$ tend vers zéro. On va montrer que pour presque tout i , $a_i \in X$.

Sinon il existe un ultrafiltre (non principal) \mathfrak{U} sur \mathbb{N} tel que $\{i \in \mathbb{N}; a_i \in X\} \notin \mathfrak{U}$. Soit R l'anneau des suites de norme p -adique bornée dans \mathbb{C}_p , et l'idéal I de R défini comme l'intersection des I_n , où $I_n = \{r \in R; \{i \in \mathbb{N}; |r(i)|_p \leq p^{-n}\} \in \mathfrak{U}\}$. Alors on peut montrer que $D = R/I$ est un corps, on a un plongement (diagonal) naturel $j : \mathbb{C}_p \hookrightarrow D$ et σ induit un automorphisme de D qu'on appellera aussi σ . Soit a^* l'image de la suite (a_0, \dots, a_m, \dots) dans D ; on voit que $a^* \in A(D)$, que $F(\sigma)(a^*) = 0$ et que $a^* \in X(D)$.

On passe dans un gros corps de différence générique (L, σ) étendant (D, σ) . Le corollaire 2.13 et le lemme 3.1 nous disent que $X(L) \cap \text{Ker}(F(\sigma))$ est contenu dans une réunion finie de translatées de sous-variétés semi-abéliennes de A , chaque translatée étant incluse dans X . En particulier $a^* \in c+B \subset X$, B sous-variété semi-abélienne de A . Soit $\pi : A \rightarrow A/B$. Alors la suite des $\pi(a_i)$ doit se rapprocher de $\pi(a^*)$. On a donc que, pour presque tout i , les $\pi(a_i)$ ont même image dans le corps résiduel. Par hypothèse de bonne réduction (appliquée à A/B), on a que la réduction est injective sur la p' -torsion de A/B . Donc la suite des $\pi(a_i)$ devient constante et égale à $\pi(a^*)$. Cela veut dire que pour presque tout i , $a_i \in c+B$, c'est-à-dire que pour presque tout i , $a_i \in X$.

4.2. Modules de Drinfeld

Enfin, plus récemment, T. Scanlon, en utilisant l'analyse modèle-théorique des corps de différence génériques de caractéristique p et le théorème de dichotomie (du type de 2.4) qui est montré dans [ChHrPe], a démontré l'équivalent de la conjecture de Manin-Mumford pour les modules de Drinfeld ([Sc3]).

BIBLIOGRAPHIE

- [Ab] A. Abbes – *Hauteurs et discrétude [d'après L. Szpiro, E. Ullmo et S. Zhang]*, Séminaire Bourbaki, exposé 825, Mars 1997.
- [AbrVo] D. Abramovic and F. Voloch – *Towards a proof of the Mordell-Lang conjecture in characteristic p* , Intern. Math. Research Notices (IMRN) No.2 (1992), 103-115.
- [Ax] J. Ax – *The elementary theory of finite fields*, Annals of Math. 88 (1968), 239-271.
- [Bo] E. Bouscaren ed. – *Model Theory and Algebraic Geometry*, Lecture Notes in Mathematics 1696, Springer 1998.
- [Bu] A. Buium – *Intersections in jet spaces and a conjecture of Serge Lang*, Annals of Math. 136 (1992), 583-593.
- [Ch1] Z. Chatzidakis – *Groups definable in ACFA*, in Algebraic Model Theory, B. Hart, A. Lachlan and M. Valeriote eds., NATO ASI Series, Kluwer Academic Publishers 1997.
- [Ch2] Z. Chatzidakis – *A survey on the model theory of difference fields*, in Model Theory, Algebra and Geometry, D. Haskell and C. Steinhorn ed., MSRI Publications 2000, 65-96.
- [ChHr] Z. Chatzidakis and E. Hrushovski – *The model theory of difference fields*, Transactions of the A.M.S, Vol. 351 (1999), 2997-3071.
- [ChHrPe] Z. Chatzidakis, E. Hrushovski and Y. Peterzil – *The model theory of difference fields II*, preprint 1999.
- [Coh] R.M. Cohn – *Difference algebra*, Tracts in Mathematics 17, Interscience Pub. 1965.
- [Col] R. Coleman – *p -adic integrals and torsion points on curves*, Annals of Math. 121 (1985), 111-168.
- [DaPh] S. David et P. Philippon – *Minorations des hauteurs normalisées des sous-variétés de variétés abéliennes*, in International Conference On Discrete Mathematics and Number Theory (Tiruchiparelli, 1996), K. Murty and M. Waldschmidt eds., Contemp. Math., 1998, 3-17.

- [Du] J.L. Duret – *Les corps faiblement algébriquement clos non séparablement clos ont la propriété d'indépendance*, in Model Theory of Algebra and Arithmetic, L. Pacholski et al. ed., Lecture Notes in mathematics 834, Springer 1980, 135-157.
- [Fa] G. Faltings – *The general case of Lang's conjecture*, in Symposium in Algebraic Geometry, V. Christante and W. Messing eds., Perspectives in Math. 15, Academic Press, 1994, 175-182.
- [Fu] W. Fulton – *Intersection Theory*, Ergebnisse 2, Springer 1984.
- [Go] J.B. Goode – *H.L.M. (Hrushovski-Lang-Mordell)*, Séminaire Bourbaki, exposé 811, Février 1996.
- [Hi1] M. Hindry – *Autour d'une conjecture de Serge Lang*, Invent. Math. 94 (1988), 575-603.
- [Hi2] M. Hindry – *Introduction to abelian varieties and the Mordell-Lang conjecture*, in Model Theory and Algebraic Geometry, E. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.
- [Hr1] E. Hrushovski – *The Mordell-Lang conjecture for function fields*, Journal of the AMS 9 (1996), 667-690.
- [Hr2] E. Hrushovski – *The Manin-Mumford conjecture and the model theory of difference fields*, preprint, 1995, à paraître dans Annals of Pure and Applied Logic.
- [Hr3] E. Hrushovski – *The first-order theory of the Frobenius*, preprint, 1995.
- [HrPi1] E. Hrushovski and A. Pillay – *Weakly normal groups*, in Logic Colloquium '85, North Holland 1987, 233-244.
- [HrPi2] E. Hrushovski and A. Pillay – *Groups definable in local fields and pseudo-finite fields*, Israel J.Math. 85 (1994), 203-262.
- [HrSo] E. Hrushovski and Ž. Sokolović – *Minimal subsets of differentially closed fields*, à paraître dans les Transactions of the AMS.
- [HrZi] E. Hrushovski and B. Zil'ber – *Zariski Geometries*, Journal of the A.M.S. 9 (1996), 1-56.
- [La1] S. Lang – *Division points on curves*, Ann. Mat. Pura Appl. (4) 70 (1965), 229-234.
- [La2] S. Lang – *Number Theory III, Diophantine Geometry*, volume 60, Encyclopaedia of Mathematical Sciences, Springer 1991.
- [Ma1] A. Macintyre – *Generic automorphisms of fields*, Annals of Pure and Applied Logic 88 vol.2-3 (1997), 165-180.
- [Ma2] A. Macintyre – *Non-standard Frobenius*, en préparation.
- [Maz] B. Mazur – *Abelian variety and the Mordell-Lang Conjecture*, in Model Theory, Algebra and Geometry, D. Haskell and C. Steinhorn ed., MSRI Publications 2000.

- [McQ] M. McQuillan – *Division points on semi-abelian varieties*, Invent. Math. 120 (1995), 143-159.
- [Mr] D. Marker – *Zariski geometries*, in Model Theory and Algebraic Geometry, E. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.
- [Mu] D. Mumford – *Abelian varieties*, Oxford University Press, Oxford 1985.
- [Pi1] A. Pillay – *The model-theoretic content of Lang’s conjecture* in Model Theory and Algebraic Geometry, E. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.
- [Pi2] A. Pillay – *ACFA and the Manin-Mumford conjecture*, in Algebraic Model Theory, B. Hart, A. Lachlan and M. Valeriote eds., NATO ASI Series, Kluwer Academic Publishers 1997.
- [Pi3] A. Pillay – *Model Theory and diophantine geometry*, Bull. Am. Math. Soc. 34 (1997), 405-422.
- [Ra1] M. Raynaud – *Courbes sur une variété abélienne et points de torsion*, Invent. Math. 71 (1983), 207-233.
- [Ra2] M. Raynaud – *Sous-variétés d’une variété abélienne et points de torsion*, in Arithmetic and Geometry, vol.I, M. Artin and J. Tate eds., Birkhäuser 1983, 327-352.
- [Sc1] T. Scanlon – *p-adic distance from torsion points of semi-abelian varieties*, Journal für die Reine und Angewandte Mathematik 499 (1998), 225-236.
- [Sc2] T. Scanlon – *The conjecture of Tate and Voloch on p-adic proximity to torsion*, Intern. Math. Research Notices (IMRN) No. 17 (1999), 909-914.
- [Sc3] T. Scanlon – *Diophantine geometry of the torsion of a Drinfeld module*, preprint 1998.
- [Se] J.P. Serre – *Oeuvres, Collected papers, Volume IV, 1985-1998*, Springer 2000.
- [TaVo] J. Tate and J.F. Voloch – *Linear forms in p-adic roots of unity*, International Mathematics Research Notices (IMRN) No.12 (1996), 589-601.
- [Voj] P. Vojta – *Integral points on subvarieties of semi-abelian varieties*, Invent. Math. 126 (1996), 133-181.
- [We] A. Weil – *Courbes algébriques et variétés abéliennes*, Hermann 1971.

Elisabeth BOUSCAREN

Université Denis-Diderot Paris 7
 Équipe de Logique Mathématique
 UFR de Mathématiques (case 7012)

2 place Jussieu

F-75251 Paris Cedex 05

E-mail : elibou@logique.jussieu.fr

RATIONAL ELLIPTIC CURVES ARE MODULAR
[after Breuil, Conrad, Diamond and Taylor]

by Bas EDIXHOVEN

1. INTRODUCTION

In 1994, Wiles and Taylor-Wiles proved that every semistable elliptic curve over \mathbb{Q} is modular, in the sense that it is a quotient of the jacobian of some modular curve (see [62], [59]). This work has been reported upon in this seminar in [49] and [40]; see especially [49, §1.2] for a historical account. As a consequence, Fermat's Last Theorem, known to be a consequence of this modularity result since work of Ribet based on a conjecture of Serre (see [39]), was finally proved. For a more detailed account of all this, see the book [14], and also [16]. Since 1994, this modularity result has been generalized by an increasing sequence of groups of authors: [23], [13], and [4].

THEOREM 1.1 (Diamond). — *Every elliptic curve over \mathbb{Q} that is semistable at 3 and 5 is modular.*

THEOREM 1.2 (Conrad, Diamond, Taylor). — *Every elliptic curve over \mathbb{Q} that acquires semistable reduction over a tame extension of \mathbb{Q}_3 is modular.*

THEOREM 1.3 (Breuil, Conrad, Diamond, Taylor). — *Every elliptic curve over \mathbb{Q} is modular.*

The method of the proofs is basically that of Wiles, i.e., for a given elliptic curve E over \mathbb{Q} one tries to prove that the mod l Galois representation $\bar{\rho}_{E,l}$ on $E(\overline{\mathbb{Q}})[l]$ is modular for some prime number l , and then that all lifts of $\bar{\rho}_{E,l}$ to l -adic representations of a suitable type are modular. The second step involves studying deformations of Galois representations, the systematic theory of which was initiated by Mazur, triggered by work of Hida. The key result for the first step is the celebrated theorem of Langlands [35] and Tunnell [60] that says that $\bar{\rho}_{E,3}$ is modular, as $\mathrm{GL}_2(\mathbb{F}_3)$ is solvable and has a faithful two-dimensional complex representation. The complications that arise in the proofs of the theorems above simply come from having to prove results as in Wiles and Taylor-Wiles, but with fewer hypotheses. In particular, choosing the right deformations of the

restriction of $\bar{\rho}_{E,l}$ to $G_l := \text{Gal}(\bar{\mathbb{Q}}_l/\mathbb{Q}_l)$ becomes much more complicated if E does not have semistable reduction at l .

The aim of this report is to give a reasonable sketch of the proofs of the theorems above, to describe the relation to some conjectures by Fontaine and Mazur and by Langlands, and to mention some related results. For some applications of the modularity results above, we refer to [15]. The author of this report does not claim to have checked the computations in [4], but he has studied [4] quite seriously and has not encountered any real problem. Let us also state the following theorem (Theorem B of [4]), whose proof is intricately linked to that of Theorem 1.3 above.

THEOREM 1.4 (Breuil, Conrad, Diamond, Taylor). — *Every irreducible continuous representation $\bar{\rho}: G_{\mathbb{Q}} \rightarrow \text{GL}_2(\mathbb{F}_5)$ with cyclotomic determinant is modular.*

2. RELATION WITH CONJECTURES BY LANGLANDS, FONTAINE AND MAZUR

The Langlands program predicts, among many other things, that all L -functions coming from algebraic geometry are in fact automorphic, i.e., arise from automorphic representations. More precisely, every absolutely irreducible motive of rank n over a number field F and with coefficients in a subfield E of $\bar{\mathbb{Q}}$ should correspond to a cuspidal algebraic automorphic representation of $\text{GL}_n(\mathbb{A}_F)$ with coefficients in E : see [9, Question 4.16], and the paragraph after that.

In that paragraph, Clozel explains how this conjecture relates to the conjecture of Hasse–Weil type that says that the L -function of such a motive extends meromorphically to all of \mathbb{C} and satisfies a certain functional equation. He finishes by remarking that the only cases for which the Hasse–Weil conjecture has been proved are cases where one actually proves the stronger conjecture, i.e., the existence of an automorphic representation; this remains true after the work of Wiles and its generalizations.

Of course, if E is an elliptic curve over \mathbb{Q} , the Langlands program predicts that E is modular. Hence the modularity theorem for elliptic curves over \mathbb{Q} is just a tiny part of the Langlands program.

Fontaine and Mazur stated the following conjecture (Conjecture 1 of [31]).

CONJECTURE 2.1. — *Let l be prime, $n \geq 0$, and let $\rho: \text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{GL}_n(\mathbb{Q}_l)$ be an irreducible continuous representation. Then ρ is isomorphic to a subquotient of some étale cohomology group $H^i(X_{\bar{\mathbb{Q}}}, \mathbb{Q}_l(r))$ with X a smooth projective variety over \mathbb{Q} , if and only if ρ satisfies the following two conditions:*

- (1) ρ is ramified at only finitely many primes;
- (2) the restriction $\rho|_{G_l}$ to a decomposition group at l is potentially semistable (see [30] for this notion).

In one direction, this conjecture has been proved: the $H^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_l(r))$ are known to be unramified at almost all primes, and the restriction to G_l is known to be potentially semistable by work of Tsuji and de Jong (see [1, §6.3.3]). It is the other direction that is even more spectacular: it is amazing that just these two conditions should imply, for example, that the Frobenius elements at almost all primes have eigenvalues that are algebraic numbers, and even Weil numbers, and that ρ should be part of a compatible system of l -adic representations. The evidence that one has today for this direction of the conjecture consists of the potentially abelian cases (treated in [31, §6]; ρ occurs in the tensor category generated by representations with finite image and representations which arise from potentially CM abelian varieties), and the cases treated by Wiles' method. However, see [38] for a representation that does satisfy the two conditions above, but for which one does not know if it satisfies Conjecture 2.1.

Combined with the Langlands program, Conjecture 2.1 implies (Conjecture 3c of [31]) that every 2-dimensional ρ satisfying the two conditions, up to Tate twist, either has a finite image, or arises from a modular form of weight at least two.

Since the space of modular forms of a given weight and level is finite dimensional, one also expects certain finiteness results concerning ρ as in Conjecture 2.1, which become semistable over a given extension of \mathbb{Q} , and are of fixed Hodge-Tate type: see [31, §3]. Most of [31] is in fact concerned with a deformation theoretic study of these finiteness conjectures.

Suppose now that $l > 2$. For a given absolutely irreducible continuous $\bar{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\mathbb{F}_l)$ one considers all lifts $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\mathbb{Z}_l)$ that are unramified outside a fixed set of primes. The l -adic variety (over \mathbb{Q}_l) of such lifts is conjecturally three dimensional. Now suppose moreover that $\bar{\rho}|_{G_l}$ is absolutely irreducible. Then the variety of lifts of $\bar{\rho}|_{G_l}$ is smooth and of dimension five by [41, Thm 4.1]. Since one expects the locus of global lifts that are potentially semistable and of a given type (i.e., Hodge-Tate type at l , and semistable over a fixed extension of \mathbb{Q}) to be zero dimensional, one expects that the locus of such local lifts is of codimension three in the five dimensional variety. Indeed, in the crystalline case with Hodge-Tate weights in the interval $[0, l-1]$, this was proved in [41] (moreover, the two-dimensional space is smooth). We note that, by [3], “potentially Barsotti-Tate” is equivalent to “potentially crystalline with Hodge-Tate weights in $[0, 1]$ ” (we recall that $l > 2$).

Of course, the computations done by Ramakrishna and by Fontaine and Mazur are not directly in terms of representations of G_l . Ramakrishna uses the results of Fontaine and Laffaille, and Fontaine and Mazur work with filtered (ϕ, N) -modules. We note that by recent work of Colmez and Fontaine, [10], one actually has an equivalence of tensor categories between semistable l -adic representations of G_l and weakly admissible filtered (ϕ, N) -modules, which makes it possible to translate problems on the Galois side into

problems in linear algebra, even more than before the equivalence between “weakly admissible” and “admissible” was known. On the other hand, what is still not available in this generality is a theory that works for \mathbb{Z}_l -lattices instead of \mathbb{Q}_l -vector spaces.

3. REVIEW OF WILES’ METHOD

Before turning to the work of Breuil, Conrad, Diamond and Taylor, let us review Wiles’ method. Good references for this part are [16], [14], [49], [40], and of course [62] (the introduction of which gives the story of the proof) and [59]. For simplicity, we only discuss this method in a case that suffices for modularity of semistable elliptic curves.

Let E be a semistable elliptic curve over \mathbb{Q} . The first observation is that there are many elliptic curves E' over \mathbb{Q} such that $E[5]$ and $E'[5]$ are symplectically isomorphic; this is due to the fact that the modular curve that parameterizes such E' (over \mathbb{Q} -schemes) is a non-empty open part of $\mathbb{P}_{\mathbb{Q}}^1$. One proves that there is such an E' , semistable, and such that the representation $\bar{\rho}_{E',3}: G_{\mathbb{Q}} \rightarrow \text{Aut}(E'[3](\bar{\mathbb{Q}})) \cong \text{GL}_2(\mathbb{F}_3)$ is surjective (see [49, §3]). By Langlands and Tunnell, the representation $\bar{\rho}_{E',3}$ is modular. The aim is now to show that $\rho_{E',3}: G_{\mathbb{Q}} \rightarrow \text{Aut}(E'(\bar{\mathbb{Q}})[3^{\infty}]) \cong \text{GL}_2(\mathbb{Z}_3)$ is modular, by showing that all 3-adic lifts of $\bar{\rho}_{E',3}$ with reasonable properties are modular, and hence so is E' . Before we discuss how that works, let us see how one then establishes the modularity of E itself.

Of course, if $\bar{\rho}_{E,3}$ is surjective, then we could have taken $E' = E$, so let us assume that $\bar{\rho}_{E,3}$ is not surjective. Then $E[3]$ is in fact reducible (this uses the semistability at all primes; see [49, Proposition 1]). But then $\bar{\rho}_{E,5}$ is irreducible, or $E_{\bar{\mathbb{Q}}}$ is isogeneous to the elliptic curve E_1 over $\bar{\mathbb{Q}}$ that has j -invariant $-5 \cdot 29^3 / 2^5$, as one sees by looking at the modular curve $X_0(15)$, which has genus one and exactly eight rational points, four of which are cusps (see [45, §2.1]). The elliptic curve E_1 has a model over \mathbb{Q} with conductor 50, which can be checked to be modular. Since modularity is invariant under isogeny and twisting, we may now assume that $\bar{\rho}_{E,5}$ is irreducible, and hence surjective ([49, Proposition 1]). In this case, we already know that $\bar{\rho}_{E,5}$ is modular, because E' is, and one proves the same type of result for modularity of 5-adic liftings as in the 3-adic case.

Let us now give a precise statement of these lifting results. We need some terminology and notation, adapted to the type of representations that we are interested in, i.e., those coming from modular forms of weight two. For each prime p , we choose an embedding $\bar{\mathbb{Q}} \rightarrow \bar{\mathbb{Q}}_p$, and we let G_p and I_p denote the corresponding decomposition and inertia subgroups of $G_{\mathbb{Q}}$. We let $\varepsilon: G_{\mathbb{Q}} \rightarrow \mathbb{Z}_l^*$ denote the l -cyclotomic character, given by the action on the elements of l -power order in $\bar{\mathbb{Q}}^*$.

DEFINITION 3.1. — *Let l be a prime number, and k a finite field of characteristic l . Let R be a complete local noetherian ring with residue field k , and let M be a free R -module of rank 2 with a continuous action by $G_{\mathbb{Q}}$; a choice of basis then gives a continuous*

representation $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(R)$. For p prime and different from l , M is called semistable at p if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$. The representation M is called Barsotti-Tate (at l) if for each finite quotient \overline{M} of M there exists a finite group scheme \overline{M} over \mathbb{Z}_l such that M and $\overline{M}(\overline{\mathbb{Q}}_l)$ are isomorphic as $\mathbb{Z}_l[G_l]$ -modules. The representation M is called semistable at l if it is Barsotti-Tate or if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$.

THEOREM 3.2 (Wiles, Taylor-Wiles). — Let $l \neq 2$ be a prime number. Let K be a finite extension of \mathbb{Q}_l , O its ring of integers, and k its residue field. Let $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(O)$ be an odd continuous representation such that:

- (1) its reduction $\overline{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(k)$ is modular and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible;
- (2) $\rho|_{G_l}$ is semistable;
- (3) ρ is ramified at only finitely many primes;
- (4) $\det(\rho) = \varepsilon$;
- (5) for every $p \equiv -1 \pmod{l}$ such that $\overline{\rho}|_{I_p}$ is reducible, $\overline{\rho}|_{G_p}$ is reducible too.

Then ρ is modular.

In view of what has been said above, this result implies that all semistable elliptic curves over \mathbb{Q} are modular. Wiles' strategy to prove Theorem 3.2 is to compare systematically all deformations of $\overline{\rho}$ with certain properties when restricted to decomposition groups to those coming from modular forms of a given level. For simplicity, we will now assume that ρ is semistable at all primes, and follow the exposition in [40], with some modifications, anticipating our discussion of [23], [13] and [4].

So suppose that $\overline{\rho}$ is as in Theorem 3.2, and moreover that ρ is semistable at all primes. We will now forget about ρ , for the moment, but keep $\overline{\rho}$. So $\overline{\rho}$ is a continuous representation of $G_{\mathbb{Q}}$ on a 2-dimensional k -vector space, with k a finite extension of \mathbb{F}_l with $l \neq 2$, and has the following properties: it is modular, absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$, semistable at all primes, and $\det(\overline{\rho}) = \overline{\varepsilon}$. As nothing about these hypotheses changes if we replace k by a finite extension of it, we may suppose, for example, that the characteristic polynomials of the $\overline{\rho}(\sigma)$, σ in $G_{\mathbb{Q}}$, are all split. Let O be the ring of integers in a finite extension K of \mathbb{Q}_l , with residue field k . (Later in the proof, we need a modular form of "minimal level" giving rise to $\overline{\rho}$, and with coefficients in O .) For any finite set Σ of primes we define two O -algebras $R_{O,\Sigma}$ and $\mathbb{T}_{O,\Sigma}$, as follows.

DEFINITION 3.3. — Let R be a complete local noetherian O -algebra with residue field k . A deformation of $\overline{\rho}$ to R is a free R -module M of rank two, with a continuous $G_{\mathbb{Q}}$ action, such that $k \otimes_R M$ is isomorphic to $\overline{\rho}$. A deformation ρ is said to be of type Σ if $\det(\rho) = \varepsilon$, and ρ is semistable at l and "minimally ramified" outside Σ :

- (1) if $l \notin \Sigma$ and $\overline{\rho}$ is Barsotti-Tate, then ρ is Barsotti-Tate;

- (2) if $p \notin \Sigma \cup \{l\}$ and $\bar{\rho}$ is unramified at p , then ρ is unramified at p ;
 (3) if $p \notin \Sigma \cup \{l\}$ and $\bar{\rho}$ is ramified at p (and hence semistable, with our hypotheses), then ρ is semistable at p .

With these definitions, there is, for each Σ , a universal deformation ring $R_{O,\Sigma}$ that represents the functor that sends R to the set of isomorphism classes of deformations of type Σ over R . A very good reference for this is [19]. If $K \rightarrow K'$ is a finite extension, then $R_{O',\Sigma} = O' \otimes_O R_{O,\Sigma}$.

Let us now turn to the definition of $\mathbb{T}_{O,\Sigma}$. The reader is referred to Appendix A for certain properties of the Galois representation ρ_f associated to a modular form f of weight two with coefficients in $\overline{\mathbb{Q}}_l$. We define \mathcal{N}_Σ to be the set of weight two newforms f with coefficients in $\overline{\mathbb{Q}}_l$ such that $\rho_f: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(O_f)$ is of type Σ , where O_f is the sub- O -algebra of $\overline{\mathbb{Q}}_l$ generated by the coefficients of f . The results in Appendix A imply that there is an integer N_Σ , such that for each f in \mathcal{N}_Σ , the level of f divides N_Σ . This implies that \mathcal{N}_Σ is a finite set. One can take N_Σ as follows:

$$N_\Sigma := l^\delta \cdot \prod_{p|N(\bar{\rho})} p \cdot \prod_{p \in \Sigma - \{l\}} p^2,$$

where δ is 0 if $\bar{\rho}$ is Barsotti-Tate and l not in Σ , and δ is 1 otherwise, and where $N(\bar{\rho})$ is the level associated to $\bar{\rho}$ by Serre in [48] (i.e., $N(\bar{\rho})$ is given by the usual formula for the Artin conductor of a representation, in terms of the ramification subgroups at all $p \neq l$). The reason that such a N_Σ suffices is that the wild parts of the conductors of ρ_f and $\bar{\rho}_f$ are equal. For simplicity, we will now only consider Σ that do not contain primes dividing $N(\bar{\rho})$ (this suffices for the application to semistable lifts of $\bar{\rho}$). For each f in \mathcal{N}_Σ , we have a morphism $R_{O,\Sigma} \rightarrow O_f$, and we define $\mathbb{T}_{O,\Sigma}$ to be the image of $R_{O,\Sigma}$ in the product of the O_f . Since R_Σ is generated, as O -algebra, by the traces of elements in the universal representation, $\mathbb{T}_{O,\Sigma}$ is generated by the elements $a_p: f \mapsto a_p(f)$ for p not dividing lN_Σ .

The method of Wiles is now to show that the surjections $R_{O,\Sigma} \rightarrow \mathbb{T}_{O,\Sigma}$ are isomorphisms, by studying how they change as Σ varies. The first step in this is to use what has been proved about Serre's conjectures on modularity of mod l representations in [48]: \mathcal{N}_\emptyset is not empty (see [44] and [21]). This implies that we can suppose (and we will) that we have a section $\pi = \pi_\emptyset: \mathbb{T}_\emptyset \rightarrow O$. We let P_Σ denote the corresponding O -valued points of $\mathrm{Spec}(\mathbb{T}_{O,\Sigma})$ and $\mathrm{Spec}(R_{O,\Sigma})$, for each Σ . Wiles introduced the following O -modules associated to each Σ : on the one hand the cotangent spaces at the P_Σ , i.e., $P_\Sigma^* \Omega_{R_{O,\Sigma}/O}^1$ and $P_\Sigma^* \Omega_{\mathbb{T}_{O,\Sigma}/O}^1$, and on the other hand the "module of congruences" $P_\Sigma^* \mathcal{O}_{Z_\Sigma}$, defined as follows. Since $\mathbb{T}_{O,\Sigma}$ is finite free as O -module, $\mathrm{Spec}(\mathbb{Q} \otimes \mathbb{T}_{O,\Sigma})$ is the disjoint union of two open and closed subschemes $P_{\Sigma,K}$ and $Z_{\Sigma,K}$, with $P_{\Sigma,K}$ consisting of the point $P_\Sigma(\mathrm{Spec}(K))$. We let Z_Σ be the scheme theoretic closure of $Z_{\Sigma,K}$ in $\mathrm{Spec}(\mathbb{T}_{O,\Sigma})$ (note that the $\mathbb{T}_{O,\Sigma}$ are reduced by construction). These modules, that will intervene only via their lengths, are usually introduced as $\ker(P_\Sigma^*)/\ker(P_\Sigma^*)^2$ and $P_\Sigma^* \mathrm{Ann}_{\mathbb{T}_{O,\Sigma}}(\ker(P_\Sigma^*))$. (This last module has finite length if and only if $\mathrm{Spec}(\mathbb{Q} \otimes \mathbb{T}_{O,\Sigma})$ is reduced at $P_{\Sigma,K}$.)

The fact that $R_{O,\Sigma}$ represents the functor of isomorphism classes of deformations of type Σ implies the following Galois cohomological description of $P^*\Omega^1_{R_{O,\Sigma}/O}$:

$$\text{Hom}_O(P^*\Omega^1_{R_{O,\Sigma}/O}, K/O) = H^1_\Sigma(G_\mathbb{Q}, \text{ad}^0(\rho) \otimes K/O),$$

where ρ is the representation corresponding to $P = P_\Sigma$, where $\text{ad}^0(\rho)$ is the representation of $G_\mathbb{Q}$ on the sub- O -module of trace zero elements of $\text{End}_O(M_\rho)$, and where $H^1_\Sigma(G_\mathbb{Q}, \text{ad}^0(\rho) \otimes K/O)$ denotes the subgroup of $H^1(G_\mathbb{Q}, \text{ad}^0(\rho) \otimes K/O)$ of classes that map, at all p , to the subgroups $L_{\Sigma,p}$ of the $H^1(G_p, \text{ad}^0(\rho) \otimes K/O)$ that reflect the conditions for deformations to be of type Σ . To be explicit, these $L_{\Sigma,p}$ are:

- $L_{\Sigma,p} = H^1(G_p/I_p, (\text{ad}^0(\rho) \otimes K/O)^{I_p})$ if $p \notin \Sigma \cup \{l\}$;
- $L_{\Sigma,p} = H^1(G_p, \text{ad}^0(\rho) \otimes K/O)$ if $p \in \Sigma$ and $p \neq l$;
- $L_{\Sigma,l}$ is the subspace of $H^1(G_l, \text{ad}^0(\rho) \otimes K/O)$ that corresponds to deformations that are Barsotti-Tate, if $l \notin \Sigma$;
- $L_{\Sigma,l}$ is the subspace of $H^1(G_l, \text{ad}^0(\rho) \otimes K/O)$ that corresponds to deformations that are semistable at l , if $l \in \Sigma$.

The results of Poitou-Tate on local duality and global Euler characteristic show that, for M a finite discrete $G_\mathbb{Q}$ -module, with a Selmer datum $L_v \subset H^1(G_v, M)$ at all places v of \mathbb{Q} , one has:

$$\frac{\#H^1_L(G_\mathbb{Q}, M)}{\#H^1_{L^\perp}(G_\mathbb{Q}, M^*)} = \frac{\#H^0(G_\mathbb{Q}, M)}{\#H^0(G_\mathbb{Q}, M^*)} \cdot \prod_v \frac{\#L_v}{\#H^0(G_v, M)},$$

where M^* is the Cartier dual $\text{Hom}(M, \overline{\mathbb{Q}}^*)$ of M , and where, for each v , L_v^\perp is the orthogonal of L_v . Moreover, if $L \subset L'$ are two Selmer data for M , then one has an exact sequence:

$$0 \rightarrow H^1_L(G_\mathbb{Q}, M) \rightarrow H^1_{L'}(G_\mathbb{Q}, M) \rightarrow \prod_v L'_v/L_v \rightarrow H^1_{L^\perp}(G_\mathbb{Q}, M^*)^\vee \rightarrow H^1_{L'^\perp}(G_\mathbb{Q}, M^*)^\vee \rightarrow 0.$$

Having established this, Wiles first proves that $R_{O,\emptyset} \rightarrow \mathbb{T}_{O,\emptyset}$ is an isomorphism, and then that this remains so if one enlarges Σ . The argument for the first step is really amazing, he somehow manages to “patch”, for a suitable sequence of Σ_n , the R_{O,Σ_n} and the \mathbb{T}_{O,Σ_n} into power series rings, with the same number of generators, and deduce from that that $R_{O,\emptyset} \rightarrow \mathbb{T}_{O,\emptyset}$ is an isomorphism. (This patching argument was introduced in [59], and used to show that $\mathbb{T}_{O,\emptyset}$ is a complete intersection, but Faltings pointed out that one could also use the argument directly in proving $R_{O,\emptyset} \rightarrow \mathbb{T}_{O,\emptyset}$ to be an isomorphism.) We will now take a closer look at this argument, in order to see which conditions have to be satisfied by the type of deformations that one considers for it to work.

So suppose that one wants to do this argument for $R_{O,\Sigma}$. The primes p that one wants to add to Σ are congruent to 1 modulo a high power of l , and such that $\bar{\rho}$ is unramified at p with distinct Frobenius eigenvalues in k^* . For such a p , and for Σ' containing p , $\rho_{O,\Sigma'}^{\text{univ}}|_{G_p}$ is a direct sum of two characters, whose restrictions to I_p factor through the l -part Δ_p of $\text{Gal}(\mathbb{Q}_p^{\text{unr}}(\mu_p)/\mathbb{Q}_p^{\text{unr}}) = (\mathbb{Z}/p\mathbb{Z})^*$. Choosing one of the two Frobenius eigenvalues gives

$R_{O,\Sigma'}$ the structure of an $O[\Delta_p]$ -algebra. The Σ' that one wants to consider are of the form $\Sigma' = \Sigma \cup Q$, with Q a set of r elements, for some fixed r , and such that $R_{O,\Sigma'}$ can also be topologically generated by r elements. (Note that $R_{O,\Sigma'}$ is an algebra over $O[\Delta_Q]$ with $\Delta_Q = \prod_{p \in Q} \Delta_p$, and that $O[\Delta_Q]$ looks more and more as a power series ring in r variables, as the primes p are closer to 1, l -adically.) Let L and L' be the Selmer data corresponding to Σ and Σ' . Then $\dim_k \prod_v L'_v/L_v = r$, so one finds, by the exact sequence above, that $\dim_k H_{L^\perp}^1(G_{\mathbb{Q}}, \text{ad}^0(\bar{\rho})^*) \geq \dim_k (H_L^1(G_{\mathbb{Q}}, \text{ad}^0(\bar{\rho})))$. But in the displayed formula above, one has $\dim_k L_p \geq \dim_k H^1(G_p/I_p, \text{ad}^0(\bar{\rho})^{I_p}) = \dim_k (H^0(G_p, \text{ad}^0(\bar{\rho})))$, for all $p \neq l$, whereas $\dim_k L_\infty = 0$ and $\dim H^0(G_\infty, \text{ad}^0(\bar{\rho})) = 1$. Moreover, in that formula one has $\#H^0(G_{\mathbb{Q}}, M) = 1$ (since $\bar{\rho}$ is absolutely irreducible), and $\#H^0(G_{\mathbb{Q}}, M^*) = 1$ (since the restriction of $\bar{\rho}$ to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible). It follows that this setup can only work if $L_p = H^1(G_p/I_p, \text{ad}^0(\bar{\rho})^{I_p})$ for all $p \neq l$, and $\dim_k(L_l) \leq 1 + \dim_k \text{ad}^0(\bar{\rho})^{G_l}$. This means that Σ must be \emptyset , and that L_l must be of dimension 1, unless $\bar{\rho}|_{I_l} \cong \bar{\varepsilon} \oplus 1$. This last condition puts a very strong restriction on the type of local deformations at l that one can use.

In order to prove that $R_{O,\emptyset} \rightarrow \mathbb{T}_{O,\emptyset}$ is an isomorphism, Taylor and Wiles use that, in their situation, the localization at $\bar{\rho}$ of $H^1(X_0(N_\emptyset)(\mathbb{C}), O)$ is a free $\mathbb{T}_{O,\emptyset}$ -module, and similarly for the Σ 's that they choose. Such results are quite delicate to prove. In the next section we will discuss how Diamond and Fujiwara have gotten around this, and actually obtain such freeness results as a consequence of the method. In [16], the freeness assumption is not used, but the given proof still relies on q -expansions (see [16, Remark 4.15]).

Let us now briefly discuss how Wiles proved that the $R_{O,\Sigma} \rightarrow \mathbb{T}_{O,\Sigma}$ are isomorphisms. This is done by induction on the number of elements of Σ , but, in order to carry out this induction, one actually proves more, namely, that these O -algebras are complete intersections. Indeed, Wiles found a criterion for doing the induction, in terms of the changes of the lengths of $P^*\Omega_{R_{O,\Sigma}/O}^1$ and $P^*\mathcal{O}_{Z_\Sigma}$ when comparing between Σ and $\Sigma' := \Sigma \cup \{p\}$. On the Galois side, the exact sequence above gives an upper bound for the length of $P^*\Omega_{R_{O,\Sigma'}/O}^1$. On the Hecke side, [16, §4.4] gives a new proof of the lower bound for the length of $P^*\mathcal{O}_{Z_{\Sigma'}}$ that was proved by Wiles. This proof does not use freeness, and it nicely relates this change of length to the residue at 2 of the L -function of the symmetric square of the system of representations associated to P , giving a relation to the Bloch-Kato conjectures. Wiles' argument, which is to compute the composite $J_0(N_\Sigma) \rightarrow J_0(N_{\Sigma'}) \rightarrow J_0(N_\Sigma)$, is also sketched in [16, §4.4].

4. IMPROVEMENTS OF THE COMMUTATIVE ALGEBRA PART

The results in commutative algebra that are used in [13] and [4] are improvements of those in [62] and [59]. These improvements were found independently by Diamond [24] and

Fujiwara [32], motivated by Fujiwara's work on modularity over totally real number fields. We also note that Lenstra, de Smit, Rubin, and Schoof have established isomorphism and complete intersection criteria as in Wiles, without the Gorenstein hypothesis, and without the limiting process, see [20]. Let us now state the criteria as in [24, Theorems 2.1 and 2.4].

THEOREM 4.1. — *Let k be a finite field, and $r \geq 0$ an integer. Let $A := k[[S_1, \dots, S_r]]$, $B := k[[X_1, \dots, X_r]]$, let R be a k -algebra, and let H be a non-zero R -module that has finite k -dimension. Suppose that for every $n \geq 1$ one has a commutative diagram:*

$$\begin{array}{ccc} A & \xrightarrow{\phi_n} & B \\ \downarrow & & \downarrow \psi_n \\ k & \longrightarrow & R \end{array}$$

and a B -module H_n with a morphism $\pi_n: H_n \rightarrow H$ such that as an A -module, H_n is free over A/m_A^n , and such that the morphism $k \otimes_A H_n \rightarrow H$ induced by π_n is an isomorphism. Then H is free over R , and R is a (zero dimensional) complete intersection.

In the application of this result, k is as above, A is a projective limit of k -algebras of the form $k[\Delta_Q]$, with Q a set of r distinct primes $p \equiv 1 \pmod{l^n}$ and Δ_Q the product of the $(\mathbb{Z}/p\mathbb{Z})_l^*$, B is a projective limit of $k \otimes_O R_{O,Q}$'s, $R = k \otimes_O R_{O,\emptyset}$, and H and H_n come from (co)homology groups of modular curves. The freeness of H_n over A/m_A^n basically comes from standard facts about cohomology of locally constant sheaves and unramified covers of affine Riemann surfaces. The Hecke algebra $k \otimes_O \mathbb{T}_{O,\emptyset}$ is the image T of R in $\text{End}_k(M)$, so the conclusion that H is free over R implies that $R = T$. The freeness version of Wiles' numerical criterion is as follows.

THEOREM 4.2. — *Let O be a complete discrete valuation ring with finite residue field k , and let R be a complete noetherian local O -algebra. Let H be an R -module, finite free over O , let $\phi: R \rightarrow T$ be the quotient by $\text{Ann}_R(H)$, and suppose that T has a section $\pi_T: T \rightarrow O$. Put $\pi_R := \phi\pi_T$. Define $\Omega := H/(H[\ker(\pi_T)] + H[\text{Ann}_T(\ker(\pi_T))])$. Let d be the O -rank of $H[\ker(\pi_R)]$. If Ω has finite length over O , then the following are equivalent:*

- (1) $\text{rank}_O H \leq d \cdot \text{rank}_O T$ and $\text{length}_O \Omega \geq d \cdot \text{length}_O (\ker(\pi_R) / \ker(\pi_R)^2)$;
- (2) $\text{rank}_O H = d \cdot \text{rank}_O T$ and $\Omega \cong (O / \text{Fitt}_O(\ker(\pi_R) / \ker(\pi_R)^2))^d$;
- (3) R is a complete intersection and H is free (of rank d) over R .

5. THE WORK OF BREUIL, CONRAD, DIAMOND, AND TAYLOR

We are now ready to discuss the work of the four authors mentioned above in [23], [13], and [4]. Before getting into any details, let us see what problems were solved in each of these three articles. In [23, Theorem 5.3], Diamond gets rid of condition (5) in Theorem 3.2. To be precise, let us state his result.

THEOREM 5.1 (Diamond). — *Let $l > 2$ be prime, K a finite extension of \mathbb{Q}_l , O its ring of integers, k its residue field, and $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(O)$ an odd continuous representation such that:*

- (1) *its reduction $\bar{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(k)$ is modular, and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible;*
- (2) *$\rho|_{G_l}$ is Barsotti-Tate and $\det(\rho)|_{I_l} = \varepsilon|_{I_l}$, or, with respect to a suitable basis, $\rho|_{G_l}$ is of the form $\begin{pmatrix} \phi & * \\ 0 & \psi \end{pmatrix}$, with ψ unramified, $\bar{\psi} \neq \bar{\phi}$, and $\phi|_{I_l} = \chi\varepsilon^{k-1}|_{I_l}$ for some integer $k \geq 2$ and χ of finite order;*
- (3) *ρ is ramified at only finitely many primes.*

Then ρ is modular.

We should note here that Theorem 3.2 is weaker than the result that was proved by Wiles. What is proved in [62] is the theorem above, with the extra condition (5) of Theorem 3.2. Let us now explain what the problem is in a case that does not satisfy this condition (5).

So suppose that $\bar{\rho}$ satisfies the conditions of Theorem 5.1, that $p \neq l$, that $\bar{\rho}_p = \bar{\rho}|_{G_p}$ is irreducible, but $\bar{\rho}|_{I_p}$ is reducible. Then $\bar{\rho}_p$ is of the form $\mathrm{Ind}_{\mathbb{Q}_{p^2}}^{\mathbb{Q}_p} \psi$, with \mathbb{Q}_{p^2} the unramified extension of degree two of \mathbb{Q}_p , and $\psi: G_{\mathbb{Q}_{p^2}} \rightarrow k^*$ a continuous character such that $\psi^\sigma \neq \psi$. (To prove this, note that $\bar{\rho}(I_p)$ must have exactly two fixed points on $\mathbb{P}^1(\mathbb{F}_p)$, that are interchanged by Frob_p , since otherwise $\bar{\rho}_p$ would be reducible.) But then, if l divides $p+1$, there are nontrivial deformations of $\bar{\rho}_p$ of the form $\rho_p = \mathrm{Ind}_{\mathbb{Q}_{p^2}}^{\mathbb{Q}_p}(\psi\mu)$ from G_p to $\mathrm{GL}_2(O')$, with O' a finite extension of O , $\psi: G_p \rightarrow O'^*$ the Teichmüller lift of $\psi: G_p \rightarrow k^*$, and with $\mu: G_{\mathbb{Q}_{p^2}} \rightarrow O'^*$ of order l . One checks that $\det(\rho)|_{I_p}$ is the Teichmüller lift of $\det(\bar{\rho})|_{I_p}$ (use that Frob_p acts on the tame inertia group $I_p^{\mathrm{tame}} = \hat{\mathbb{Z}}^{(p)}(1)$ by multiplication by p). The whole problem arises from the fact that, on the one hand, $\bar{\rho}_p$ and ρ_p have the same Artin conductor, namely, the square of the conductor of ψ , but that, on the other hand, $\bar{\rho}_p$ admits different lifts with this conductor. This means that if we consider lifts of $\bar{\rho}$ to be minimally ramified at p if they have Artin conductor $\mathrm{cond}(\psi)^2$ at p , then we get an $L_p \subset H^1(G_p, \mathrm{ad}^0(\bar{\rho}))$, in the notation of Section 3, that is nonzero, making already Wiles' method at the minimal level impossible.

The conclusion is that, on the automorphic side, levels of newforms are not fine enough invariants to work with; one should impose finer conditions on the restrictions to the G_p (at the Galois side), and corresponding conditions on the irreducible admissible representations on the other side. Wiles already notes this in the second remark following Conjecture 2.16 in [62].

The finer conditions that will be imposed are in terms of what are called “types” and “extended types” in the articles that we discuss here. An extended type, at a prime p ($p = l$ is allowed) is simply an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}_l}$ of the Weil-Deligne group W'_p of \mathbb{Q}_p (see Appendix A), and then types are isomorphism

classes of restrictions to I_p of extended types. The local Langlands correspondence makes extended types correspond to isomorphism classes of infinite dimensional irreducible admissible representations of $\mathrm{GL}_2(\mathbb{Q}_p)$, over $\overline{\mathbb{Q}_l}$. We will not discuss the proof of Theorem 5.1 here, as it is repeated in [13] and [4], with some changes, however. Diamond used, in order to simplify the representation theory at the automorphic side, the Jacquet-Langlands correspondence to work with a quaternion algebra over \mathbb{Q}_p instead of the matrix algebra. In the two subsequent articles, one works directly with modular curves. We recommend [22] for an overview of [23], that does not become too technical. But, as the reader can already guess, the rest of this section will get more technical, especially on the automorphic side, because of these finer restrictions.

With Theorem 5.1 above, it is not hard to prove that all elliptic curves over \mathbb{Q} that are semistable at 3 and 5 are modular; we refer to [23, §5] for details. So the only remaining problem to get modularity for all elliptic curves over \mathbb{Q} is to get rid of the semistability conditions at 3 and 5. Since modularity is invariant under twisting, Diamond's result actually implies that the only elliptic curves E that remain to be dealt with are those that have potentially good reduction at 3 and 5, but that do not have a twist with good reduction at 3 and 5. Since one knows that the two j -invariants that correspond to E with more than two automorphisms are modular, the only twists to consider are quadratic twists.

The first step in the direction of relaxing the conditions at 3 and 5 was made in [13], where it is proved that E is modular if it acquires good reduction over a tame extension of \mathbb{Q}_3 . The main new ingredient of this paper, compared to [23], is a new type of deformation problem, for a mod l representation of G_l . Roughly speaking, one considers deformations ρ over R of $\bar{\rho}$ over k such that the restriction of ρ to G_F , or a twist of it by a fixed quadratic character, is Barsotti-Tate, for F a fixed finite extension of \mathbb{Q}_l with ramification index $e \leq l-1$. We have seen, in Section 3, that it is crucial that the tangent space over k of the universal deformation ring of the type of deformations of $\bar{\rho}|_{G_l}$ that one considers be of dimension at most one. This crucial result for [13] was proved by Conrad in [12], using his description of finite free group schemes over the rings of integers of such F obtained in [11], generalizing earlier work of Fontaine in [29]. Using Conrad's result, it was then proved ([13, Theorem 7.1.2]) that any elliptic curve E over \mathbb{Q} that acquires good reduction over a tame extension of \mathbb{Q}_3 is modular.

The final step in relaxing the conditions at 3 is done in [4]. It is the work of Breuil [3], summarized in [2], that gives a workable enough description of certain finite free group schemes over rings of integers of arbitrary finite extensions F of \mathbb{Q}_l , that makes this possible. With this tool available, it is then proved that the remaining E , i.e., those that acquire good reduction only after a wild extension of \mathbb{Q}_3 , are modular. The article [4] consists for about 70% (of 77 pages) of the proof that, in the cases that are needed, the local deformation space at l has dimension at most one.

In order to keep this section of reasonable length, we postpone the discussion of these results at l to the next one, and in this one we focus more on the global aspects of the proof, and on Conjecture 1.3.1 of [4], which says for which $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(O)$ one hopes to be able to prove modularity.

We will now follow [4, §1], in order to introduce the necessary terminology, and to state the conjecture just mentioned. We suppose $l > 2$. An extended l -type is defined as an isomorphism class of two-dimensional representations of W'_l over $\overline{\mathbb{Q}}_l$ (with open kernel), and l -types are isomorphism classes of restrictions to I_l of extended l -types.

Suppose that τ' is an extended l -type, with restriction τ to I_l .

DEFINITION 5.2. — *A continuous representation $\rho: G_l \rightarrow \mathrm{GL}_2(O)$ (with O the ring of integers in a finite extension K of \mathbb{Q}_l) is said to be of extended type τ' (resp. of type τ) if:*

- (1) ρ is potentially Barsotti-Tate;
- (2) $\mathrm{WD}(\rho)$ (as in Appendix B) is in τ' (resp. $\mathrm{WD}(\rho)|_{I_l}$ is in τ);
- (3) the character $\varepsilon^{-1} \det(\rho)$ has finite order prime to l .

Now fix a finite extension K of \mathbb{Q}_l , let O be its ring of integers and k be its residue field. Let $\bar{\rho}$ be a two-dimensional continuous representation of G_l over k , say on a vector space V , such that $\mathrm{End}_{G_l}(V) = k$ (i.e., either $\bar{\rho}$ is absolutely irreducible, or it is a non-split extension of a character by another character). Under this last hypothesis, we have a universal deformation ring R^l_O representing the functor of deformations of $\bar{\rho}$ to complete noetherian local O -algebras with residue field k . (The superscript l is there to indicate that we are considering representations of G_l .) We remark that extended types will only be considered if their restriction to I_l is irreducible. Now consider $\mathrm{Spec}(R^l_O)$. In it, we have a minimal closed subset that contains all deformations of $\bar{\rho}$ to finite extensions O' of O that are of type τ , and similarly for extended type τ' . These minimal closed subsets correspond to (reduced) quotients $R^l_{O,\tau}$ and $R^l_{O,\tau'}$. A deformation ρ over R of $\bar{\rho}$ is said to be *weakly of type τ* (resp. *weakly of extended type τ'*), if the corresponding morphism $R^l_O \rightarrow R$ factors through $R^l_{O,\tau}$ (resp. through $R^l_{O,\tau'}$).

DEFINITION 5.3. — *A type τ (resp. an extended type τ') is weakly acceptable for $\bar{\rho}$ if there exists a surjection of O -algebras $O[[X]] \rightarrow R^l_{O,\tau}$ (resp. $O[[X]] \rightarrow R^l_{O,\tau'}$). A type τ (resp. an extended type τ') is acceptable for $\bar{\rho}$ if moreover $R^l_{O,\tau} \neq 0$ (resp. $R^l_{O,\tau'} \neq 0$), i.e., if there exists at least one l -adic deformation of type τ (resp. of extended type τ'). We will also speak of $\bar{\rho}$ accepting τ (resp. τ').*

Of course, with these definitions, it is very hard to check whether a given $\bar{\rho}$ accepts a given τ or τ' . It is precisely this kind of verifications that occupy the most of [4], and it is there that crucial use is made of Conrad's and Breuil's results on finite group schemes. We note that [4] conjectures that an l -adic lift of $\bar{\rho}$ is of type τ (resp. extended

type τ') if and only if it is weakly of that kind (Conjecture 1.1.1 of [4]), but this has no importance for what follows. What is much more important, is that [4, Conjecture 1.3.1] tries to predict acceptability purely in computable, representation theoretic terms. In order to state this conjecture, [4] needs about 4 pages of preparation, consisting mostly of definitions. Instead of trying to state all these definitions, we will try to see where they come from.

The question one should ask oneself is: if f is a newform over $\overline{\mathbb{Q}}_l$, then what can one say about $\overline{\rho}_{f,l}: G_l \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_l)$ in terms of $\pi_{f,l}$, assuming $\overline{\rho}_f$ irreducible? In particular, for a given $\overline{\rho}$, what are the irreducible admissible representations that occur as $\pi_{f,l}$, for f with $\overline{\rho}_{f,l} \cong \overline{\rho}$?

An answer to this question will then say for which τ and τ' there does exist an l -adic lift of $\overline{\rho}$ of that type. Moreover, from the mechanism that is used to find this, one may guess under what conditions one expects $R_{O,\tau}^l$ or $R_{O,\tau'}^l$ to be topologically generated by one element.

To find the answer to the question (and for other reasons as well), [4] constructs certain l -adic sheaves on certain modular curves, that pick out a non-zero part of exactly those π_f such that $\pi_{f,l}$ has a prescribed type (or extended type). For each τ and for each τ' , with $\tau'|_{I_l}$ irreducible, one defines open subgroups U_τ of $\mathrm{GL}_2(\mathbb{Z}_l)$ and $U_{\tau'}$ of $\mathrm{GL}_2(\mathbb{Q}_l)$, and irreducible representations σ_τ and $\sigma_{\tau'}$ on finite dimensional $\overline{\mathbb{Q}}_l$ -vector spaces, with open kernel. The choice of these subgroups and representations is justified by [4, Lemma 1.2.1]: for every irreducible admissible representation π of $\mathrm{GL}_2(\mathbb{Q}_l)$ over $\overline{\mathbb{Q}}_l$ one has

- $\mathrm{Hom}_{U_\tau}(\sigma_\tau, \pi) = \overline{\mathbb{Q}}_l$ if $\mathrm{WD}(\pi)|_{I_l} \cong \tau$, and $\mathrm{Hom}_{U_\tau}(\sigma_\tau, \pi) = 0$ if $\mathrm{WD}(\pi)|_{I_l} \not\cong \tau$;
- $\mathrm{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = \overline{\mathbb{Q}}_l$ if $\mathrm{WD}(\pi) \cong \tau'$, and $\mathrm{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = 0$ if $\mathrm{WD}(\pi) \not\cong \tau'$.

The fact that such subgroups and representations exist is not particular to our situation. There is a general theory, called (no surprise) type theory, whose goal it is to describe smooth representations of p -adic groups in terms of their restrictions to compact open subgroups; see [5]. Before we go on, let us mention that Khare has also asked and answered the question above, at least in the case of types, in [33].

With these (U_τ, σ_τ) and $(U_{\tau'}, \sigma_{\tau'})$ one constructs sheaves on modular curves as follows. One defines U_l to be U_τ if one considers a type, and $U_{\tau'} \cap \mathrm{GL}_2(\mathbb{Z}_l)$ if one considers an extended type. In each case, one has a representation σ_l of U_l , namely, σ_τ and the restriction of $\sigma_{\tau'}$. Let $U^{(l)}$ be a sufficiently small open subgroup of $\mathrm{GL}_2(\hat{\mathbb{Z}}^{(l)})$, and let σ be the representation of $U := U_l U^{(l)}$ given by σ_l . Then the modular curve and the sheaf are:

$$Y_U := \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}) / UC^*, \quad \mathcal{F}_\sigma := \mathrm{GL}_2(\mathbb{Q}) \backslash (\mathrm{GL}_2(\mathbb{A}) \times M_\sigma^\vee) / UC^*,$$

with M_σ the $\overline{\mathbb{Q}}_l[U]$ -module given by σ . In the case where one considers an extended type, one also gets an automorphism w_l of the pair $(Y_U, \mathcal{F}_\sigma)$. By construction, the duals of the two-dimensional Galois representations ρ that occur in $H_1^1(Y_U, \mathcal{F}_\sigma)$ if one considers a type (resp. in $H_1^1(Y_U, \mathcal{F}_\sigma)^{w_l=1}$ if one considers an extended type) for some $U^{(l)}$ are precisely

those that correspond to newforms f such that $\pi_{f,l}$ are of the prescribed kind that is described by τ or τ' . Let $U' := \mathrm{GL}_2(\mathbb{Z}_l)U^{(l)}$, and consider the morphism $\pi: Y_U \rightarrow Y_{U'}$. Then one has $H_1^1(Y_U, \mathcal{F}_\sigma) = H_1^1(Y_{U'}, \pi_*\mathcal{F}_\sigma)$. In order to get information on the corresponding $\bar{\rho}_l$'s, one now reduces the sheaf modulo l , i.e., one chooses a $\overline{\mathbb{Z}}_l$ -lattice for σ , and one reduces modulo the maximal ideal. By construction, the Jordan-Hölder constituents of this reduction are of the form $\mathcal{F}_{n,m} := \mathrm{Sym}^n(\overline{\mathcal{F}}) \otimes \det(\overline{\mathcal{F}})^{\otimes m}$, with $0 \leq n < l$ and m in $\mathbb{Z}/(l-1)\mathbb{Z}$, and with $\overline{\mathcal{F}}$ the locally constant sheaf of two-dimensional $\overline{\mathbb{F}}_l$ -vector spaces given by the standard representation of $\mathrm{GL}_2(\overline{\mathbb{F}}_l)$, or, if one wishes, by the dual of the l -torsion of the universal elliptic curve. Moreover, the (n, m) that occur, and their multiplicities, can be computed by representation theory (Brauer characters). But now one can use the results of Deligne and Fontaine stated in [28, Theorems 2.5 and 2.6] that describe the $\bar{\rho}_{f,l}$ for newforms of prime to l level, and weight between 2 and $l+1$, in order to see which $\bar{\rho}$ occur in the $H_1^1(Y_{U'}, \mathcal{F}_{n,0})$. Since $\det(\overline{\mathcal{F}})$ is simply $\overline{\mathbb{F}}_l(\varepsilon)$, $H_1^1(Y_{U'}, \mathcal{F}_{n,m})$ is just $H_1^1(Y_{U'}, \mathcal{F}_{n,0})(\varepsilon^{-m})$. Let us note that in the case of an extended type, one also has to deal with the automorphism w_l ; this is done in [4, §1.4]. The role played by the $\mathcal{F}_{n,m}$ explains that the dependence upon $\bar{\rho}$ of “ τ (resp. τ') admits $\bar{\rho}$ ” is via its properties that determine the weight that Serre has attached to $\bar{\rho}|_{I_l}$ (see [28, Sections 2–4]).

Having seen this, we can understand what goes on behind the definition of the notion of “ τ (resp. τ') admits $\bar{\rho}$ ” in [4, §1.3]: this means that there exist newforms f of that type such that ρ_f is a lift of $\bar{\rho}$ of the required type. What is harder to understand, is what is behind the corresponding two notions “simply admits”, because [4] defines this by simply listing all elements of this relation. A reasonable guess seems that this condition means that $\overline{\pi_*\mathcal{F}_\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\bar{\rho}^\vee$. (In fact, the freeness results in Theorems 4.1 and 4.2 and the definition of the Hecke modules that are used imply that in the situations where these theorems can be applied, $\overline{\pi_*\mathcal{F}_\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\bar{\rho}^\vee$.) It would be interesting to know how much of the relation between $\pi_{f,l}$ and $\bar{\rho}_{f,l}$ can be computed using Fontaine’s functors. Let us now state this Conjecture 1.3.1, and the two main theorems (1.4.1 and 1.4.2) of [4].

CONJECTURE 5.4. — *Let k be a finite subfield of $\overline{\mathbb{F}}_l$, $\bar{\rho}: G_l \rightarrow \mathrm{GL}_2(k)$ a continuous representation, τ an l -type and τ' an extended l -type with irreducible restriction to I_l . Suppose that the centraliser of the image of $\bar{\rho}$ is k and that the image of τ is not contained in the center of $\mathrm{GL}_2(\overline{\mathbb{Q}}_l)$.*

- (1) τ (resp. τ') admits $\bar{\rho}$ if and only if $R_{O,\tau}^l \neq \{0\}$ (resp. $R_{O,\tau'}^l \neq \{0\}$), i.e., if and only if there is a finite extension K' of \mathbb{Q}_l in $\overline{\mathbb{Q}}_l$ and a continuous representation $\rho: G_l \rightarrow \mathrm{GL}_2(O_{K'})$ which reduces to $\bar{\rho}$ and has type τ (resp. has extended type τ').
- (2) τ (resp. τ') simply admits $\bar{\rho}$ if and only if τ (resp. τ') is acceptable for $\bar{\rho}$.

THEOREM 5.5. — *Let $l > 2$ be prime, K a finite extension of \mathbb{Q}_l in $\overline{\mathbb{Q}_l}$ and k its residue field. Let $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(K)$ be an odd continuous representation, ramified at only finitely many primes. Assume that its reduction $\bar{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(k)$ is absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$, and is modular. Further, suppose that:*

- $\bar{\rho}|_{G_l}$ has centraliser k ;
- $\rho|_{G_l}$ is potentially Barsotti-Tate with l -type τ (resp. with extended l -type τ');
- τ (resp. τ') admits $\bar{\rho}$;
- τ (resp. τ') is weakly acceptable for $\bar{\rho}$.

Then ρ is modular.

The proof of this theorem is very parallel to the proof of [13, Theorem 7.1.1], and is just written in terms of the changes to make. The strategy is of course the same as Wiles', especially in the way that we have described it, but now one imposes, at all primes where this is required (i.e., l and the so-called vexing primes of [22]), these finer restrictions to define the right notion of minimally ramified deformations. The commutative algebra that is used consists of the results that we have described in Section 4. The required Hecke modules are constructed as cohomology groups of sheaves on modular curves just as the \mathcal{F}_σ above, with the difference that one will also have types τ_p at some primes $p \neq l$.

Of course, in order to apply this theorem, one has to prove that the last condition holds, i.e., that there exists a surjection $O[[X]] \rightarrow R_{O,\bar{\rho}}^l$. This condition has indeed been proved in sufficiently many cases in order to prove Theorem 1.4, by Conrad in [12] for tamely ramified types with small image, and in [4] for some more types, using Breuil's work. We will come back to this question of proving weak acceptability in the next section.

Let us now see what is still required in order to prove the theorem that all elliptic curves E over \mathbb{Q} are modular. The proof of this is divided into three cases:

- (1) $\bar{\rho}_{E,5}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is irreducible;
- (2) $\bar{\rho}_{E,5}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is reducible, but $\bar{\rho}_{E,3}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{-3}))}$ is absolutely irreducible;
- (3) the remaining cases.

First of all, the last case corresponds to rational points on a few modular curves of small level. It is proved in [13, Lemma 7.2.3], with help of Elkies, that the set of all such elliptic curves has, up to isogeny and twist, just three elements, which are known to be modular by calculations. Let us consider the second case. Then E acquires semistable reduction over a tame extension of \mathbb{Q}_3 because $\bar{\rho}_{E,5}(I_3)$ has order dividing $(5-1)^2 5$. If a quadratic twist E' of E is semistable at 3, one switches to E' , and one is in the situation considered in [23]. If not, then any ramified quadratic twist of E_K with K a ramified quadratic extension of \mathbb{Q}_3 has good reduction, so that one can use [12, Theorem 4.2.1]. Let us finally consider the first case. In this case, Theorem 1.4 says that $\bar{\rho}_{E,5}$ is modular. Moreover, since $5 > 3$, E acquires semi-stable reduction over a tame extension of \mathbb{Q}_5 of degree dividing 4 or dividing 6; in the first case, where E is potentially ordinary at 5,

Theorem 5.3 of [23] applies, in the second case, there is a ramified extension K of \mathbb{Q}_5 of degree 3, such that a ramified quadratic twist of E_K has good, supersingular reduction, and [12, Theorem 4.2.1] applies.

So it remains to explain how Theorem 1.4 is proved. Let $\bar{\rho}$ be as in that theorem. One first twists $\bar{\rho}$ by a suitable quadratic character, such that $\bar{\rho}|_{G_3}$ falls into one of the 6 cases of [4], page 3, whose Artin conductors at 3 are 3^i , $0 \leq i \leq 5$. Then one considers elliptic curves E over \mathbb{Q} such that $\bar{\rho}_{E,5}$ is isomorphic to $\bar{\rho}$. The moduli space of these is the union of two non-empty open subschemes of $\mathbb{P}_{\mathbb{Q}}^1$, hence there are plenty of such E . Using Hilbert irreducibility, and some computations by Manoharmayum [36], one can show that there exists such an E such that $\bar{\rho}_{E,3}$ is surjective on $\mathrm{GL}_2(\mathbb{F}_3)$, and such that, in the cases of conductor 3^i with $i \geq 3$, $\rho_{E,3}$ is such that Theorem 5.5 above can be applied to it, i.e., such that the type, or extended type, can be proved to be weakly acceptable for $\bar{\rho}_{E,3}$. These results are proved in [4, §2.1]. The use of an extended type is required only in the case of conductor 3^5 .

6. DEFORMATION PROBLEMS AT l

Let us now discuss the results concerning weak acceptability, obtained in [12], [13], and in Sections 4–9 of [4]. We recall what this means. Let $l > 2$ be prime, K a finite extension of \mathbb{Q}_l , with ring of integers O and residue field k . Let $\bar{\rho}$ be a two-dimensional representation of G_l over k , with centraliser k . Let τ be an l -type, and τ' an extended l -type with irreducible restriction to I_l . Then the quotients $R_{O,\tau}^l$ and $R_{O,\tau'}^l$ of the universal deformation ring R_O^l of $\bar{\rho}$ have been defined in the previous section, by taking the Zariski closures in $\mathrm{Spec}(R_O^l)$ of the sets of l -adic lifts of type τ (resp. extended type τ'). Weak admissibility of τ (resp. τ') just means that there exists a surjection $O[[X]] \rightarrow R_{O,\tau}^l$ (resp. $O[[X]] \rightarrow R_{O,\tau'}^l$). So what one wants to compute is the dimension over k of the space of deformations over $k[t]/(t^2)$ of $\bar{\rho}$ that are weakly of type τ (or extended type τ'). But the way that these kind of deformations have been defined makes this impossible. So, in order to deal with this problem, one defines other deformation problems, whose universal deformation rings surject to the ones above, and for which one then proves that they admit a surjection from $O[[X]]$. The aim of this section is just to describe these new deformation problems, and to sketch the tools that are used in their study. Let us mention that there seems to be some hope that one can deal directly with the rings $R_{O,\tau}^l$ and $R_{O,\tau'}^l$ (see the beginning of [4, §4]). In what follows, we drop the superscript l from the notation, as we are only considering representations of G_l .

We follow [4, §4]. We first discuss some generalities, and then what happens in the worst case, i.e., the conductor 3^5 case. Let F be a finite Galois extension of \mathbb{Q}_l , with group Γ . We let R be the ring of integers in F , and k its residue field (if we need to refer to the residue field of O , we will call it k_O). For \mathcal{G} a finite free group scheme over R ,

of l -power order, we let $\mathbb{D}(\mathcal{G})$ denote the contravariant Dieudonné module of \mathcal{G}_k ; it is a $W(k)$ -module, equipped with operators \mathbf{F} and \mathbf{V} , such that $\mathbf{FV} = \mathbf{VF} = l$ and, for all x in $W(k)$: $\mathbf{F}x = \text{Frob}_l(x)\mathbf{F}$. A *descent datum* for a finite free group scheme \mathcal{G} over R is a right action of Γ on \mathcal{G} , compatible with its action on $\text{Spec}(R)$, i.e., for each γ in Γ , one has a commutative diagram:

$$\begin{array}{ccc} \mathcal{G} & \xrightarrow{[\gamma]} & \mathcal{G} \\ \downarrow & \sim & \downarrow \\ \text{Spec}(R) & \xrightarrow[\text{Spec}(\gamma)]{\sim} & \text{Spec}(R) \end{array}$$

such that $[\gamma_1\gamma_2] = [\gamma_2][\gamma_1]$ for all γ_1 and γ_2 in Γ . Note that, since $\mathbb{Z}_l \rightarrow R$ may be ramified, this is not what one should call a descent datum; however, it is a descent datum after restriction to F . In particular, we can associate in this way, to a pair $(\mathcal{G}, [\cdot])$, a group scheme over \mathbb{Q}_l . A descent datum as above gives an action of Γ on $\mathbb{D}(\mathcal{G})$, compatible with the action of Γ on $W(k)$, and commuting with \mathbf{F} and \mathbf{V} , i.e., it becomes a module over the ring $W(k)[\mathbf{F}, \mathbf{V}][\Gamma]$ (with suitable commutation relations). The idea is now that to τ and τ' , one can associate ideals I and I' of this ring, that will impose the right conditions on l -adic lifts of $\bar{\rho}$ to be of type τ or of extended type τ' .

Let us now describe the kind of deformation problems that are considered. The extension F of \mathbb{Q}_l should be taken such that the type τ (or the extended type τ') becomes unramified over it. Then one fixes a *model* over R for $\bar{\rho}$, i.e., a pair $(\mathcal{G}_0, [\cdot])$ as above, giving $\bar{\rho}$ as the module of its $\overline{\mathbb{Q}_l}$ -points, and such that I (or I') annihilates $\mathbb{D}(\mathcal{G}_0)$. Once a model is chosen, one can consider all deformations ρ of $\bar{\rho}$, say to artinian rings, that admit a model $(\mathcal{G}, [\cdot])$ with $\mathbb{D}(\mathcal{G})$ killed by I (or by I'), with a filtration in which each successive quotient is isomorphic to $(\mathcal{G}_0, [\cdot])$. A nice condition to impose is then that such models are unique, and indeed, this can be realized in the situations that are needed (this is what [4, §4.2] is about). Let us denote by $R_{O,I}$ and $R_{O,I'}$ the universal deformation rings thus obtained.

This is where the generalities end, and where one considers each of the 3 cases (conductor 3^3 , 3^4 and 3^5) separately. Let us just describe what happens in the worst case: conductor 3^5 . In that case, Conjecture 1.3.1 of [4] suggests 3 extended types τ'_i , $i \in \mathbb{Z}/3\mathbb{Z}$. One can take $O = \mathbb{Z}_3$. The restrictions of the τ'_i to $W_{\mathbb{Q}_3(\sqrt{-3})}$ are given by a morphism

$$\mathbb{Q}_3(\sqrt{-3})^* \longrightarrow \mathbb{Q}_3(\zeta)^*, \quad \left\{ \begin{array}{ll} \sqrt{-3} & \mapsto \zeta - \zeta^{-1} \\ -1 & \mapsto -1 \\ 4 & \mapsto 1 \\ 1 + 3\sqrt{-3} & \mapsto \zeta \\ 1 + \sqrt{-3} & \mapsto \zeta^i \end{array} \right.$$

with ζ of order 3. This defines abelian extensions F_i of $\mathbb{Q}_3(\sqrt{-3})$ of degree 12, that are Galois over \mathbb{Q}_3 . The ideals I_i that one uses are all three generated by: $\mathbf{F} + \mathbf{V}$, $[\gamma_4^2] + 1$, and $([\gamma_3] - [\gamma_3^{-1}])[\gamma_2] - \mathbf{F}$, with γ_2, γ_3 and γ_4 certain elements of $\text{Gal}(F_i/\mathbb{Q}_3)$. In particular, γ_2 is not in the inertia subgroup of $\text{Gal}(F_i/\mathbb{Q}_3)$, hence the last generator of the I_i reflects that one works with an extended type. Without this condition, or even with a similar condition coming from other possible extended types, one wouldn't expect the tangent space of the deformation problem to be of dimension one.

Theorem 4.6.1 of [4] says that, for each i , there exist four models of $\bar{\rho}$ with the property that models for deformations as above are unique, and, moreover, that each deformation of $\bar{\rho}$ as above has a filtration with all successive quotients isomorphic to one of these four. Theorem 4.6.2 of [4] then says that all four have the property that the universal deformation ring is topologically generated by one element, but three of them have universal deformation ring $\mathbb{F}_3[[X]]$ ([4, Theorem 4.6.3]). Using this, it is finally proved that every 3-adic lift of $\bar{\rho}$ that is of extended type τ'_i , say over a finite extension O of \mathbb{Z}_3 , comes from a morphism $R_O \rightarrow R_{O, \tau'_i}$ that factors through the universal deformation ring R_{O, τ'_i} associated to this last model. Hence, finally, this proves that each of the τ'_i weakly accepts $\bar{\rho}$.

To finish this section, let us point out that proving the theorems in the preceding paragraph takes about 45 pages in [4], with about 30 of them filled with computations with Breuil's ϕ_1 -modules. It is my hope that these notes will encourage readers to take a look at those pages, and understand what is going on. In order to say at least something about these modules, let us give the definition of the l -torsion ϕ_1 -modules over R , the category of which is anti-equivalent to that of finite free group schemes over R that are killed by l .

Let R etc. be as in the second paragraph of this section. Let π be a uniformizer of R , and $E_\pi(u) = u^e - lG_\pi(u)$ its minimal polynomial over $W(k)$. Let ϕ denote the l -th power map on $k[u]/(u^{el})$. An l -torsion ϕ_1 -module is then a triple (M, M_1, ϕ_1) with M a finite free $k[u]/(u^{el})$ -module, with M_1 a $k[u]/(u^{el})$ -submodule containing $u^e M$, and with $\phi_1: M_1 \rightarrow M$ ϕ -semilinear, such that $\phi_1(M_1)$ generates M as $k[u]/(u^{el})$ -module.

We note that the category just described does not depend on the choice of the uniformizer π , but that the functors giving the anti-equivalence mentioned above do depend on that choice. This fact causes a lot of trouble in [4], as one has to study the action of $\text{Gal}(F/\mathbb{Q})$ on models \mathcal{G} over $\text{Spec}(R)$, and, of course, $\text{Gal}(F/\mathbb{Q})$ will not fix the choice of π .

7. TWO RELATED RESULTS

The aim of this section is to briefly state two modularity results on two-dimensional Galois representations that are obtained by others than those mentioned in the title. They can be found in [50] and [6].

THEOREM 7.1 (Skinner, Wiles). — *Let $l > 2$ be prime, and let $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$ be an odd continuous representation, ramified at only finitely many primes, with $\det(\rho) = \psi\varepsilon^{k-1}$ with ψ of finite order, $\varepsilon: G_{\mathbb{Q}} \rightarrow \mathbb{Z}_l^*$ the cyclotomic character and $k \geq 2$ an integer. Suppose that $\rho|_{G_l}$ is the extension of a character ψ_2 by a character ψ_1 with $\psi_2|_{I_l}$ of finite order and such that $\overline{\psi_1}|_{G_l} \neq \overline{\psi_2}|_{G_l}$. If $\overline{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_l)$ is irreducible, then suppose that $\overline{\rho}$ is modular. Then ρ is modular.*

In fact, they even prove a stronger result, with \mathbb{Q} replaced by an arbitrary totally real field. This generality is actually even necessary for the theorem above, since the proof involves changing the field \mathbb{Q} . Let us note that the representations ρ considered by Skinner and Wiles are very different from those considered by Breuil, Conrad, Diamond, and Taylor, which mostly have irreducible restrictions to any open subgroup of G_l .

THEOREM 7.2 (Buzzard, Dickinson, Shepherd-Barron, Taylor). — *Let ρ be a continuous, irreducible, odd representation from $G_{\mathbb{Q}}$ to $\mathrm{GL}_2(\mathbb{C})$ with unsolvable image. Suppose that ρ is unramified at 2 and at 5, and that the image of $\rho(\mathrm{Frob}_2)$ in $\mathrm{PGL}_2(\mathbb{C})$ has order 3. Then ρ is modular.*

Let us note that all continuous representations $\rho: G_F \rightarrow \mathrm{GL}_2(\mathbb{C})$ with solvable image and with F a number field are known to be associated to automorphic representations, by Hecke (in the dihedral case) and Langlands and Tunnell ([35], [60]). The strategy of the proof is explained in [56], and carried out in [7], [53], and [27]. The paper [6] mainly pulls everything together, and provides slight technical but needed improvements of previously obtained results.

In a nutshell, the strategy consists in realizing a suitable twist of ρ over a number field in \mathbb{C} , such that it has a reduction $\overline{\rho} \bmod 2$ with values in $\mathrm{GL}_2(\mathbb{F}_4)$. Then one shows that $\overline{\rho}$ is modular, that ρ arises from an overconvergent 2-adic modular form, and finally that ρ arises from a weight one form.

8. LATEST NEWS

This section has been added at the time the final version of this text was written (June 2000). Its aim is just to direct the reader to some developments that took place after the lecture (March).

Taylor has released two preprints [57] and [58]. In the first one, he proves, using work of Skinner and Wiles ([50], [51], [52]), and of many other people, the following result.

THEOREM 8.1 (Taylor). — *Let l be an odd prime, and $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$ a continuous irreducible representation such that:*

- (1) ρ is unramified at all but finitely many primes;
- (2) ρ is odd (i.e., $\det \rho(c) = -1$);
- (3) $\rho|_{G_l}$ is an extension of χ_2 by $\varepsilon^n \chi_1$, with $\chi_1|_{I_l}$ and $\chi_2|_{I_l}$ of finite order and n a non-zero positive integer, such that $\varepsilon^n \chi_1 \chi_2^{-1}(I_l)$ is not pro- l .

Then there is a totally real number field E , a regular algebraic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_E)$ and a place λ of the field of coefficients of π above l such that $\rho_{\pi, \lambda}$ (the λ -adic representation associated to π) is equivalent to $\rho|_{G_E}$.

As a consequence, such a ρ has an L -function, and this L -function is meromorphic and satisfies the expected functional equation. Also, under some mild hypothesis, it follows that ρ occurs in some $H^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_l(r))$, as in Conjecture 2.1. The idea that allows one to use the results of Skinner and Wiles, and others, concerning ρ such that $\bar{\rho}$ has a soluble image, is to use abelian varieties with real multiplications such that the $\bar{\rho}$ of the Theorem above is related to the l -torsion, and such that the p -torsion (for some other prime p) gives a suitable soluble image. The existence of such abelian varieties is proved by applying Skolem type results to Hilbert-Blumenthal modular varieties (see for example [37]).

Ramakrishna proved in [42] that, under mild hypotheses, a continuous mod l representation $\bar{\rho}: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(k)$ (not supposed to be modular) can be lifted to a representation $\rho: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(W(k))$ over the Witt vectors of k , with ρ unramified at almost all primes. More recently, in [43], he has proved that one can even obtain that ρ is semi-stable or crystalline at l . His main innovation is to consider deformation problems that lead to a universal deformation ring $W(k)$. He requires $\rho|_{G_p}$ to be of the form $\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$ for suitable p . Note that this is stronger than a condition on $\rho|_{I_p}$ (until now, only conditions on $\rho|_{I_p}$ were imposed), which makes it reasonable that the deformation ring will be small. One finds a slight generalization of Ramakrishna's results in [58], where they are used to prove some more cases of the Artin conjecture. Combined with Theorem 8.1, one obtains that $\bar{\rho}$ as above becomes modular after restriction to G_E with E a suitable totally real extension of \mathbb{Q} (see [57]). This can be seen as a “potential” version of Serre's conjecture.

Khare has used Ramakrishna's work ([34]) to give another proof of certain “ $R = T$ ” theorems. Assuming $\bar{\rho}$ to be modular, one gets an “ $R = T$ ” theorem for Ramakrishna's deformation problem (the main result of [26] implies that Ramakrishna's lift ρ is modular). Starting with this result, Khare proves that for large enough Σ , $R_{O, \Sigma} \rightarrow \mathbb{T}_{O, \Sigma}$ is an isomorphism; his proof avoids the special arguments of Wiles and Taylor-Wiles in the minimal case. Of course, this last result suffices for proving modularity results. (It seems that from this one also easily obtains the result for all Σ .)

Appendix A. GALOIS REPRESENTATIONS ASSOCIATED TO MODULAR FORMS

The aim of this section is to recall what we need about the Galois representations associated to modular forms. For simplicity, we only discuss the case of forms of weight two, so that we only need to deal with the cohomology of the constant sheaf on modular curves. We use the now standard point of view that was initiated by Deligne in [17]. As a general reference for this section, we recommend [25].

The object from which everything originates here is the Shimura datum $(\mathrm{GL}_2, \mathbb{H}^\pm)$, with $\mathrm{GL}_2(\mathbb{R})$ acting on $\mathbb{H}^\pm = \mathbb{P}^1(\mathbb{C}) - \mathbb{P}^1(\mathbb{R})$ in the usual way. Let \mathbb{A} denote the ring of adèles of \mathbb{Q} , and \mathbb{A}_f its subring of the finite adèles. For every compact open subgroup U of $\mathrm{GL}_2(\mathbb{A}_f)$, let $X_U^0(\mathbb{C})$ denote the complex analytic variety $\mathrm{GL}_2(\mathbb{Q}) \backslash (\mathbb{H}^\pm \times \mathrm{GL}_2(\mathbb{A}_f)/U)$; it can be compactified, by adding a finite number of points, to a smooth compact Riemann surface (usually not connected) $X_U(\mathbb{C})$. We denote the associated complex algebraic curve by $X_{U,\mathbb{C}}$. The interpretation of these curves as moduli spaces of elliptic curves with level structures give models $X_{U,\mathbb{Q}}$ over \mathbb{Q} . The inverse limit $X_{\mathbb{Q}}$ of the $X_{U,\mathbb{Q}}$ has an action, from the right, by $\mathrm{GL}_2(\mathbb{A}_f)$. For l prime, the $\overline{\mathbb{Q}}_l$ -vector space:

$$H_l := \lim_U H^1(X_{U,\overline{\mathbb{Q}},\mathrm{et}}, \overline{\mathbb{Q}}_l)$$

has an action by $G_{\mathbb{Q}} \times \mathrm{GL}_2(\mathbb{A}_f)$. In order to understand the decomposition of H_l as a representation of $\mathrm{GL}_2(\mathbb{A}_f)$, one uses the Hodge decomposition:

$$H^1(X_U(\mathbb{C}), \mathbb{C}) = \Omega^1(X_U(\mathbb{C})) \oplus \overline{\Omega^1(X_U(\mathbb{C}))}.$$

On $\Omega^1(X_U(\mathbb{C}))$ one no longer has an action of $\mathrm{GL}_2(\mathbb{A}_f)$, but is still is a module over the Hecke algebra associated to U : the convolution algebra of compactly supported bi- U -invariant functions on $\mathrm{GL}_2(\mathbb{A}_f)$ (say that $\mathrm{GL}_2(\hat{\mathbb{Z}})$ has measure one). The q -expansion principle and some theory of smooth irreducible representations of the $\mathrm{GL}_2(\mathbb{Q}_p)$ show that $\Omega^1(X(\mathbb{C}))$ decomposes into a direct sum of irreducible ones, each one occurring only once:

$$\Omega^1(X(\mathbb{C})) = \lim_U \Omega^1(X_U(\mathbb{C})) = \bigoplus_f \pi_f,$$

where f runs through the set of newforms of weight two with coefficients in \mathbb{C} . We recall that for all p we have a chosen embedding $\overline{\mathbb{Q}} \rightarrow \overline{\mathbb{Q}}_p$. It follows that H_l decomposes as a direct sum:

$$H_l \cong \bigoplus_f \rho_f^\vee \otimes \pi_f,$$

with f running through the set of weight two newforms with coefficients in $\overline{\mathbb{Q}}_l$, and with $\rho_f: G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$ a continuous representation. We note that ρ_f is realized over any finite extension over which π_f is defined. The representation π_f is a restricted tensor product $\otimes'_p \pi_{f,p}$ over all primes, with each π_p an infinite dimensional irreducible admissible representation of $\mathrm{GL}_2(\mathbb{Q}_p)$. On the Galois side, we define, for each prime p , $\rho_{f,p} := \rho_f|_{G_p}$.

With these definitions, one knows that, for $p \neq l$, $\pi_{f,p}$ and $\rho_{f,p}$ determine each other via a suitably normalized local Langlands correspondence. (This was first proved at the unramified places by Eichler and Shimura, then for $\pi_{f,p}$ principal series or special by Langlands, then for $p \neq 2$ by Deligne, and finally for all p by Carayol, and simplified by Nyssen.) The representation $\rho_{f,l}$ is usually not determined by $\pi_{f,l}$ (just think of the case where f corresponds to an elliptic curve with split multiplicative reduction at l , where $\rho_{f,l}$ almost determines the elliptic curve), but Saito has shown in [46] that the (ϕ, N, G_l) -module obtained by forgetting the filtration of the filtered (ϕ, N, G_l) -module corresponding to $\rho_{f,l}$ via Fontaine's functor ([31, §10]) corresponds to $\pi_{f,l}$. (Actually, in the weight two case that we consider this is in fact easily deduced from the results for $p \neq l$; see [13, Appendix B].)

In order to fix notation, let us give a precise description of this local correspondence, so that there is no ambiguity about the normalization. To do this, we first recall that the best way to formulate the local Langlands correspondence is in terms of the Weil-Deligne group (see [55, §4] and [18]). For p prime, the Weil group W_p of \mathbb{Q}_p is the subgroup of G_p consisting of elements whose image in $G_{\mathbb{F}_p}$ is in $\text{Frob}_p^{\mathbb{Z}}$ ($\text{Frob}_p: x \mapsto x^p$ is the arithmetic Frobenius). The Weil-Deligne group is an object W'_p that is defined so that a representation of W'_p on a finite dimensional E -vector space ($E \supset \mathbb{Q}$) is a pair (V, N) with V a continuous representation of W_p (with the discrete topology on V), and a nilpotent endomorphism N of V such that $wNw^{-1}v = pNv$ for all v in V and w in W_p mapping to Frob_p . Such a pair is called F -semisimple if V is semisimple as a representation of W_p . With these definitions, there are canonical bijections between the set of (isomorphism classes of) infinite dimensional irreducible admissible representations of $\text{GL}_2(\mathbb{Q}_p)$ over $\overline{\mathbb{Q}}$, and the set of 2-dimensional F -semisimple representations of W'_p over $\overline{\mathbb{Q}}$. These bijections are such, that local L and ε -factors on both sides (suitably normalized) correspond, and they are compatible with the action of $G_{\mathbb{Q}}$ on both sides. For $p \neq 2$, this is easy, since one can easily write down the elements on both sides (on the Galois side, one uses that the wild inertia acts reducibly). For $p = 2$, this is harder; the general case was worked out by Kutzko. If V is a finite dimensional K -vector space, with K a finite extension of \mathbb{Q}_l , and $p \neq l$, there is an equivalence between representations of W'_p on V as above, and continuous representations of W_p on V .

Following [13], we normalize the local Langlands correspondence WD in such a way that $\rho_{f,p}|_{W_p}$, viewed as a representation of W'_p over $\overline{\mathbb{Q}}_l$, is isomorphic to $\overline{\mathbb{Q}}_l \otimes_{\overline{\mathbb{Q}}} \text{WD}(\pi_{f,p})$, for each newform f with coefficients in $\overline{\mathbb{Q}}$. If $\sigma(\pi_{f,p})$ is as in [8], then we have $\text{WD}(\pi_{f,p}) = \sigma(\pi_{f,p}) \otimes \chi$, where χ is the unramified character that sends Frob_p to p . If $\pi_{f,p}$ is unramified, i.e., if p does not divide the level of f , then $\rho_{f,p}$ is unramified and $\rho_{f,p}(\text{Frob}_p)$ is semisimple (remember that we are in weight two) and has characteristic polynomial $X^2 - t_p X + p s_p$, where t_p and s_p are the eigenvalues of f for the Hecke and diamond operators T_p and S_p that are defined by the double cosets $U \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} U$ and $U \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} U$,

with $U = \mathrm{GL}_2(\mathbb{Z}_p)$. The determinant of $\rho_{f,p}$ is $\varepsilon\chi_{\pi_{f,p}}$, where $\chi_{\pi_{f,p}}$ is the central character of $\pi_{f,p}$, viewed as a character of W_p^{ab} via the isomorphism of class field theory under which the image of p in $\mathbb{Q}_p^*/\mathbb{Z}_p^*$ corresponds to Frob_p . If χ is a continuous character $\mathbb{Q}_p^* \rightarrow \overline{\mathbb{Q}}^*$, then $\mathrm{WD}(\pi_{f,p} \otimes (\chi \det)) \cong \mathrm{WD}(\pi_{f,p}) \otimes \chi$.

To finish this section, let us recall some facts about the classification of the two-dimensional semisimple representations of W_p , over \mathbb{C} , say. Let ρ be such a representation, on a \mathbb{C} -vector space V , say. If ρ is reducible, it is a sum of two characters. Suppose ρ irreducible. Since the wild inertia subgroup I_p^{wild} of I_p acts on V via a finite p -group, it acts via two characters (possibly equal), unless $p = 2$ (recall that the dimension of an irreducible complex representation of a finite group divides the order of the group). Suppose that $\rho|_{I_p^{\mathrm{wild}}}$ splits as a sum of two distinct characters. Then W_p acts on the set of the two corresponding lines in V , and non-trivially because ρ is irreducible. It follows that ρ becomes reducible over a quadratic extension of \mathbb{Q}_p . Suppose now that I_p^{wild} acts via scalars on V . Then, considering the action of W_p/I_p (which is the semi-direct product of \mathbb{Z} by I_p^{tame}) on $\mathbb{P}(V)$, one sees that, again, ρ becomes reducible over a quadratic extension of \mathbb{Q}_p . So the conclusion is this: the two-dimensional complex semisimple representations of W_p are sums of two characters, or induced from a character of W_K with K quadratic over \mathbb{Q}_p , or such that $p = 2$ and I_p^{wild} acts irreducibly (these latter ones were first classified by Weil [61], in his “exercices dyadiques”; clearly, the title of [4] is inspired by this).

Appendix B. TYPES ASSOCIATED TO l -ADIC REPRESENTATIONS AND TO ELLIPTIC CURVES

Let l be any prime. We recall that an extended l -type (over $\overline{\mathbb{Q}}_l$) is an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}}_l$ of the Weil-Deligne group W'_l of \mathbb{Q}_l (see Appendix A), and that types are isomorphism classes of restrictions to I_l of extended l -types. We want to describe how one attaches an extended type to a continuous representation $\rho: G_l \rightarrow \mathrm{GL}_2(O)$, with O the ring of integers of a finite extension K of \mathbb{Q}_l contained in $\overline{\mathbb{Q}}_l$, under the assumption that ρ is potentially Barsotti-Tate. So let ρ be such a representation, and let F be a finite extension of \mathbb{Q}_l over which ρ becomes Barsotti-Tate, i.e., such that $\rho|_{G_F}$ is isomorphic to $\mathcal{G}(\overline{\mathbb{Q}}_l)$ for some l -divisible group with O -action over the ring of integers O_F of F . Of course, one solution to this is simply to apply Fontaine’s $D_{\mathrm{st},F}$ functor as in [31, §10(b)], but in this simple case of p -divisible group schemes one can be more explicit. Another reason for doing this more explicitly is that one wants to do computations in the case of elliptic curves. For more details we refer to [13, Appendix B].

The representation ρ we have corresponds to an l -divisible group $\mathcal{G}_{\mathbb{Q}_l}$ over \mathbb{Q}_l , with O -action. Let F be a finite Galois extension of \mathbb{Q}_l such that \mathcal{G}_F extends (uniquely, by [54, Theorem 4]) to an l -divisible group \mathcal{G}_{O_F} over O_F , with O -action. Let Γ denote $\mathrm{Gal}(F/\mathbb{Q}_l)$.

For every σ in Γ , we have commutative diagrams:

$$\begin{array}{ccc}
 \mathcal{G}_{O_F} & \xrightarrow[\sim]{[\sigma]} & \mathcal{G}_{O_F} \\
 \downarrow & & \downarrow \\
 \mathrm{Spec}(O_F) & \xrightarrow[\mathrm{Spec}(\sigma)]{\sim} & \mathrm{Spec}(O_F)
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathcal{G}_{k_F} & \xrightarrow[\sim]{[\sigma]} & \mathcal{G}_{k_F} \\
 \downarrow & & \downarrow \\
 \mathrm{Spec}(k_F) & \xrightarrow[\mathrm{Spec}(\sigma)]{\sim} & \mathrm{Spec}(k_F)
 \end{array}$$

with k_F the residue field of O_F . The last diagram gives a right action of G_l on \mathcal{G}_{k_F}/k_F . Let $d: W_l \rightarrow \mathbb{Z}$ be the morphism such that σ in W_l induces the $d(\sigma)$ th power of the absolute Frobenius on $\overline{\mathbb{F}}_l$. Then we get a morphism from W_l to $(\mathbb{Q} \otimes \mathrm{End}_{k_F}(\mathcal{G}_{k_F}))^*$ by sending σ to $[\sigma]^{-1} \mathrm{Frob}_{\mathrm{abs}}^{d(\sigma)}$. Now let \mathbb{D} denote the contravariant Dieudonné module functor. Then $\mathbb{Q} \otimes \mathbb{D}(\mathcal{G}_{k_F})^\vee$, with $^\vee$ denoting $W(k_F)$ -dual, is a free $K \otimes_{\mathbb{Z}_l} W(k_F)$ -module of rank two, with a left action by W_l . The extended l -type $\mathrm{WD}(\rho)$ associated to ρ is then the two-dimensional $\overline{\mathbb{Q}}_l$ vector space obtained by base change via $K \otimes_{\mathbb{Z}_l} W(k_F) \rightarrow \overline{\mathbb{Q}}_l$ (note that both K and F are subfields of $\overline{\mathbb{Q}}_l$); the monodromy operator is defined to be zero, as we have good reduction.

We can repeat the construction above, with $\mathcal{G}_{\mathbb{Q}_l}$ replaced by an elliptic curve E over \mathbb{Q}_l , with good reduction E_{O_F} over O_F . Then one gets morphisms:

$$W_l \longrightarrow (\mathbb{Z}[1/l] \otimes \mathrm{End}_{k_F}(E_{k_F}))^* \longrightarrow (\overline{\mathbb{Q}} \otimes \mathrm{End}_{k_F}(E_{k_F}))^* \cong \mathrm{GL}_2(\overline{\mathbb{Q}}).$$

More generally, one can start with a newform f with coefficients in $\overline{\mathbb{Q}}$, and then one has:

$$\mathrm{WD}(\rho_{f,l}) = \overline{\mathbb{Q}}_l \otimes_{\overline{\mathbb{Q}}} \mathrm{WD}(\pi_{f,l}),$$

by the results of Appendix A.

Acknowledgements

I would like to thank Christophe Breuil, Fred Diamond, Reinie Ern e, Eyal Goren, Chandrashekhara Khare, Rutger Noot and Richard Taylor for their comments on an earlier version of this text, and Michel Gros for the reference to [5]. Attending the conference ‘‘Modularity of Elliptic Curves and Beyond’’ at the MSRI in Berkeley, December 6–10, 1999, was very useful, as well as the series of talks given by Brian Conrad on this subject in Rome, July 1999.

REFERENCES

- [1] P. BERTHELOT – *Alt erations de vari et es alg ebriques*. S eminaire Bourbaki, 1995–96, expos e 815, Ast erisque **241**, S.M.F. (1997), 273–311.
- [2] C. BREUIL – *Sch emas en groupes et modules filtr es*. C. R. Acad. Sci. Paris **328** (1999), 93–97.

- [3] C. BREUIL – *Groupes p -divisibles, groupes finis et modules filtrés*. Preprint, 1999. To appear in *Annals of Mathematics*.
- [4] C. BREUIL, B. CONRAD, F. DIAMOND, R. TAYLOR – *On the modularity of elliptic curves over \mathbb{Q} : wild 3-adic exercises*. Preprint, <http://www.math.harvard.edu/>
- [5] C.J. BUSHNELL – *Smooth representations of p -adic groups: the role of compact open subgroups*. *Proceedings of the international congress of mathematicians*, Vol. 1, 2 (Zürich, 1994), 770–779, Birkhäuser, Basel, 1995.
- [6] K. BUZZARD, M. DICKINSON, N. SHEPHERD-BARRON, R. TAYLOR – *On icosahedral Artin representations*. Preprint, <http://www.math.harvard.edu/>
- [7] K. BUZZARD, R. TAYLOR – *Companion forms and weight one forms*. *Ann. of Math.* (2) **149** (1999), no. 3, 905–91
- [8] H. CARAYOL – *Sur les représentations l -adiques associées aux formes modulaires de Hilbert*. *Ann. Sci. Éc. Norm. Sup.* **19** (1986), 409–468.
- [9] L. CLOZEL – *Motifs et formes automorphes*. *Automorphic forms, Shimura varieties, and L -functions*, Vol. I, edited by L. Clozel and J.S. Milne, *Perspectives in Mathematics* 10, Academic Press (1990), 77–159.
- [10] P. COLMEZ, J-M. FONTAINE – *Construction des représentations p -adiques semi-stables*. Preprint Ecole Normale Supérieure, LMENS-99-18.
- [11] B. CONRAD – *Finite group schemes over bases with low ramification*. *Compositio Math.* **119** (1999), 239–320.
- [12] B. CONRAD – *Ramified deformation problems*. *Duke Math. J.* **97** (1999), 439–513.
- [13] B. CONRAD, F. DIAMOND, R. TAYLOR – *Modularity of certain potentially Barsotti-Tate Galois representations*. *J.A.M.S.* **12** (1999), 521–567.
- [14] G. CORNELL, J. SILVERMAN, G. STEVENS, editors – *Modular Forms and Fermat’s Last Theorem* Springer-Verlag, 1997.
- [15] H. DARMON – *A proof of the full Shimura-Taniyama-Weil conjecture is announced*. *Notices of the A.M.S.*, December 1999.
- [16] H. DARMON, F. DIAMOND, R. TAYLOR – *Fermat’s Last Theorem*. *Elliptic curves, modular forms and Fermat’s last theorem* (Hong Kong, 1993). Second edition. International Press, Cambridge, MA, 1997, 2–140.
- [17] P. DELIGNE – *Formes modulaires et représentations l -adiques*. *Séminaire Bourbaki*, 1968–69, exposé 355, *Lecture Notes in Mathematics* **179**, Springer (1969), 139–172.
- [18] P. DELIGNE – *Formes modulaires et représentations de $GL(2)$* . *Antwerp II, Lecture Notes in Math.* **349**, Springer (1973), 55–106.
- [19] B. DE SMIT, H. LENSTRA – *Explicit construction of universal deformation rings*. In [14], pp. 313–326.

- [20] B. DE SMIT, K. RUBIN, R. SCHOOF – *Criteria for complete intersections*. In [14], 343–355.
- [21] F. DIAMOND – *The refined conjecture of Serre*. In *Elliptic curves, modular forms and Fermat’s last theorem* (Hong Kong, 1993). International Press, 1995, 22–37.
- [22] F. DIAMOND – *An extension of Wiles’ results*. In [14], 475–498.
- [23] F. DIAMOND – *On deformation rings and Hecke rings*. *Ann. Math.* **144** (1996), 137–166.
- [24] F. DIAMOND – *The Taylor–Wiles construction and multiplicity one*. *Invent. math.* **128** (1997), 379–391.
- [25] F. DIAMOND, J. IM – *Modular forms and modular curves*. In “Seminar on Fermat’s last theorem”, Canadian Mathematical Society Conference Proceedings 17, 1995 (V. Kumar Murty, editor).
- [26] F. DIAMOND, R. TAYLOR – *Lifting modular mod l representations*. *Duke Math. J.* **74** (1994), no. 2, 253–269.
- [27] M. DICKINSON – *On the modularity of certain 2-adic Galois representations*. Preprint.
- [28] S.J. EDIXHOVEN – *The weight in Serre’s conjectures on modular forms*. *Invent. math.* **109** (1992), 563–594.
- [29] J-M. FONTAINE – *Groupes p -divisibles sur les corps locaux*. *Astérisque* **47–48**, S.M.F., 1977.
- [30] J-M. FONTAINE – *Représentations p -adiques semi-stables*. In “Périodes p -adiques,” *Astérisque* **223**, S.M.F., 1994, p. 113–184.
- [31] J-M. FONTAINE, B. MAZUR – *Geometric Galois representations*. In “Elliptic curves, modular forms and Fermat’s Last Theorem” (Hong Kong, 1993), International Press, 1995, pp. 41–78.
- [32] K. FUJIWARA – *Deformation rings and Hecke rings in the totally real case*. Preprint.
- [33] C. KHARE – *A local analysis of congruences in the (p, p) -case: Part II*. To appear in *Invent. math.*
- [34] C. KHARE – *On isomorphisms between deformation and Hecke rings*. Work in progress.
- [35] R.P. LANGLANDS – *Base change for $GL(2)$* . *Ann. of Math. Studies* **96**, Princeton University Press, Princeton, 1980.
- [36] J. MANOHARMAYUM – *Pairs of mod 3 and mod 5 representations arising from elliptic curves*. Thesis, Cambridge University, 1998.
- [37] L. MORET-BAILLY – *Groupes de Picard et problèmes de Skolem II*. *Ann. Sci. ENS* **22** (1989), 181–194.
- [38] R. NOOT – *Abelian varieties arising from Mumford’s Shimura curves*. Preprint, Rennes, 1997.

- [39] J. OESTERLÉ – *Nouvelles approches du “théorème” de Fermat*. Séminaire Bourbaki, 1987–88, exposé 694, Astérisque **161–162**, S.M.F. (1988), 175–186.
- [40] J. OESTERLÉ – *Travaux de Wiles (et Taylor, ...), partie II*. Séminaire Bourbaki, 1994–95, exposé 804, Astérisque **237**, S.M.F. (1996), 333–355.
- [41] R. RAMAKRISHNA – *On a variation of Mazur’s deformation functor*. *Compositio Math.* **87** (1993), 269–286.
- [42] R. RAMAKRISHNA – *Lifting Galois representations*. *Invent. math.* **138** (1999), 537–562.
- [43] R. RAMAKRISHNA – *Deforming Galois representations and the conjectures of Serre and Fontaine-Mazur*. Preprint.
- [44] K.A. RIBET, – *On modular representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms*. *Invent. math.* **100** (1990), 431–476.
- [45] K. RUBIN – *Modularity of mod 5 representations*. In [14], 463–474.
- [46] T. SAITO – *Modular forms and p -adic Hodge theory*. *Invent. math.* **129** (1997), 607–620.
- [47] J-P. SERRE – *Représentations l -adiques*. Kyoto Int. Symposium on Algebraic Number Theory, Japan Soc. for the Promotion of Science (1977), pp. 177–193, (Oeuvres, t. III, pp. 384–400).
- [48] J-P. SERRE – *Sur les représentations modulaires de degré 2 de $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$* . *Duke Math. J.* **54** (1987), 179–230.
- [49] J-P. SERRE – *Travaux de Wiles (et Taylor, ...), partie I*. Séminaire Bourbaki, 1994–95, exposé 803, Astérisque **237**, S.M.F. (1996), 319–332.
- [50] C. SKINNER, A. WILES – *Residually reducible representations and modular forms*. Preprint. To appear in *Pub. Math. IHES*.
- [51] C. SKINNER, A. WILES – *Base change and a problem of Serre*. Preprint.
- [52] C. SKINNER, A. WILES – *Nearly ordinary deformations of irreducible residual representations*. Preprint.
- [53] N. SHEPHERD-BARRON, R. TAYLOR – *Mod 2 and mod 5 icosahedral representations*. *J. Amer. Math. Soc.* **10** (1997), 281–332.
- [54] J. TATE – *p -divisible groups*. Proceedings of a conference on local fields, (Driebergen, 1966), Springer, 1967, 158–183.
- [55] J. TATE – *Number theoretic background*. Proceedings of Symposia in Pure Mathematics **33** (1979), part 2, 3–26.
- [56] R. TAYLOR – *Icosahedral Galois representations*. *Pacific J. Math.*: Special issue: Olga Taussky-Todd: in memoriam (1997), 337–347.
- [57] R. TAYLOR – *Remarks on a conjecture of Fontaine and Mazur*. Preprint, <http://www.math.harvard.edu/>
- [58] R. TAYLOR – *On icosahedral Artin representations II*. Preprint, <http://www.math.harvard.edu/>

- [59] R. TAYLOR, A. WILES – *Ring-theoretic properties of certain Hecke algebras*. Ann. Math. **141** (1995), 553–572.
- [60] J. TUNNELL – *Artin’s conjecture for representations of octahedral type*. Bull. A.M.S. **5** (1981), 173–175.
- [61] A. WEIL – *Exercices dyadiques*. Invent. math. **27** (1974), 1–22.
- [62] A. WILES – *Modular elliptic curves and Fermat’s Last Theorem*. Ann. Math. **141** (1995), 443–551.

Bas EDIXHOVEN¹

I.R.M.A.R., U.M.R. 6625 du CNRS

Université de Rennes I

Campus de Beaulieu

F-35042 RENNES Cedex

E-mail : edix@maths.univ-rennes1.fr

¹partially supported by the Institut Universitaire de France, and by the European TMR Network Contract ERB FMRX 960006 “arithmetic algebraic geometry”.

VARIÉTÉS DE FANO RÉELLES [d'après C. Viterbo]

par Viatcheslav KHARLAMOV

1. INTRODUCTION

1.1. Structures algébriques sur une variété différentiable

Toute variété compacte différentiable est difféomorphe à la partie réelle d'une variété algébrique réelle sans singularités. Ce théorème est l'aboutissement des travaux de H. Seifert, J. Nash, R. Thom, A. Wallace, A. Tognoli, H. King et d'autres. Ici, une variété algébrique réelle non singulière est un ensemble défini dans un espace réel, affine ou projectif au choix, par un système d'équations polynomiales réelles sans aucun point singulier réel (voir [20] et [5] pour les détails et les problèmes encore ouverts ; notons aussi que les démonstrations se font par approximation).

Pour nous, une variété (algébrique) réelle est une variété (algébrique) complexe compacte lisse $M(\mathbb{C})$ munie d'une involution antiholomorphe, appelée *structure réelle* ou *conjugaison complexe*. L'ensemble des points fixes de cette involution est une sous-variété analytique réelle, appelée *partie réelle* de $M(\mathbb{C})$ et notée $M(\mathbb{R})$. Nous souhaitons étudier les relations entre les propriétés topologiques et géométriques de $M(\mathbb{C})$ et $M(\mathbb{R})$. Sur ce sujet, de nombreux résultats sont connus grâce à l'étude du 16-ième problème de Hilbert qui pose la question des restrictions sur la topologie de $M(\mathbb{R})$ imposées par le degré des équations le définissant ; citons en particulier les travaux d'Arnol'd, Rokhlin et leurs successeurs, voir les surveys [41], [37] et [14].

Après résolution des singularités, le théorème de Nash-Tognoli-King montre que, pour toute variété différentiable L , il existe une variété algébrique projective non singulière $M(\mathbb{C})$ stable par la conjugaison complexe et dont L est l'ensemble des points réels.

Quelques remarques concernant ce résultat s'imposent. Ici, la variété L considérée peut avoir un nombre fini quelconque de composantes connexes, alors que $M(\mathbb{C})$ peut toujours être choisie connexe (voir par exemple [36]). Notons aussi que si $M(\mathbb{C})$ est une variété projective munie d'une structure réelle, elle peut être dotée d'un plongement projectif $M(\mathbb{C}) \rightarrow P_n(\mathbb{C})$ qui conserve la structure réelle : partant d'un plongement quelconque

dans $P_n(\mathbb{C})$, on restreint le plongement de Segre $P_n(\mathbb{C}) \times P_n(\overline{\mathbb{C}}) \rightarrow P_{n^2+2n}(\mathbb{C})$ au graphe de la structure réelle de M vu comme sous-variété de $M(\mathbb{C}) \times M(\overline{\mathbb{C}})$. (Notez que la structure réelle sur $P_{n^2+2n}(\mathbb{C})$ qui commute avec le plongement de Segre est équivalente à la structure réelle usuelle.)

Fixons dans la suite une variété algébrique réelle dont on note M l'ensemble des points complexes, et L une composante connexe de la partie réelle. Il est nécessaire d'avoir choisi dans \mathbb{C} une racine carrée de -1 , notée i , ce qui équipe M de l'orientation complexe.

Un voisinage tubulaire de L dans M ne dépend que de L , à difféomorphisme près. En effet, la multiplication par i induit un isomorphisme entre le fibré tangent $T(L)$ de L et son fibré normal $N(L)$ dans M . De plus, les nombres d'Euler de ces fibrés satisfont la formule d'adjonction

$$(1) \quad e(N(L)) = (-1)^{\frac{1}{2}n(n-1)} e(T(L)),$$

ce qui provient du choix de l'orientation complexe sur M et de la relation $e_1 \wedge \cdots \wedge e_n \wedge i e_1 \wedge \cdots \wedge i e_n = (-1)^b e_1 \wedge i e_1 \wedge \cdots \wedge e_n \wedge i e_n$, où $b = \frac{1}{2}n(n-1)$. Notons que dès que l'on a choisi une métrique sur L , le fibré cotangent $T^*(L)$ s'identifie à $T(L)$ puis au voisinage tubulaire de L . L'orientation complexe de M restreinte à $T^*(L)$ donne l'orientation canonique (symplectique) de $T^*(L)$.

Du point de vue analytique complexe, le germe de M le long de L est déterminé par la structure analytique réelle de L . On peut noter que les voisinages tubulaires de L contiennent des tubes réels de Grauert [18]. Ce sont des variétés de Stein difféomorphes à $T^*(L)$; voir l'article de D. Burns [7] qui traite des structures réelles sur les tubes de Grauert.

1.2. Une curieuse conjecture de Nash

Il est naturel de chercher à réaliser les variétés différentiables comme variétés réelles en imposant des restrictions sur le complexifié. C'est suivant cette logique que J. Nash [29] (1952) demanda, tout en émettant des réserves, s'il est possible de réaliser toute variété différentiable compacte comme un ensemble algébrique réel sans singularités ayant un complexifié rationnel (ici, rationnel signifie birationnellement isomorphe, sur \mathbb{R} , à l'espace projectif ; notez qu'une variété algébrique réelle peut être rationnelle sur \mathbb{C} sans l'être sur \mathbb{R}). En fait cette conjecture avait été réfutée, avant même d'avoir été formulée, par le théorème de Comessatti [9] (1914) de classification des modèles minimaux des surfaces réelles et rationnelles sur \mathbb{C} . Il en découle en effet que toute composante orientable de la partie réelle d'une surface réelle rationnelle sur \mathbb{C} est une sphère ou un tore.

On peut en donner une démonstration rapide (c'est la même démarche qui mène aux inégalités d'Arnold [21]). Elle est basée sur (1). Soient M une surface réelle et rationnelle sur \mathbb{C} , $f \in H_2(M(\mathbb{C}); \mathbb{R})$ la classe d'homologie réalisée par une composante réelle L orientable et orientée, et h la classe d'homologie de la section hyperplane. La

forme d'intersection est négative sur l'orthogonal de h , parce que la surface est rationnelle ($H^2 = H^{1,1}$). De plus $fh = 0$ puisque f et h appartiennent à deux espaces propres différents de l'action induite par la conjugaison complexe, et (1) montre que $f^2 = -\chi(L)$. Donc, $\chi(L) \geq 0$ et $g(L) \leq 1$.

En dimension trois, la conjecture est encore fautive. C'est une conséquence de la classification des modèles minimaux des variétés réelles rationnelles de dimension 3, donnée par J. Kollár [23]. Plus précisément, J. Kollár montre que parmi des composantes connexes des parties réelles de toutes les telles variétés, il n'y a au plus qu'un nombre fini de variétés à courbure sectionnelle strictement négative (voir [23] pour des résultats encore plus précis ; signalons seulement que les composantes orientables sont des sommes connexes $X \# aP_3(\mathbb{R}) \# b(S^1 \times S^2)$ où X est soit une somme connexe des espaces lenticulaires, soit une fibration de Seifert, soit un $S^1 \times S^1$ -fibré au-dessus de S^1 , soit son $\mathbb{Z}/2$ -quotient non ramifié). En dimension supérieure, la conjecture est toujours ouverte.

Notons deux résultats positifs : toute variété compacte de dimension trois est homéomorphe à la partie réelle d'une variété algébrique projective réelle et rationnelle sur \mathbb{R} , mais singulière ; et toute telle variété est aussi difféomorphe à la partie réelle d'une variété algébrique sans singularité réelle, rationnelle sur \mathbb{R} , mais non projective (le premier est dû à R. Benedetti et A. Marin [4] ; le second à J. Kollár [24]). Signalons aussi la solution d'une version purement topologique de la conjecture de Nash en dimension trois par S. Akbulut et H. King [1], ainsi que par R. Benedetti et A. Marin [4], puis en toutes dimensions par G. Mikhalkin [28].

1.3. Variétés fortement de Fano

Les techniques employées par C. Viterbo sont celles de la géométrie symplectique. Par conséquent, dans ce qui suit, nous préférons le langage de la géométrie symplectique à celui de la géométrie algébrique, qui n'interviendra à nouveau qu'à la fin de la démonstration.

Ainsi, *une variété de Fano* est une variété compacte complexe lisse kählérienne M dont la classe de Kähler ω est égale à celle de Chern : $\omega = c_1(M)$ (variété symplectique monotone ou, en d'autres termes, à fibré anticanonique ample ; pour démontrer l'équivalence on peut faire appel au critère de Nakai-Moishezon). Toute variété de Fano M est *uniréglée*, c'est-à-dire qu'il existe deux variétés compactes complexes F et N avec une application régulière $N \rightarrow F$ à fibre générique $P_1(\mathbb{C})$, ainsi qu'une application régulière de degré fini $N \rightarrow M$. D'autre part, toute variété de Fano est simplement connexe et satisfait $\text{Pic}(M) = H^2(M)$ (voir le rapport de O. Debarre [11] ; ceux qui préfèrent une démonstration moins algébrique de $\pi_1 = 1$ peuvent utiliser la métrique de Calabi-Yau et le théorème de Myers [27]).

Comme exemples de variétés de Fano, on peut citer les hypersurfaces lisses de degré $\leq n$ dans P_n , et plus généralement les intersections complètes lisses de multidegré (d_1, \dots, d_k) dans P_n avec $n \geq d_1 + \dots + d_k$. Dans ces exemples, $-K = (n + 1 - \sum d_i)h$ où K est le

diviseur canonique et h le diviseur de la section hyperplane. Par conséquent, si la partie réelle n'est pas vide, la structure réelle se relève dans le fibré associé à h , et se prolonge donc en une structure réelle sur P_n ayant des points réels. Or toute structure réelle sur P_n ayant des points réels est conjuguée à la structure usuelle, donc finalement les seules structures réelles sur les variétés ci-dessus sont les structures induites par celle de P_n (rappelons que lorsque n est impair, il y a des structures réelles sur P_n sans points réels : pour $n = 1$, un exemple est fourni par la conique $x^2 + y^2 + 1 = 0$).

Notons aussi que, si un tube de Grauert est une partie affine d'une variété projective lisse, cette variété projective est de Fano.

Le théorème de Viterbo ne s'applique qu'aux variétés de Fano satisfaisant deux hypothèses supplémentaires. Une variété est dite *minimalement uniréglée*, si elle est uniréglée par des courbes rationnelles dont l'aire est minimale parmi celles des courbes holomorphes de M . Une variété est *fortement de Fano*, si elle est minimalement uniréglée et s'il existe un diviseur ample H de M dont l'intersection avec une courbe rationnelle uniréglatrice est égale à 1.

Ces deux hypothèses sont vérifiées pour les intersections complètes dans P_n de multi-degré (d_1, \dots, d_k) si $d_1 + \dots + d_k < n$ (qui sont uniréglées par des droites), mais elles excluent le cas $d_1 + \dots + d_k = n$. En dimension trois, le premier cas non couvert est celui des quartiques dans P_4 .

Si M est une variété munie d'une structure réelle conj et d'une métrique de Kähler g , quitte à remplacer g par $\frac{1}{2}(g + g^{\text{conj}})$, on peut se ramener au cas où la forme de Kähler Ω est anti-invariante par l'action de conj^* . Pour cette structure, $M(\mathbb{R})$ est lagrangienne.

Dans tout ce qui suit, les métriques de Kähler seront ainsi choisies.

1.4. Théorème de Viterbo et Conjecture de Kollár

THÉORÈME 1.1. — *Aucune composante connexe du lieu réel d'une variété réelle fortement de Fano n'admet de métrique à courbure sectionnelle strictement négative.*

Ce résultat de C. Viterbo [39] confirme la conjecture, toujours ouverte, de J. Kollár d'après laquelle aucune variété réelle uniréglée de *dimension* ≥ 3 n'a de composantes réelles à courbure sectionnelle strictement négative. (En dimension deux, déjà les éclatés de $P_2(\mathbb{R})$ en au moins deux points portent une métrique hyperbolique.)

D'après J. Kollár [23], la conjecture est vraie en dimension trois sauf pour un nombre fini de classes de variétés à déformation sur \mathbb{C} près. Il pose aussi la question : l'existence d'une métrique hyperbolique implique-t-elle que la variété soit de type général ?

Notons deux résultats intermédiaires, d'intérêt indépendant, obtenus au cours de la démonstration du théorème 1.1 :

PROPOSITION 1.2. — *Si une composante connexe L de la partie réelle d'une variété fortement de Fano réelle admet une métrique sans géodésique contractile, il existe parmi*

les courbes d'aire minimale et d'intersection 1 avec H (voir 1.3) une courbe réelle (lisse) dont la partie réelle est non vide et contenue dans L .

PROPOSITION 1.3. — *Soit M une variété fortement de Fano dont les courbes unirégultrices sont d'aire α . Toute sous-variété lagrangienne L admet une membrane holomorphe à bord dans L et d'aire inférieure à α . Si de plus, L n'a pas de métrique à courbure sectionnelle strictement négative, la membrane peut être choisie d'aire inférieure à $\frac{1}{2}\alpha$.*

La courbe réelle de la proposition 1.2 est une courbe rationnelle régulatrice si H est très ample, comme il l'est, par exemple, dans le cas d'intersections complètes. Par ailleurs, si une telle courbe réelle régulatrice existe, L contient un fermé semi-algébrique d'intérieur non vide couvert par les parties réelles des courbes régulatrices réelles.

Notons aussi que parmi les courbes algébriquement équivalentes à deux fois une courbe régulatrice, il y a toujours, sans aucune hypothèse sur L et pour toute variété de Fano, suffisamment de courbes réelles lisses pour couvrir un ouvert dense de la partie réelle de M (voir [25]).

Si on avait pu étendre 1.1 et 1.2 à toutes les variétés de Fano et toutes les familles régulatrices, ces résultats auraient entraîné les deux corollaires suivants : *Le tube de Grauert (voir 1.1) d'une variété possédant une métrique à courbure sectionnelle strictement négative n'est jamais la partie affine d'une variété projective lisse. Une variété fortement de Fano réelle a au plus une composante connexe réelle admettant une métrique sans géodésique contractile.*

Dans cet exposé nous expliquerons les grandes lignes de la démonstration sans prétendre entrer dans les détails, en particulier lorsqu'il s'agit des arguments de mise en position générale.

1.5. Classification des surfaces de Fano réelles

Le cas des surfaces est extrêmement simple (et très probablement assez différent du cas des variétés de dimensions supérieures). Comme toutes les variétés de Fano de dimension donnée, les surfaces de Fano réelles (plus traditionnellement appelées surfaces de Del Pezzo), forment un nombre fini de classes à déformation (réelle) près. Leur classification à déformation près découle de celle de Comessatti des surfaces rationnelles, à l'aide des modèles anti(-bi)canoniques des surfaces. Pour les détails de la démonstration, voir [13] (et pour des informations supplémentaires, voir [34]).

Suivant la terminologie traditionnelle, nous appelons *degré* d'une surface de Del Pezzo le nombre $d = c_1^2$. Une *simplification de Morse* est une opération topologique qui consiste à enlever soit une composante sphérique, soit une anse d'une composante. Notons V_q la somme connexe de q copies de $P_2(\mathbb{R})$.

PROPOSITION 1.4. — *À une exception près, chaque surface de Del Pezzo réelle (M, conj) de degré $d \geq 1$ est déterminée à déformation près par la topologie de $M(\mathbb{R})$. Dans le cas*

exceptionnel, $d = 8$ et $M(\mathbb{R})$ est vide, le nombre de classes de déformation est deux, et dans l'une M/conj est Spin, alors que dans l'autre non.

En chaque degré, les types topologiques de $M(\mathbb{R})$ sont les suivants et ceux obtenus par des suites de simplifications de Morse à partir d'eux : $M(\mathbb{R}) = V_1$ en degré 9; $M(\mathbb{R}) = V_2$ et $S^1 \times S^1$ en degré 8; $M(\mathbb{R}) = V_3$ en degré 7; $M(\mathbb{R}) = V_4$ et $S^1 \times S^1$ en degré 6; $M(\mathbb{R}) = V_5$ en degré 5; $M(\mathbb{R}) = V_6, S^1 \times S^1$ et $2S^2$ en degré 4; $M(\mathbb{R}) = V_7$ et $V_1 \sqcup S^2$ en degré 3; $M(\mathbb{R}) = V_8, 2V_1, V_2 \sqcup S^2, S^1 \times S^1$ et $4S^2$ en degré 2; $M(\mathbb{R}) = V_9, V_2 \sqcup V_1, V_3 \sqcup S^2$ et $V_1 \sqcup 4S^2$ en degré 1. (Ici, \sqcup est la réunion disjointe et nS^2 est la réunion disjointe de n sphères de dimension 2.)

Parmi les surfaces de Del Pezzo, les seules variétés fortement de Fano sont le plan projectif ($d = 9$) et les surfaces géométriquement réglées ($d = 8$). Les deux structures exceptionnelles du théorème précédent sont les deux structures réelles sur $P_1(\mathbb{C}) \times P_1(\mathbb{C})$ sans point réel.

1.6. Remerciements

Je voudrais remercier V. Blanloeil, O. Debarre, T. Delzant, M. Paun et J. Y. Welschinger qui m'ont aidé à décortiquer la démarche de C. Viterbo et m'ont apporté une aide précieuse au cours de la rédaction de ces notes. Je remercie aussi C. Viterbo qui a patiemment répondu à mes multiples questions, et Y. Eliashberg qui m'a fait découvrir une autre approche aux problèmes traités dans ce rapport.

2. SYSTÈMES HAMILTONIENS ASSOCIÉS

Soit M une variété complexe compacte lisse kählérienne (c'est-à-dire, une variété munie d'une forme symplectique Ω et d'une structure complexe J telle que $\Omega(\cdot, J(\cdot))$ est une métrique riemannienne). Notons n la dimension complexe de M . Considérons une sous-variété lagrangienne L de M (une sous-variété différentiable réelle de dimension réelle n dont les espaces tangents sont isotropes pour Ω ; l'exemple essentiel pour nous est donné par la partie réelle d'une variété kählérienne munie d'une structure réelle, voir 1.3). D'après le théorème de Darboux-Weinstein, il existe un voisinage U de L dans M symplectomorphe à un voisinage de la section nulle dans $T^*(L)$ (avec sa structure symplectique canonique $dp \wedge dq$, comparer 1.1). À partir de maintenant, nous les identifions et considérerons U comme situé simultanément dans M et dans $T^*(L)$.

Fixons une métrique g sur L (sans aucun rapport avec la métrique kählérienne de M). Il est bien connu que les géodésiques γ de (L, g) sont les points critiques de l'énergie

$$E(\gamma) = \int_0^\pi |\gamma'(t)|^2 dt$$

et que les graphes des géodésiques (trajectoires géodésiques) dans $T^*(L)$ sont les points critiques de l'action hamiltonnienne

$$A(\gamma) = \int (pdq - Hdt), \quad \text{où } H = \frac{1}{2} \sum g^{ij} p_i p_j \text{ et } pdq = \sum p_i dq^i.$$

En associant à H le champ X_H défini par $\Omega(\xi, X_H) = dH(\xi)$ (ce qui est équivalent à $X_H = J\nabla H$), on retrouve le flot géodésique sur le bord ∂T_r^* des tubes $T_r^* = \{(q, p) \in T^*(L) \mid \sum g^{ij} p_i p_j \leq r^2\}$ de rayon r .

L'objectif est d'étendre ce tableau à M en mélangeant le flot géodésique sur le bord d'un voisinage tubulaire de L avec la géométrie de l'opérateur de Cauchy-Riemann loin de ce bord. Plus précisément, on considère l'espace

$$\tilde{\Lambda}_0 = \{(\gamma, u) \mid u \in C^\infty(D^2, M), \gamma = \partial u\}$$

et on introduit sur cet espace une famille d'actions dépendant d'un paramètre $\lambda \geq 0$, $\lambda \in \mathbb{R}$,

$$A_\lambda(\gamma, u) = \int_{D^2} u^* \Omega - \lambda \int_{\partial D^2} H_{\epsilon, \rho} \circ \gamma dt, \quad \text{où } dt = \frac{dz}{2\pi i}.$$

La fonction $H_{\epsilon, \rho} : M \rightarrow \mathbb{R}$ ne dépend que de la distance à L , elle est choisie égale à 1 au dehors de $U = T_r^*$ et égale à $h_{\epsilon, \rho}(|p|)$, $|p|^2 = \sum g^{ij} p_i p_j$, dans U . La fonction $h_{\epsilon, \rho}$ est de classe C^∞ , identiquement nulle sur $[0, \epsilon]$, croissante et strictement convexe sur $[\epsilon, \frac{1}{2}(\epsilon + \rho)]$, croissante et strictement concave sur $[\frac{1}{2}(\epsilon + \rho), \rho]$ et égale à 1 sur $[\rho, \infty[$ (bien entendu, $0 < \epsilon < \rho < r$). On suppose que $\frac{1}{2}(\epsilon + \rho)$ n'est pas parmi les longueurs des géodésiques fermées de L .

Pour un couple (γ, u) tel que γ soit un lacet réduit à un point y , on a

$$A_\lambda(\gamma, u) = \int_\alpha \Omega - \lambda H_{\epsilon, \rho}(y)$$

où α est la classe d'homologie réalisée par $u : S^2 = D^2/S^1 \rightarrow M$.

Le calcul direct de la première variation de A_λ donne

$$dA_\lambda(\delta\gamma) = - \int_{\partial D^2} \langle \delta\gamma, J\gamma' + \lambda \nabla H_{\epsilon, \rho} \rangle dt$$

où $\delta\gamma$ est un champ de vecteurs tangents à M le long de γ . Comme conséquence, les points critiques de A_0 sont les couples (γ, u) tels que γ soit un lacet constant, et $u : S^2 = D^2/S^1 \rightarrow M$ une sphère holomorphe ; et la valeur critique de A_0 correspondante à (γ, u) est $\int_\alpha \Omega$ où α est la classe d'homologie de u . Pour $\lambda > 0$, il y a trois types de points critiques suivant les trois types de γ : les lacets constants dans $M \setminus T_{< \rho}^*$, les lacets constants dans T_ϵ^* , et les trajectoires géodésiques périodiques dans $T_{< \rho}^* \setminus T_\epsilon^*$. Les valeurs critiques correspondantes aux deux premiers types de points critiques sont respectivement : $\int_\alpha \Omega - \lambda$ si $\lambda > 0$ et le lacet est au dehors de T_ρ^* , et $\int_\alpha \Omega$ si $\lambda > 0$ et le lacet est dans T_ϵ^* .

Toutes les trajectoires critiques γ sont des trajectoires périodiques de période 1 du flot X_H . Ce flot sur ∂T_τ^* , $\epsilon < \tau < \rho$, est le flot (co)géodésique reparamétré. Les projections des

trajectoires périodiques de période 1 contenues dans ∂T_τ^* sont les géodésiques de longueur $h'(\tau)$.

3. TRAJECTOIRES DE FLOER ET COURBES HOLOMORPHES

Les trajectoires de Floer sont les analogues des séparatrices (trajectoires du flot du gradient reliant les points critiques) dans la théorie des fonctions de Morse sur les variétés de dimension finie. Étant guidé par cette analogie et le calcul de la première variation de A_λ , on définit une trajectoire de Floer reliant deux points critiques, l'un correspondant à un lacet constant sur L et l'autre à un lacet constant au dehors de U , comme une solution $u : \mathbb{R} \times S^1 \rightarrow M$ de l'équation

$$(2) \quad \frac{\partial u}{\partial s} + J \frac{\partial u}{\partial t} + \lambda \nabla H_{\epsilon, \rho}(u) = 0$$

telle que

$$\lim_{s \rightarrow +\infty} u(s, t) = y \in M \setminus U \quad \text{et} \quad \lim_{s \rightarrow -\infty} u(s, t) = x \in L.$$

Une écriture équivalente de (2) est $\bar{\partial}u = -\lambda \nabla H_{\epsilon, \rho}(u)$. En particulier, pour $\lambda = 0$ ces trajectoires de Floer sont les courbes rationnelles holomorphes $S^2 \rightarrow M$ qui passent par un point x de L et un point y de $M \setminus U$. De plus, les courbes rationnelles minimales représentent les trajectoires de Floer qui relient deux valeurs critiques consécutives de A_0 .

Le groupe \mathbb{C}^* agit de façon usuelle sur les solutions de l'équation (2). Dans le cas des variétés minimalement réglées, cette action est libre sur l'ensemble des solutions avec $[u] = \alpha$, α étant la classe des courbes régulatrices.

L'action est souvent réduite à S^1 en normalisant u comme suit : $\int_{s \leq 0} u^* \Omega = \frac{1}{2} \int_\alpha \Omega$. À partir de maintenant, nous identifions la sphère S^2 et le compactifié de $\mathbb{R} \times S^1$ par deux points $\pm\infty$. Ainsi, sur S^2 il y aura toujours deux points marqués.

3.1. Sélection des courbes holomorphes

Considérons une variété fortement de Fano M minimalement réglée par des courbes rationnelles $u : S^2 \rightarrow M$ de classe $\alpha = [u] \in H_2(M; \mathbb{Z}) = \pi_2(M)$. Posons

$$\text{Mor}_\alpha = \{u : S^2 \rightarrow M \mid \bar{\partial}u = 0, [u] = \alpha\}$$

et notons $\text{ev}_{\pm\infty}$ le morphisme d'évaluation $\text{Mor}_\alpha \rightarrow M$, $u \mapsto u(\pm\infty)$.

L'affirmation suivante est une conséquence du lemme de cassage de Mori et du théorème de compacité de Bishop.

PROPOSITION 3.1. — *$\text{Mor}_\alpha / \mathbb{C}^*$ est une variété projective, et pour tout couple de points $x \neq y$ de M , il n'existe à reparamétrisation près qu'un nombre fini (éventuellement nul) de courbes rationnelles $u : S^2 \rightarrow M$ avec $u(-\infty) = x$ et $u(+\infty) = y$.*

Notons $M(x)$ la sous-variété de M des points qui peuvent être joints à x par $u \in \text{Mor}_\alpha$ (c'est-à-dire, $y \in M(x)$ si et seulement s'il existe $u \in \text{Mor}_\alpha$ tel que $u(-\infty) = x$ et $u(+\infty) = y$). Par des arguments standard de position générale, on déduit de 3.1 le résultat suivant (rappelons que H est le diviseur ample tel que $H \circ \alpha = 1$, voir 1.3).

COROLLAIRE 3.2. — *Il existe un entier $k > 0$ tel que pour tout point générique x de M et toute intersection complète générique A d'hypersurfaces H_1, \dots, H_k avec $H_i \in |d_i H|$ l'intersection de $\text{ev}_{-\infty}^{-1}(x)$ et $\text{ev}_{+\infty}^{-1}(A)$ est non vide, finie et transverse.*

(On peut compléter 3.1 et 3.2 et démontrer que $\dim_{\mathbb{C}} \text{Mor}_\alpha / \mathbb{C}^* = c_1(M) \cap \alpha + n - 1$ et $k = \dim_{\mathbb{C}} M(x) = c_1(M) \cap \alpha - 1 > 0$.)

Notons le nombre de points d'intersection de $\text{ev}_{-\infty}^{-1}(x)$ et $\text{ev}_{+\infty}^{-1}(A)$ par $\nu(x, A)$. En fait, A peut être remplacée par une variété différentiable dans la même classe d'homologie, et le nombre de points d'intersection est alors remplacé par le nombre algébrique.

Si L est une sous-variété lagrangienne de M , le point générique x peut être choisi dans L , et la variété A dans le complémentaire d'un voisinage tubulaire U de L (aucune sous-variété complexe de M ne contient L ; et si d est pair et suffisamment grand, il existe une hypersurface $H' \in |dH|$ qui ne rencontre pas L).

Dans ce qui suit on fixe $x \in L$ et $A \subset (M \setminus U)$ ainsi choisis.

4. CASSAGE DES COURBES HOLOMORPHES

Le théorème suivant produit une géodésique fermée munie d'une membrane partiellement holomorphe et dont on contrôle l'indice i_{CZ} dit de Conley-Zehnder-Duistermaat. C'est ce résultat qui permettra à la fin soit de construire une courbe réelle, soit de réfuter l'existence d'une métrique hyperbolique.

THÉORÈME 4.1. — *Si L admet une métrique sans géodésique contractile, il existe $\lambda > 0$, une trajectoire périodique γ de période λ du flot Hamiltonien $X_{\epsilon, \rho}$, et un disque $u : D^2 \rightarrow M$ tels que $\gamma = \partial u$, $u(0) = x$, $0 < A_\lambda(\gamma, u) < \int_\alpha \Omega$ et enfin, si la métrique est de plus à courbure sectionnelle strictement négative, tels que $i_{CZ}(\gamma, u) \geq n$.*

Idée de la démonstration. Supposons que le point $x \in L$ et la variété A sont fixés comme dans 3.1. Considérons l'espace $\mathcal{M}_\lambda(\alpha, x, A)$ des trajectoires de Floer de classe α reliant x à A et introduisons le diagramme de bifurcation :

$$\Delta = \{(\lambda, m) | \lambda \geq 0, m \in \mathcal{M}_\lambda(\alpha, x, A) / \mathbb{C}^*\}.$$

Notons que $\lambda \leq \int_\alpha \Omega$ pour tout $(\lambda, m) \in \Delta$ parce que $\int_\alpha \Omega - \lambda$ est la valeur critique de A_λ au-dessus de 0 sur la trajectoire de Floer m .

LEMME 4.2. — Si Δ n'est pas compact, il existe une suite finie de trajectoires périodiques γ_i , $1 \leq i \leq k$, et une suite de trajectoires de Floer $(\gamma_{i-1}, u_i, \gamma_i)$, $1 \leq i \leq k+1$, reliant γ_{i-1}

et γ_i , telles que γ_0 soit la trajectoire constante $x \in L$, γ_{k+1} soit une trajectoire constante $y \in A$, les autres trajectoires soient dans $T_\rho^* \setminus T_\epsilon^*$, et $[u_1 \# \cdots \# u_{k+1}] = \alpha$.

Pour la démonstration de 4.2, qui nécessite de fines renormalisations et perturbations, on renvoie à [39].

Dès qu'une telle suite de trajectoires de Floer est obtenue, on prend $u = u_1$. L'inégalité $A_\lambda(\gamma, u) > 0$ est due à la croissance de A_λ le long des trajectoires de Floer, et $A_\lambda < \int_\alpha \Omega$ découle de $k \geq 1$ et $\sum_{i=1}^{k+1} A_\lambda(\gamma_{i-1}, u_i, \gamma_i) = \int_\alpha \Omega - \lambda$ (rappelons que $H_{\epsilon, \rho}(x) = 0$ et $H_{\epsilon, \rho}(M \setminus T_r^*) = 1$).

Si la métrique est à courbure sectionnelle strictement négative, la trajectoire γ est non dégénérée. On peut alors perturber $H_{\epsilon, \rho}$ en $\tilde{H}_{\epsilon, \rho}$ générique et de Morse sur L , et obtenir, de la même façon que dans la démonstration de 4.2, une trajectoire de Floer $(\tilde{\gamma}, \tilde{u})$ avec $i_{CZ}(\tilde{\gamma}, \tilde{u}) = i_{CZ}(\gamma, u)$ et \tilde{u} qui relie $\tilde{\gamma}$ à un point qui réalise le minimum de $\tilde{H}_{\epsilon, \rho}$ sur L . On obtient donc l'inégalité $i_{CZ}(\tilde{\gamma}) \geq n - \text{ind Hess } \tilde{H}_{\epsilon, \rho}|_L = n$ (voir [39] pour les détails techniques de cette perturbation).

Pour montrer que Δ n'est pas compact, notons que, pour x et A suffisamment génériques, la structure complexe de M est régulière en chaque trajectoire de Floer $(0, m) \in \Delta$ (le fibré normal des courbes unirégulatrices génériques est ample). Donc la projection $\Delta \rightarrow \mathbb{R}_+$, $(\lambda, m) \mapsto \lambda$ est propre au-dessus d'un voisinage de $\lambda = 0$, et la partie de Δ pour laquelle λ est voisin de 0 est une variété lisse de dimension 1 dont le bord est constitué des $\nu(x, A)$ points satisfaisant $\lambda = 0$. Cette variété est orientée près de ces points et ils sont tous positifs. Si Δ était compact, la perturbation décrite dans [19] aurait donné une variété compacte orientée (l'orientation provenant de [17]) à bord orienté positif. \square

Si les géodésiques de L ne sont pas contractiles (cas des métriques à courbure négative), le disque D construit dans la démonstration du théorème 4.1 ainsi que les autres trajectoires de Floer $(\gamma_{i-1}, u_i, \gamma_i)$ (un disque supplémentaire plus, éventuellement, des cylindres) sortent tous de T_ρ^* . Donc, chacun contient une partie $C_{\epsilon, \rho}^i$ non vide à bord sur laquelle $u : C_{\epsilon, \rho}^i \rightarrow M$ est une courbe holomorphe propre de $M \setminus T_\rho^*$. Il conviendra d'attribuer l'indice $i = 1$ au disque du théorème 4.1 et de poser $C(u) = C_{\epsilon, \rho} = \bigcup C_{\epsilon, \rho}^i$.

PROPOSITION 4.3. — *L'aire des courbes $C_{\epsilon, \rho}^k$ est uniformément séparée de 0 et ∞ : il existe $\delta > 0$ tel que $\delta < \text{aire } C_{\epsilon, \rho}^k < \text{aire } C(u) < \int_\alpha \Omega$ pour tout ϵ, ρ .*

Preuve. — On a $A_\lambda(\gamma, u) < \int_\alpha \Omega$ et

$$\begin{aligned} A_\lambda(\gamma, u) &= \frac{1}{2} \sum_i \int_{(\gamma_{i-1}, u, \gamma_i)} |\partial_s y|^2 + |J\partial_t u + X_H(u_s)|^2 \\ &\geq \int_{C(u)} \frac{1}{2} (|\partial_s u|^2 + |J\partial_t u|^2) = \int_{C(u)} \Omega = \text{aire } C(u), \end{aligned}$$

d'où la borne supérieure. Comme les géodésiques fermées différentes ne sont jamais homologues, chaque courbe $C_{\epsilon, \rho}^k$ sort de U . En particulier, chacune coupe $\partial T_{\frac{2}{3}\rho}^*$ et on obtient une minoration uniforme en appliquant l'inégalité de Bishop-Lelong-Wirtinger. \square

5. COMPARAISON DES INDICES

Il s'agit de l'indice $i_M \in \mathbb{Z}_+$ de Morse des géodésiques, de la classe de Maslov $\mu(L) \in H^2(M, L) = H^2(M, T_\rho)$ des sous-variétés lagrangiennes L , et de l'indice $i_{CZ} \in \mathbb{Z}_+$ de Conley-Zehnder-Duistermaat des orbites périodiques non dégénérées d'un flot hamiltonien. Pour que ce dernier indice soit bien défini, il suffit que l'orbite périodique γ soit équipée d'une membrane, c'est-à-dire d'une surface orientée $u : S \rightarrow M$ (pas forcément un disque) qui a l'orbite pour bord et dont l'orientation du bord coïncide avec la direction du flot. Une telle membrane définit alors, à homotopie près, une unique trivialisation symplectique du fibré u^*T_*M et par définition i_{CZ} vaut "l'indice spectral" du flot hamiltonien linéarisé le long de γ , voir [10] et [15]. Finalement, i_{CZ} ne dépend que de la classe de u dans $H_2(M, \gamma)$. À l'aide de la même trivialisation de u^*T_*M , on peut définir la classe de Maslov comme l'obstruction à une extension sur u de la trivialisation symplectique sur $\partial u \subset L$. Rappelons que l'image de $\mu(L)$ dans $H^2(M)$ coïncide avec $2c_1(M)$.

Signalons quelques spécificités de ce choix de normalisation de l'indice de Conley-Zehnder-Duistermaat (la normalisation change d'un auteur à l'autre). Pour le graphe $\gamma \subset T^*(L)$ d'une géodésique fermée non dégénérée contractile et u dans $T^*(L)$, on a $i_{CZ}(\gamma, u) = i_M$; pour γ et u constants et placés en un point critique non dégénéré de H , on a $i_{CZ} = n - \text{ind}(\text{Hess } H)$ si le hessien de H est suffisamment petit.

THÉORÈME 5.1. — *Soit γ une trajectoire périodique du flot de $X_{\epsilon, \rho}$ associée à une géodésique fermée γ de L . Si u est une surface dans M dont le bord est γ ,*

$$i_{CZ}(\gamma, u) = i_M(\gamma) + \langle \mu(L), u \rangle \quad \text{ou} \quad i_{CZ}(\gamma, u) = i_M(\gamma) + \langle \mu(L), u \rangle - 1.$$

Les trajectoires γ traitées dans ce théorème sont contenues dans ∂T_x^* avec $\epsilon < x < \rho$, $x \neq \frac{1}{2}(\epsilon + \rho)$. La première formule, sans -1 , a lieu si $x < \frac{1}{2}(\epsilon + \rho)$, et la deuxième pour $x > \frac{1}{2}(\epsilon + \rho)$.

Dans le cas de $M = \mathbb{C}^n$, les formules du théorème 5.1 sont établies dans [38]. La démonstration dans le cas général s'effectue de la même façon.

6. MEMBRANES HOLOMORPHES

Supposons que $M = M(\mathbb{C})$ est une variété fortement de Fano réelle et L est une composante de $M(\mathbb{R})$ qui admet une métrique à courbure sectionnelle strictement négative. (Les résultats de cette section se généralisent aux sous-variétés lagrangiennes L réelles analytiques dans une variété M fortement de Fano, sous la même hypothèse sur la métrique.)

PROPOSITION 6.1. — *Il existe une courbe holomorphe compacte C' dont le bord s'appuie sur L et qui vérifie*

$$(3) \quad \langle \mu(L), C' \rangle \geq n \quad \text{et} \quad \text{aire}(C') = \frac{1}{2} \int_\alpha \Omega.$$

La démonstration s'appuie sur le phénomène remarquable de prolongement des courbes holomorphes à travers les variétés analytiques totalement réelles maximales (par exemple les variétés lagrangiennes analytiques réelles, ce qui inclut les parties réelles des variétés algébriques). Ce résultat est spécifique à cette classe de sous-variétés et il est purement local (étrangement, la preuve s'appuie à nouveau sur l'étude, cette fois-ci locale, des familles de courbes rationnelles).

THÉORÈME 6.2. — *Soient W un ouvert de \mathbb{C}^n et A un sous-ensemble analytique complexe de $W \setminus \mathbb{R}^n$ de dimension 1 en tout point (une courbe holomorphe propre dans $W \setminus \mathbb{R}^n$). Si A est invariant par la conjugaison complexe, son adhérence $\bar{A} \cap W$ est un sous-ensemble analytique complexe de W .*

Pour les courbes d'aire bornée ce résultat est dû à B. Shiffman [33], et sans cette hypothèse, à H. Alexander [2].

Démonstration de 6.1. D'après le théorème 6.2, si L est la partie réelle $L = M(\mathbb{R})$ d'une variété compacte complexe $M = M(\mathbb{C})$ munie d'une structure réelle, il nous suffit de construire dans $M \setminus L$ une courbe complexe analytique C propre vérifiant (3). (Si L n'est qu'une variété lagrangienne analytique réelle, la conjugaison complexe peut être introduite près de L et le théorème s'applique de nouveau.) Une telle courbe C s'obtient à partir des courbes $C_{\epsilon, \rho'}^i \subset M \setminus T_\rho^*$, avec $\rho' \leq \rho$ et $\rho', \rho \rightarrow 0$, par le procédé diagonal, en utilisant l'inégalité de Bishop-Lelong-Wirtinger et le théorème de Bishop de compacité appliqués aux courbes $C_{\epsilon, \rho'}^i \cap (M \setminus T_\rho^*)$. Ce procédé peut être appliqué au moins deux fois, à $i = 1$ et à $i = k + 1$ (voir 4.2). Ainsi nous obtenons dans $M \setminus L$ deux courbes holomorphes propres C' et C'' telles que, d'après 4.3,

$$(4) \quad n, \delta < \text{aire}(C'), \quad \delta < \text{aire}(C''), \quad \text{aire}(C') + \text{aire}(C'') \leq \int_\alpha \Omega.$$

(En particulier, C' et C'' sont non vides.) Dans le cas de variétés réelles, on double C' à l'aide de la conjugaison complexe. L'adhérence C_1 de la courbe doublée $C' \cup \text{conj } C'$ est une courbe holomorphe compacte (éventuellement singulière). Évidemment, $\text{aire}(C_1) = 2 \text{aire}(C')$. Grâce à (4), on a l'inégalité $\text{aire}(C_1) < 2 \int_\alpha \Omega - 2\delta$, et par l'hypothèse de minimalité de α

$$\text{aire}(C_1) = \int_\alpha \Omega.$$

De plus, le même raisonnement s'applique à $C_2 = C'' \cup \text{conj } C''$ et $\text{aire}(C_1) + \text{aire}(C_2) \leq 2 \int_\alpha \Omega$. Ceci montre que la trajectoire de Floer ne s'est cassée qu'en deux et que $C' \cup C''$ en est la limite au sens des courants (pas de multiplicités dans les limites). Par suite, $[C', \partial C'] = u_*[D, \partial D] \in H_2(M, U)$, et par conséquent $\langle \mu(L), C' \rangle = \langle \mu(L), u \rangle$. D'après 5.1 et 4.1, et comme l'indice de Morse des géodésiques est nul dans le cas des métriques hyperboliques,

$$\langle \mu(L), u \rangle = i_{CZ}(\gamma) - i_M(\gamma) = i_{CZ}(\gamma) \geq n.$$

□

Fin de la démonstration de 1.1. Par la naturalité de la classe de Maslov-Chern, $\text{conj}_* \mu = -\mu$, et on déduit de (3) que

$$(5) \quad c_1(M) \cap C_i \geq n, \quad \text{pour } i = 1, 2.$$

Maintenant, il ne reste qu'à combiner quelques résultats connus de la théorie des variétés de Fano.

Supposons que M est de dimension ≥ 3 (en dimension 2 le résultat est trivial modulo la classification des surfaces de Del Pezzo). D'abord, on déduit de (5) l'inégalité $c_1(M) \cap \alpha \geq n$. D'après la minimalité de α , on obtient $c_1(M) \cap C \geq n$ pour toute courbe rationnelle C (de même que pour toute courbe holomorphe). Le théorème de Wiśnevski [40] nous dit que dans ce cas $\text{Pic}(M) = \mathbb{Z}$. Puisque $H \cap \alpha = 1$, on en déduit que $c_1(M) = mH$ avec $m \geq n$. Finalement, on est dans le cadre du théorème de Kobayashi-Ochiai [22] qui dit que M est soit $P_n(\mathbb{C})$, soit une quadrique dans $P_{n+1}(\mathbb{C})$. Sur ces variétés, les seules structures réelles ayant des points réels sont les structures usuelles (voir 1.3), donc L est ou bien $P_n(\mathbb{R})$, ou bien un produit croisé de deux sphères. Dans tous les cas le revêtement universel de L n'est pas contractile et, d'après le théorème de Cartan-Hadamard, L n'admet pas de métrique à courbure sectionnelle négative.

7. REMARQUES

7.1. Théorèmes de finitude

On peut déduire des théorèmes de finitude sur \mathbb{C} (voir le rapport [11]) que les variétés de Fano réelles de dimension donnée ne forment qu'un nombre fini de classes à déformation réelle près.

Si on considère une classe quelconque de variétés à déformation sur \mathbb{C} près, les variétés réelles de cette classe, considérées à déformation réelle près, sont-elles en nombre fini ? Cette question est à notre connaissance ouverte. La réponse est positive pour les courbes, les variétés de type général, les variétés abéliennes et toutes les surfaces sauf peut-être les surfaces rationnelles.

Une question étroitement liée est la question de finitude des structures réelles sur une variété donnée à conjugaison près. Comme ci-dessus, la réponse est positive pour les courbes, les variétés de type général, les variétés abéliennes et toutes les surfaces sauf peut-être les surfaces rationnelles.

Pour les courbes, ces résultats sont très classiques. Pour les variétés abéliennes, ils sont dûs à A. Borel et J.-P. Serre [6]. Pour les surfaces, nous les avons démontrés avec I. Itenberg, après des discussions avec C. Viterbo, en nous appuyant sur la classification des surfaces et les connaissances de leurs groupes d'automorphismes.

7.2. Espaces linéaires contenus dans des intersections complètes réelles

Si la proposition 1.2 garantit l'existence de droites réelles sur des intersections complètes sous une hypothèse géométrique, il existe un tout autre critère qui utilise une forte hypothèse d'imparité de degré, mais en revanche qui se généralise aux espaces linéaires de dimensions quelconques.

La variété des r -plans contenus dans une sous-variété V de $P^n(\mathbb{C})$ définie par des équations de multidegré $d = (d_1, \dots, d_k)$ est le lieu des zéros d'une section d'un fibré vectoriel sur la grassmannienne $G(r, P^n)$. La décomposition en classes de Schubert de la classe de Chern de degré maximal de ce fibré est donnée dans [12].

Grâce au théorème de Bézout (et le fait que les cycles de Schubert sont réels), pour montrer l'existence d'espaces linéaires réels contenus dans des intersections complètes réelles, il suffit de montrer qu'un des coefficients dans la décomposition de la classe de Chern est impair. C'est ainsi que O. Debarre et L. Manivel démontrent (sans aucune hypothèse d'hyperbolicité) que si chaque composante de d est impaire, et que toute intersection complète de multidegré d contient des r -espaces linéaires complexes (condition qu'on peut exprimer en terme d'une simple inégalité sur n, r et d), alors les intersections réelles contiennent des r -espaces réels.

Comme conséquence, la partie réelle des intersections complètes fortement de Fano de degré impair (sans aucune hypothèse d'hyperbolicité) contient un fermé semi-algébrique d'intérieur non vide couvert par des droites réelles. Le cas le plus simple est celui d'hypersurfaces cubiques dans P^4 .

7.3. Classe de Maslov, orientabilité et les formes de Rokhlin-Guillou-Marin et de Viro

La classe de Maslov-Chern apparaissait déjà en géométrie algébrique réelle, partiellement grâce à ses rapports avec l'orientabilité du lieu réel. Rappelons quelques-unes de ces apparitions.

Soient $M = M(\mathbb{C})$ une variété compacte complexe munie d'une structure réelle, et $L = M(\mathbb{R})$ sa partie réelle. Le critère d'orientabilité de L le plus simple, et le plus typique, est le suivant (découvert et redécouvert par différents auteurs, voir en particulier [35], [16], [26]) : L est orientable si $c_1(M)$ est divisible par 2 et $H^1(M; \mathbb{Z}/2) = 0$.

Cette observation est étroitement liée à la construction de l'indice de Maslov. En effet, la classe de Maslov est une classe de cohomologie relative $\mu \in H^2(M, L; \mathbb{Z})$, et si $H^1(M; \mathbb{Z}/2) = 0$, elle se relève en une unique classe appartenant à $H^1(L; \mathbb{Z}/2)$ laquelle n'est autre que la classe de Stiefel-Whitney. Ainsi, L est orientable si $c_1 = 0 \in H^2(M; \mathbb{Z}/2)$ et $H^1(M; \mathbb{Z}/2) = 0$. Même mieux, sous ces hypothèses la classe de Maslov $\mu \in H^2(M, L)$ et sa moitié $\tilde{\mu} \in H^2(M, L)$ (la dernière étant définie à l'aide des lagrangiens orientés) se relèvent (après réduction mod 4) en respectivement $\mu_4 \in H^1(L; \mathbb{Z}/4)$ et $\mu_2 \in H^1(L; \mathbb{Z}/2)$, avec $\mu_4 = 2\mu_2$.

En dimension deux, comme c'était remarqué par N. Netsvetaev [30], μ_2 est la différence entre les formes quadratiques de Rokhlin [32] et de Viro [37]. Supposons pour simplifier que $H^1(M; \mathbb{Z}/2) = 0$. Si $[L] = 0 \in H^2(M; \mathbb{Z}/2)$, la forme de Viro $q_V : H_1(L; \mathbb{Z}/2) \rightarrow \mathbb{Z}/2$ est définie comme suit : prenons une courbe lisse simple γ de L , multiplions son champ de vecteurs tangents par $\sqrt{-1}$ et déplaçons γ le long de ce champ dans $M \setminus L$, le nombre $q_V[\gamma]$ est alors égal à 1 plus le nombre d'enlacement entre γ déplacé et L . La forme de Rokhlin $q_R : H_1(L; \mathbb{Z}/2) \rightarrow \mathbb{Z}/2$ est définie si L est orientable et $[L] = w_2(M)$. Si de plus $w_2(M) = 0$, elle diffère de q_V par $\chi(F) + \nu(F)$, où F est une surface plongée dans M dont le bord est réduit à γ et qui est normale à L le long de γ . D'où $q_V = q_R + \mu_2$ si $[L] = w_2(M) = 0$.

Si $[L] = w_2(M) \neq 0$ (et L n'est pas forcément orientable), la forme de Rokhlin peut être remplacée par celle de Guillou-Marin $q_{RGM} : H_1(L; \mathbb{Z}/2) \rightarrow \mathbb{Z}/4$. En présence d'une courbe réelle C avec $[C] = [L] = w_2(M)$, la forme de Viro se généralise en $q_V : H_1(L; \mathbb{Z}/2) \rightarrow \mathbb{Z}/4$ et on obtient aisément l'identité $q_V = q_{RGM} + \mu_4$.

Si $H^1(M) = 0$ et $c_1(M) = 0$, on définit de la même façon la classe de Maslov résiduelle $\mu_0 \in H^1(L)$. Dans le cas des variétés de Calabi-Yau, à cause de l'existence de n -formes holomorphes réelles, cette classe résiduelle est nulle.

Dès que le lieu réel est orientable on s'intéresse à des orientations privilégiées. Dans le cas des courbes, lorsque $C(\mathbb{R})$ décompose $C(\mathbb{C})$, deux orientations opposées canoniques, dites orientations complexes, apparaissent sur $C(\mathbb{R})$. Ces orientations, introduites par V. Rokhlin dans les années 70, jouent depuis un rôle extrêmement important dans l'étude des courbes algébriques sur les surfaces, en particulier, dans l'étude des courbes planes. L'outil principal est la formule d'orientations complexes de Rokhlin (voir [14] pour des références, un survey bref du cas des courbes sur les surfaces, une formule complémentaire découverte récemment par S. Orevkov [31] et sa généralisation due à J. Y. Welschinger [42]).

7.4. Approche d'Eliashberg

Alors que nous étions en train d'achever la préparation de ce rapport, Y. Eliashberg a proposé une autre approche au théorème de Viterbo, utilisant des techniques récemment développées avec A. Givental et H. Hofer (voir l'exposé de Y. Eliashberg au congrès de Berlin). Il nous est difficile, voire impossible, d'entrer dans les détails de ce projet.

Y. Eliashberg identifie un voisinage tubulaire de L avec un tube T_r^* du cotangent T^*L , puis $T^*L \setminus L$ avec la symplectification $(\mathbb{R} \times V, d(e^\tau \alpha))$ de $V = \partial T_r^*$, où α est la structure de contact de ∂T_r^* . Ensuite il déforme la structure symplectique Ω et la structure presque complexe J . Les variétés déformées (M^t, J^t, Ω^t) , où Ω^t n'est définie qu'à un facteur constant près, peuvent être vues soit comme le résultat du recollement de (T_r^*, Ω) , $([0, t] \times V, d(e^\tau \alpha))$ et $(M \setminus T_{<r}^*, e^t \Omega)$, soit comme celui de $(T_r^*, e^{-t} \Omega)$, $([-t, 0] \times V, d(e^\tau \alpha))$ et $(M \setminus T_{<r}^*, \Omega)$. La structure presque complexe sur T_r^* et $M \setminus T_{<r}^*$ reste la même et est choisie

invariante par translation sur $[0, t] \times V$ (respectivement $[-t, 0] \times V$). Lorsque $t \rightarrow +\infty$, à la limite, la variété M se déchire en T^*L et $M \setminus L$, les bouts de ces variétés étant équipés de la forme symplectique homogène $d(e^\tau \alpha)$ ($\tau \rightarrow +\infty$ dans T^*L , tandis que dans M , $\tau \rightarrow -\infty$). Il s'agit alors de comprendre ce que deviennent les courbes holomorphes dans la variété déchirée, c'est-à-dire d'obtenir une version appropriée du théorème de compacité de Gromov. N'utilisant pas de structure réelle globale, l'avantage de cet argument serait de donner des résultats sur toutes les sous-variétés lagrangiennes. L'inconvénient est que cette méthode ne semble pas permettre de construire des courbes réelles en présence d'une structure réelle.

BIBLIOGRAPHIE

- [1] S. AKBULUT, H. KING – *Rational structures on 3-manifolds*. Pacific J. Math., **151** (1991), 201–204.
- [2] H. ALEXANDER – *Continuing 1-dimensional analytic sets*. Math. Ann. **191** (1971), 143–144.
- [3] V. I. ARNOL'D – *On the arrangements of the ovals of real plane curves, involutions of 4-dimensional manifolds, and the arithmetic of integral quadratic forms*. Functional Anal. Appl. **5** (1971), 169–176.
- [4] R. BENEDETTI, A. MARIN – *Déchirures de variétés de dimension 3*. Comm. Math. Helv. **67** (1992), 514–545.
- [5] J. BOCHNAK, M. COSTE, M.-F. ROY – *Géométrie algébrique réelle*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 12 ; Springer-Verlag, 1987.
- [6] A. BOREL, J.-P. SERRE – *Théorèmes de finitude en cohomologie galoisienne*. Comm. Math. Helv. **39** (1964), 111–164.
- [7] D. BURNS – *On the uniqueness and characterization of Grauert tubes*. In *Complex analysis and geometry (Trento, 1993)* Lecture Notes in Pure and Appl. Math. **173**, 119–133.
- [8] F. CAMPANA – *Remarques sur le revêtement universel des variétés kählériennes compactes*. Bull. Soc. Math. France **122** (1994), n. 2, 255–284.
- [9] A. COMESSATTI – *Sulla connessione delle superficie razionali reali*. Annali di Math. **23** (1914), 215–283.
- [10] C. CONLEY, E. ZEHNDER – *Morse type index theory for flows and periodic solutions for Hamiltonian equations*. Comm. Pure Appl. Math. **37** (1984), 207–253.
- [11] O. DEBARRE – *Variétés de Fano*. Sémin. Bourbaki 1996/97, exp. n° 827, Astérisque **245** (1997), 197–221.

- [12] O. DEBARRE, L. MANIVEL – *Sur la variété des espaces linéaires contenus dans une intersection complète*. Math. Ann. **312** (1998), n. 3, 549–574.
- [13] A. DEGTYAREV, I. ITENBERG, V. KHARLAMOV – *Real Enriques surfaces*. Submitted to Lecture Notes Math.
- [14] A. DEGTYAREV, V. KHARLAMOV – *Topological properties of real algebraic varieties: du côté de chez Rokhlin*. A paraître dans Uspekhi Mat. Nauk.
- [15] J. J. DUISTERMAAT – *On the Morse index in variational calculus*. Adv. Math. **21** (1976), 173–195.
- [16] A. L. EDMONDS – *Orientability of fixed point sets*. Proc. Amer. Math. Soc. **82** (1981), n. 1, 120–124.
- [17] A. FLOER, H. HOFER, – *Coherent orientations for periodic orbit problems in symplectic geometry*. Math. Z. **212** (1993), n. 1, 13–38.
- [18] H. GRAUERT – *On Levi's problem and the imbedding of real-analytic manifolds*. Annals of Math. **68** (1958), n. 3, 460–472.
- [19] H. HOFER, C. VITERBO – *The Weinstein Conjecture in the Presence of Holomorphic Spheres*. Commun. Pure Appl. Math. **45** (1992), 582–622.
- [20] N. V. IVANOV – *Approximation of smooth manifolds by real algebraic sets*. Uspekhi Mat. Nauk **37** (1982), n. 1, 3–52.
- [21] V. M. KHARLAMOV – *Topological types of nonsingular surfaces of degree 4 in $\mathbb{R}P^3$* . Funkcional. Anal. i Priložen. **10** (1976), n. 4, 55–68.
- [22] S. KOBAYASHI, T. OCHIAI – *On complex manifolds with positive tangent bundles*. J. Math. Soc. Japan **22** (1970), 499–525.
- [23] J. KOLLÁR – *Real algebraic threefolds. II. Minimal Model Program*. J. Amer. Math. Soc. **12** (1999), 33–83.
- [24] J. KOLLÁR – *Non-projective Nash conjecture*. Exposé à Paris XI, juin 1999.
- [25] J. KOLLÁR – *Rationally connected varieties over local fields*. Ann. of Math. **150** (1999), 357–367.
- [26] V. A. KRASNOV – *Characteristic classes of vector bundles on a real algebraic variety*. Izv. Akad. Nauk SSSR Ser. Mat. **55** (1991), n. 4, 716–746.
- [27] S. B. MYERS – *Riemannian manifolds with positive mean curvature*. Duke Math. J. **8** (1941), 401–404.
- [28] G. MIKHALKIN – *Blowup equivalence of smooth closed manifolds*. Topology **36** (1997), 287–299.
- [29] J. NASH – *Real algebraic manifolds*. Annals of Math. **56** (1952), n. 3, 405–421.
- [30] N. Yu. NETSVETAEV – *An analogue of the Maslov index*. Journal of Math. Sci. **81** (1996), n. 2, 2535–2537.
- [31] S. Yu. OREVKOV – *Link theory and oval arrangements of real algebraic curves*. Topology **38** (1999), n. 4, 779–810.

- [32] V. A. ROKHLIN – *Proof of the Gudkov conjecture*. Funkts. Anal. i Prilozhen **6** (1972), n. 2, 62–64.
- [33] B. SHIFFMAN – *On the Continuation of Analytic Sets*. Math. Ann. **185** (1970), 1–12.
- [34] R. SILHOL – *Real algebraic surfaces*. Lect. Notes Math. **1392** (1989).
- [35] A. J. SOMMESE – *Real algebraic spaces*. Ann. Scuola Norm. Sup. Pisa **4** (1977), n. 4, 599–612.
- [36] A. TOGNOLI – *Quelques exemples en géométrie algébrique réelle*. In *Séminaire sur la géométrie algébrique réelle*, Publ. Math. Univ. Paris VII, **24** (1986), 29–34.
- [37] O. Ya. VIRO – *Achievements in the topology of real algebraic varieties in the last 6 years*. Uspekhi Mat. Nauk **41** (1986), n. 3, 45–67.
- [38] C. VITERBO – *A new obstruction to embedding Lagrangian tori*. Invent. Math. **100** (1990), n. 2, 301–320.
- [39] C. VITERBO – *Symplectic real algebraic geometry*. à paraître.
- [40] J. A. WIŚNIEWSKI – *On a conjecture of Mukai*. Manuscripta Math. **68** (1990), 135–141.
- [41] G. WILSON – *Hilbert’s sixteenth problem*. Topology **17** (1978), n. 1, 53–73.
- [42] J. Y. WELSCHINGER – *J-courbes réelles à nids profonds sur les surfaces réglées*. Prépublication d’IRMA-ULP, à paraître dans les oeuvres du colloque de Rokhlin, 1999, St. Petersburg.

Viatcheslav KHARLAMOV

Université Louis Pasteur

I.R.M.A.

7, rue René Descartes

67084 STRASBOURG Cedex

E-mail : kharlam@irma.u-strasbg.fr

LA CORRESPONDANCE DE LANGLANDS SUR LES CORPS DE FONCTIONS

[d'après Laurent LAFFORGUE]

par Gérard LAUMON

Lafforgue a récemment établi la correspondance de Langlands pour GL_r sur un corps de fonctions. Sa preuve suit la stratégie introduite, il y a plus de 25 ans, par Drinfeld pour traiter le cas $r = 2$. Une des innovations principales est la construction d'une compactification toroïdale du schéma simplicial $PGL_r^{\bullet+1} / PGL_r$ classifiant de PGL_r qui prolonge la compactification de De Concini et Procesi de PGL_r^2 / PGL_r .

Après avoir énoncé le théorème principal et ses conséquences, nous présenterons les grandes lignes de la démonstration, en renvoyant pour les détails aux publications de Lafforgue ([La 1] à [La 10]). Les sections 4 (Compactification du classifiant de PGL_r) et 7 (Une variante d'un théorème de Pink) sont de nature générale et peuvent être lues indépendamment du reste du texte.

Je remercie chaleureusement L. Lafforgue pour son aide dans la préparation de cet exposé.

1. ÉNONCÉ DU THÉORÈME PRINCIPAL

On fixe dans tout cet exposé une courbe X projective, lisse et géométriquement connexe sur un corps fini \mathbb{F}_q à q éléments. On note $|X|$ l'ensemble des points fermés de X .

Soit F le corps des fonctions de X . On identifie les places de F aux éléments de $|X|$. Pour chaque $x \in |X|$ on peut former le complété F_x de F en la place x . C'est un corps de valuation discrète complet. On notera encore $x : F_x^\times \rightarrow \mathbb{Z}$ sa valuation. L'anneau des entiers de F_x est l'anneau de valuation discrète $\mathcal{O}_x = \{a_x \in F_x^\times \mid x(a_x) \geq 0\} \cup \{0\}$, $\mathfrak{p}_x = \{a_x \in F_x^\times \mid x(a_x) > 0\} \cup \{0\}$ est l'unique idéal maximal de \mathcal{O}_x et le corps résiduel $\kappa(x) = \mathcal{O}_x / \mathfrak{p}_x$ est une extension finie de \mathbb{F}_q dont on notera $\deg(x)$ le degré.

L'anneau (topologique) des adèles de F est le produit restreint

$$\mathbb{A} = \{a = (a_x)_{x \in |X|} \mid a_x \in \mathcal{O}_x \text{ pour presque tout } x\} \subset \prod_{x \in |X|} F_x$$

où l'expression «pour presque tout x » signifie «pour tous les x sauf un nombre fini». Le corps F se plonge diagonalement dans \mathbb{A} et l'anneau topologique compact $\mathcal{O} := \prod_{x \in |X|} \mathcal{O}_x$ est un sous-anneau de \mathbb{A} .

On dispose d'un homomorphisme de groupes $\text{deg} : \mathbb{A}^\times \rightarrow \mathbb{Z}$ défini par

$$\text{deg}(a) = \sum_{x \in |X|} \text{deg}(x)x(a_x).$$

Cet homomorphisme est surjectif et est identiquement nul sur F^\times et aussi sur \mathcal{O}^\times . Son noyau est compact modulo F^\times : en d'autres termes, pour tout $a \in \mathbb{A}^\times$ de degré non nul, le groupe quotient $F^\times \backslash \mathbb{A}^\times / \mathcal{O}^\times a^\mathbb{Z}$ est fini.

1.1. Représentations automorphes cuspidales

Soit r un entier ≥ 1 . On considère le groupe adélique $\text{GL}_r(\mathbb{A})$ des matrices inversibles de taille $r \times r$ et à coefficients dans l'anneau \mathbb{A} , et ses sous-groupes $\text{GL}_r(F)$ et

$$K = \prod_{x \in |X|} K_x = \text{GL}_r(\mathcal{O}) = \prod_{x \in |X|} \text{GL}_r(\mathcal{O}_x).$$

On fixe une fois pour toute une mesure de Haar dg sur $\text{GL}_r(\mathbb{A})$, et une décomposition $dg = \prod_{x \in |X|} dg_x$ de cette mesure en produit de mesures de Haar locales. Il est commode de normaliser dg et les dg_x par les conditions $\text{vol}(K, dg) = 1$ et $\text{vol}(K_x, dg_x) = 1$.

Une *forme automorphe cuspidale* (pour GL_r sur F) est une fonction $\varphi : \text{GL}_r(\mathbb{A}) \rightarrow \mathbb{C}$ ayant les propriétés suivantes :

- 1) $\varphi(\gamma g) = \varphi(g)$, $\forall \gamma \in \text{GL}_r(F)$, $\forall g \in \text{GL}_r(\mathbb{A})$,
- 2) il existe un sous-groupe $K_\varphi \subset K$ d'indice fini tel que $\varphi(gk) = \varphi(g)$, $\forall g \in \text{GL}_r(\mathbb{A})$, $\forall k \in K_\varphi$,
- 3) il existe $a \in \mathbb{A}^\times$ tel que $\text{deg}(a) \neq 0$ et $\varphi(ga) = \varphi(g)$, $\forall g \in \text{GL}_r(\mathbb{A})$,
- 4) pour toute décomposition non triviale $r = r_1 + \dots + r_s$ de r en entiers strictement positifs, qui définit un sous-groupe parabolique standard $P = MU \subsetneq \text{GL}_r$ de radical unipotent U et de composante de Levi $M \cong \text{GL}_{r_1} \times \dots \times \text{GL}_{r_s}$, le *terme constant*

$$\text{GL}_r(\mathbb{A}) \rightarrow \mathbb{C}, g \mapsto \int_{U(F) \backslash U(\mathbb{A})} \varphi(ug) du,$$

est identiquement nul (ici du est n'importe quelle mesure de Haar sur le quotient compact $U(F) \backslash U(\mathbb{A})$).

Si l'on fixe un élément a_0 de degré non nul dans \mathbb{A}^\times , on a

$$L_{\text{cusp}} = \bigcup_{n \geq 1} L_{\text{cusp}}(a_0^n)$$

où, pour tout $a \in \mathbb{A}^\times$ de degré non nul, $L_{\text{cusp}}(a) \subset L_{\text{cusp}}$ est le sous-espace défini en imposant ce a particulier dans la propriété 3).

Toujours pour $a \in \mathbb{A}^\times$ de degré non nul, on montre que toute fonction $\varphi \in L_{\text{cusp}}(a)$ est à support compact sur $\text{GL}_r(F) \backslash \text{GL}_r(\mathbb{A})/a^\mathbb{Z}$, et on peut donc munir $L_{\text{cusp}}(a)$ du produit scalaire hermitien défini positif

$$(\varphi_1, \varphi_2) := \int_{\text{GL}_r(F) \backslash \text{GL}_r(\mathbb{A})/a^\mathbb{Z}} \overline{\varphi_1(g)} \varphi_2(g) dg.$$

Le groupe $\text{GL}_r(\mathbb{A})$ agit par translation à droite sur l'espace vectoriel complexe

$$L_{\text{cusp}} = L_{\text{cusp}}(\text{GL}_r(F) \backslash \text{GL}_r(\mathbb{A}))$$

des formes automorphes cuspidales. Cette représentation est *lisse* (le fixateur de tout vecteur dans L_{cusp} est un sous-groupe ouvert de $\text{GL}_r(\mathbb{A})$). Comme on a fixé une mesure de Haar dg sur $\text{GL}_r(\mathbb{A})$, la donnée de cette représentation équivaut à celle d'une structure de \mathcal{H} -module sur L_{cusp} , où l'*algèbre de Hecke*

$$\mathcal{H} = \mathcal{C}_c^\infty(\text{GL}_r(\mathbb{A}))$$

est l'algèbre de convolution des fonctions complexes, localement constantes et à support compact, sur $\text{GL}_r(\mathbb{A})$.

Pour chaque $a \in \mathbb{A}^\times$ de degré non nul, l'action de $\text{GL}_r(\mathbb{A})$ respecte le sous-espace $L_{\text{cusp}}(a)$. La représentation induite sur ce sous-espace est *admissible* (pour tout sous-groupe $K' \subset K$ d'indice fini, l'espace vectoriel des invariants sous K' dans $L_{\text{cusp}}(a)$ est de dimension finie). Elle est de plus *unitaire* pour le produit scalaire défini ci-dessus. La représentation de $\text{GL}_r(\mathbb{A})$ sur L_{cusp} est donc semi-simple : elle admet une décomposition isotypique

$$L_{\text{cusp}} \cong \bigoplus_{\pi} V_{\pi}^{\oplus m(\pi)}$$

où π parcourt un système de représentants des classes d'isomorphie de représentations complexes irréductibles admissibles de $\text{GL}_r(\mathbb{A})$, où V_{π} est l'espace de π (en général de dimension infinie) et où les multiplicités $m(\pi)$ sont des entiers ≥ 0 .

DÉFINITION. — Une représentation automorphe cuspidale irréductible (pour GL_r sur F) est une représentation complexe admissible irréductible de $\text{GL}_r(\mathbb{A})$ qui est isomorphe à un facteur direct de L_{cusp} .

On notera \mathcal{A}_r un système de représentants des classes d'isomorphie de ces représentations automorphes cuspidales irréductibles. Chaque $\pi \in \mathcal{A}_r$ admet un caractère central $\omega_{\pi} : F^\times \backslash \mathbb{A}^\times \rightarrow \mathbb{C}^\times$ qui est d'ordre fini puisque $\omega_{\pi}(a) = 1$ pour au moins un $a \in \mathbb{A}^\times$ de degré non nul.

THÉORÈME (Théorème de multiplicité un, [PS 2], [Sh]). — On a

$$L_{\text{cusp}} \cong \bigoplus_{\pi \in \mathcal{A}_r} V_{\pi}.$$

En d'autres termes, les multiplicités dans la décomposition isotypique de la représentation L_{cuspidal} de $\text{GL}_r(\mathbb{A})$ sont données par $m(\pi) = 1$ si π est automorphe cuspidale et $m(\pi) = 0$ sinon.

Pour tout $x \in |X|$, on définit aussi l'algèbre de Hecke locale

$$\mathcal{H}_x = \mathcal{C}_c^\infty(\text{GL}_r(F_x)).$$

C'est l'algèbre de convolution pour la mesure de Haar dg_x sur $\text{GL}_r(F_x)$. L'espace de toute représentation lisse de $\text{GL}_r(F_x)$ est naturellement muni d'une structure de \mathcal{H}_x -module.

On notera $e_{K_x} \in \mathcal{H}_x$ la fonction caractéristique de $K_x = \text{GL}_r(\mathcal{O}_x)$ dans $\text{GL}_r(F_x)$ et on appellera algèbre de Hecke locale *non ramifiée* la sous-algèbre

$$\mathcal{H}_x(K_x) = e_{K_x} * \mathcal{H}_x * e_{K_x} \subset \mathcal{H}_x.$$

C'est une algèbre commutative unitaire munie d'un isomorphisme, construit par Satake,

$$\mathcal{H}_x(K_x) \cong \mathbb{C}[z_1, z_1^{-1}, \dots, z_r, z_r^{-1}]^{\mathfrak{S}_r},$$

où le groupe symétrique \mathfrak{S}_r agit par permutation des indéterminées z_1, \dots, z_r . En particulier, si (π_x, V_{π_x}) est une représentation admissible irréductible de $\text{GL}_r(F_x)$ *non ramifiée*, c'est-à-dire telle que $V_{\pi_x}^{K_x} \neq (0)$, on a en fait

$$\dim(V_{\pi_x}^{K_x}) = 1$$

et le $\mathcal{H}_x(K_x)$ -module $V_{\pi_x}^{K_x}$ de rang 1 et aussi la représentation (π_x, V_{π_x}) sont (à isomorphisme près) uniquement déterminés par la donnée d'un r -uplet non ordonné

$$(z_1(\pi_x), \dots, z_r(\pi_x))$$

de nombres complexes non nuls, appelés les *valeurs propres de Hecke* de π_x .

L'algèbre de Hecke globale \mathcal{H} est le produit tensoriel restreint

$$\mathcal{H} = \varinjlim_S \left(\bigotimes_{x \in S} \mathcal{H}_x \right) \otimes \left(\bigotimes_{x \notin S} e_{K_x} \right)$$

des algèbres de Hecke locales, où S parcourt l'ensemble des parties finies de $|X|$. Une représentation admissible irréductible (π, V_π) de $\text{GL}_r(\mathbb{A})$ admet donc, pour chaque $x \in |X|$, une composante locale (π_x, V_{π_x}) qui est une représentation irréductible admissible de $\text{GL}_r(F_x)$ bien définie à isomorphisme près. Pour presque tout x , π est non ramifiée en x , c'est-à-dire admet une composante locale non ramifiée en x , et π est le produit tensoriel restreint

$$\pi \cong \bigotimes_{x \in |X|}' \pi_x$$

de ses composantes locales. (Une fois que l'on a fixé une base v_x de $V_{\pi_x}^{K_x}$ pour chaque place x non ramifiée pour π , l'espace de la représentation du membre de droite est par définition la limite inductive

$$\varinjlim_S \left(\bigotimes_{x \in S} V_{\pi_x} \right) \otimes \left(\bigotimes_{x \notin S} v_x \right)$$

pour S parcourant l'ensemble des parties finies de $|X|$ qui contiennent toutes les places ramifiées pour π .)

On définit la *fonction L partielle* de π comme le produit eulérien formel

$$L^{S_\pi}(\pi, T) = \prod_{x \notin S_\pi} L(\pi_x, T) = \prod_{x \notin S_\pi} \frac{1}{\prod_{i=1}^r (1 - z_i(\pi_x) T^{\deg(x)})} \in \mathbb{C}[[T]]$$

où $S_\pi \subset |X|$ est l'ensemble fini des places ramifiées de π . Godement et Jacquet ([G-J]) ont démontré que le produit eulérien $L^{S_\pi}(\pi, q^{-s})$ converge absolument dans un demi-plan $\operatorname{Re}(s) > \sigma$ pour un nombre réel σ assez grand, et que la série formelle en T définie par $L^{S_\pi}(\pi, T)$ est en fait le développement d'une fraction rationnelle dans $\mathbb{C}(T)$, et même d'un polynôme dans $\mathbb{C}[T]$ si $r \geq 2$.

1.2. Représentations du groupe de Galois

On fixe une clôture séparable \overline{F} de F et on note Γ_F le groupe de Galois de \overline{F} sur F . Pour chaque $x \in |X|$ on choisit arbitrairement un sous-groupe de décomposition $D_x \subset \Gamma_F$ en x . Ce groupe de décomposition est le groupe de Galois Γ_{F_x} d'une certaine clôture séparable \overline{F}_x de F_x . On note $I_x \subset D_x$ le sous-groupe d'inertie de ce groupe de décomposition. Le groupe quotient D_x/I_x est isomorphe au groupe de Galois $\Gamma_{\kappa(x)}$ de $\overline{\kappa(x)}$ sur $\kappa(x)$ pour une certaine clôture algébrique $\overline{\kappa(x)}$ de $\kappa(x)$ et est donc isomorphe à $\widehat{\operatorname{Frob}}_x$ où $\operatorname{Frob}_x \in \Gamma_{\kappa(x)}$ est l'élément de Frobenius géométrique (l'inverse de l'élevation à la puissance $|\kappa(x)| = q^{\deg(x)}$).

Muni de la topologie de Krull, Γ_F est un groupe topologique pro-fini, dont les représentations complexes continues et de dimension finie se factorisent nécessairement par un quotient fini. Suivant Serre et Grothendieck, on obtient une catégorie plus vaste de représentations de Γ_F en remplaçant le corps des coefficients \mathbb{C} par un corps ℓ -adique. Fixons donc un nombre premier auxiliaire ℓ , distinct de la caractéristique de \mathbb{F}_q , et fixons une clôture algébrique $\overline{\mathbb{Q}}_\ell$ de \mathbb{Q}_ℓ .

Une *représentation ℓ -adique* σ de Γ_F est un $\overline{\mathbb{Q}}_\ell$ -espace vectoriel V_σ de dimension finie, muni d'un homomorphisme de groupes $\sigma : \Gamma_F \rightarrow \operatorname{Aut}_{\overline{\mathbb{Q}}_\ell}(V_\sigma)$, ayant les propriétés suivantes :

- 1) il existe une base de V_σ identifiant $\operatorname{Aut}_{\overline{\mathbb{Q}}_\ell}(V_\sigma)$ à $\operatorname{GL}_r(\overline{\mathbb{Q}}_\ell)$, où r est la dimension de V_σ , et une extension finie E_λ de \mathbb{Q}_ℓ contenue dans $\overline{\mathbb{Q}}_\ell$ telles que $\sigma(\Gamma_F) \subset \operatorname{GL}_r(E_\lambda) \subset \operatorname{GL}_r(\overline{\mathbb{Q}}_\ell)$,
- 2) $\sigma : \Gamma_F \rightarrow \operatorname{GL}_r(E_\lambda)$ est continu pour la topologie de Krull sur Γ_F et la topologie ℓ -adique sur $\operatorname{GL}_r(E_\lambda)$,

- 3) pour presque tout $x \in |X|$, σ est *non ramifiée* en x , c'est-à-dire que la restriction σ_x de σ au sous-groupe de décomposition $D_x \subset \Gamma_F$ est triviale sur le sous-groupe d'inertie $I_x \subset D_x$.

Ces représentations forment une catégorie abélienne où tout objet est de longueur finie.

Soit \mathcal{G}_r un système de représentants des classes d'isomorphie de représentations ℓ -adiques irréductibles de rang r de Γ_F dont le déterminant (la puissance extérieure maximale) est d'ordre fini. Pour chaque $\sigma \in \mathcal{G}_r$, on note S_σ l'ensemble fini des places de F où σ est ramifiée. Pour chaque $x \notin S_\sigma$, on dispose de l'automorphisme $\sigma_x(\text{Frob}_x)$ de V_σ , et on note

$$(z_1(\sigma_x), \dots, z_r(\sigma_x))$$

les *valeurs propres de Frobenius* de σ en x , c'est-à-dire le r -uplet non ordonné d'éléments non nuls de $\overline{\mathbb{Q}}_\ell$ formé des valeurs propres de $\sigma_x(\text{Frob}_x)$. La *fonction L partielle* de σ est par définition le produit eulérien formel

$$L^{S_\sigma}(\sigma, T) = \prod_{x \in |X| - S_\sigma} L(\sigma_x, T) = \prod_{x \in |X| - S_\sigma} \frac{1}{\prod_{i=1}^r (1 - z_i(\sigma_x) T^{\deg(x)})} \in \overline{\mathbb{Q}}_\ell[[T]].$$

Il résulte de la formule des points fixes de Grothendieck-Lefschetz ([Gr]) que cette série formelle est le développement d'une fraction rationnelle dans $\overline{\mathbb{Q}}_\ell(T)$, et même d'un polynôme dans $\overline{\mathbb{Q}}_\ell[T]$ si $r \geq 2$.

1.3. La correspondance de Langlands et ses conséquences

On fixe un isomorphisme de corps $\iota : \overline{\mathbb{Q}}_\ell \xrightarrow{\sim} \mathbb{C}$ (il en existe d'après l'axiome du choix). On notera encore ι les isomorphismes de $\overline{\mathbb{Q}}_\ell(T)$ (resp. $\overline{\mathbb{Q}}_\ell[[T]]$) sur $\mathbb{C}(T)$ (resp. $\mathbb{C}[[T]]$) induits par ι .

THÉORÈME PRINCIPAL. — (i) (Correspondance de Langlands) *Il existe une unique bijection*

$$\mathcal{A}_r \xrightarrow{\sim} \mathcal{G}_r, \quad \pi \mapsto \sigma(\pi),$$

telle que, pour tout $\pi \in \mathcal{A}_r$ on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma(\pi)_x, T)) = L(\pi_x, T),$$

pour presque tout $x \notin S_{\sigma(\pi)} \cup S_\pi$.

(ii) (Conjecture de Ramanujan-Petersson) *Pour tout $\pi \in \mathcal{A}_r$ et toute place $x \notin S_\pi$ on a*

$$|z_i(\pi_x)| = 1, \quad \forall i = 1, \dots, r.$$

Le cas $r = 1$ de ce théorème est une reformulation de la théorie du corps de classes abélien pour les corps de fonctions. Le cas $r = 2$ a été démontré par Drinfeld ([Dr 6], [Dr 7]). Le cas général est dû à Lafforgue ([La 10]).

Les conséquences de ce théorème qui sont formulées ci-dessous étaient attendues, et leur déduction du théorème principal est bien connue.

THÉORÈME (Compatibilité avec la correspondance de Langlands locale). — *Pour tout $\pi \in \mathcal{A}_r$ et tout $x \in |X|$, la représentation de Galois locale $\sigma(\pi)_x$ de $D_x = \Gamma_{F_x}$ est l'image $\sigma_x(\pi_x)$ de la représentation locale π_x de $\mathrm{GL}_r(F_x)$ par la correspondance de Langlands locale (cf. [L-R-S]). En particulier, π et $\sigma(\pi)$ ont les mêmes facteurs L locaux en toutes les places de F .*

THÉORÈME (Conjecturé par Deligne, [De](1.2.10)). — *Soit $\sigma \in \mathcal{G}_r$.*

(i) *Le sous-corps $E = E(\sigma)$ de $\overline{\mathbb{Q}}_\ell$ engendré sur \mathbb{Q} par les coefficients des polynômes $\prod_{i=1}^r (1 - z_i(\sigma_x)T)$ pour tous les $x \notin S_\sigma$ est un corps de nombres (une extension finie de \mathbb{Q}).*

(ii) *Pour tout nombre premier ℓ' distinct de la caractéristique de \mathbb{F}_q , fixons une clôture algébrique $\overline{\mathbb{Q}}_{\ell'}$ de $\mathbb{Q}_{\ell'}$ et notons $\mathcal{G}_{r,\ell'}$ l'ensemble \mathcal{G}_r correspondant.*

Il existe alors une unique famille de représentations $\sigma_{\ell',\lambda'} \in \mathcal{G}_{r,\ell'}$, indexée par les couples (ℓ', λ') formés d'un nombre premier ℓ' distinct de la caractéristique de \mathbb{F}_q et d'un plongement $\lambda' : E \hookrightarrow \overline{\mathbb{Q}}_{\ell'}$, et ayant les propriétés suivantes :

- $\sigma = \sigma_{\ell,\lambda}$ où λ est l'inclusion de E dans $\overline{\mathbb{Q}}_\ell$,
- chaque $\sigma_{\ell',\lambda'}$ a le même ensemble de places ramifiées que σ et pour tout $x \notin S_\sigma$ on a l'égalité

$$\prod_{i=1}^r (1 - z_i(\sigma_{\ell',\lambda',x})T) = \lambda' \left(\prod_{i=1}^r (1 - z_i(\sigma_x)T) \right) \in \overline{\mathbb{Q}}_{\ell'}[T].$$

□

THÉORÈME (Conjecturé par Deligne, [De](1.2.10)). — *Soient Y un schéma normal de type fini sur le corps fini \mathbb{F}_q et \mathcal{L} un système local ℓ -adique de rang r sur Y . On suppose que \mathcal{L} est irréductible et que sa puissance extérieure maximale $\det(\mathcal{L})$ est d'ordre fini ($\det(\mathcal{L})^{\otimes N}$ est isomorphe au faisceau ℓ -adique constant de valeur $\overline{\mathbb{Q}}_\ell$ sur Y pour tout entier $N \geq 1$ assez divisible). Alors \mathcal{L} est pur de poids 0.* □

Une conséquence plus indirecte du théorème principal est l'énoncé dit «de descente» dans la première construction de Drinfeld (cf. [Dr 12], [F-G-K-V], [Lau 5], [Lau 6]) des faisceaux automorphes partout non ramifiés pour GL_r sur une courbe lisse, projective et géométriquement connexe sur un corps arbitraire (éventuellement de caractéristique nulle).

1.4. Historique

L'outil principal inventé par Drinfeld et utilisé par Drinfeld et Lafforgue pour démontrer le théorème principal, est le champ modulaire des chtoucas. Avant de découvrir ce champ, Drinfeld avait introduit des variétés modulaires sur les corps de fonctions très semblables aux variétés de Shimura sur les corps de nombres : les variétés de *modules elliptiques* (ou de *faisceaux elliptiques*). L'utilisation de ces variétés avait permis de construire $\sigma(\pi)$ sous certaines conditions sur la représentation locale π_∞ en une place donnée $\infty \in |X|$. Ainsi, Drinfeld ([Dr 9] et [Dr 10]) avait construit $\sigma(\pi)$

quand $r = 2$ et π_∞ est dans la série discrète, et Flicker et Kazhdan, puis moi-même, avons généralisé cette construction de Drinfeld, pour r arbitraire, dans les cas où π_∞ est soit supercuspidale ([F-K]), soit la représentation de Steinberg ([Lau 1], [Lau 2]).

L'assertion d'unicité du théorème principal était connue depuis longtemps : elle est en effet une conséquence immédiate du théorème de densité de Čebotarev ([Se]). Il en était de même de l'injectivité de l'application $\pi \rightarrow \sigma(\pi)$ qui est automatique d'après le théorème de multiplicité un fort de Piatetski-Shapiro ([PS 2]).

Deligne avait remarqué un principe de récurrence permettant de déduire la surjectivité de l'application $\pi \rightarrow \sigma(\pi)$ de son existence. Plus précisément, si pour tout $r' = 1, \dots, r-1$, on a construit une application $\pi' \rightarrow \sigma'(\pi')$ de $\mathcal{A}_{r'} \rightarrow \mathcal{G}_{r'}$ telle que, quel que soit $\pi' \in \mathcal{A}_{r'}$, on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma'(\pi')_x, T)) = L(\pi'_x, T)$$

pour presque tout $x \notin S_{\sigma'(\pi')} \cup S_{\pi'}$, alors l'équation fonctionnelle de Grothendieck ([Gr]), la formule du produit pour la constante de cette équation fonctionnelle ([Lau 7]) et le théorème inverse de Hecke, Weil et Piatetski-Shapiro ([PS 1], [C-PS]) permettent de définir une application

$$\mathcal{G}_r \rightarrow \mathcal{A}_r, \sigma \mapsto \pi(\sigma),$$

telle que, quel que soit $\sigma \in \mathcal{G}_r$, on ait l'égalité de facteurs L locaux

$$\iota(L(\pi(\sigma)_x, T)) = L(\sigma_x, T)$$

pour presque tout $x \notin S_\sigma \cup S_{\pi(\sigma)}$.

Lafforgue avait obtenu auparavant un énoncé très proche de la conjecture de Ramanujan-Petersson ([La 1]) comme une conséquence de la description adélique de Drinfeld des chtoucas sur les corps finis ([Dr 6]), de la formule des traces d'Arthur-Selberg ([Ar]), de la formule des points fixes de Grothendieck-Lefschetz ([Gr]), du théorème de pureté de Deligne ([De]) et des estimées de Jacquet-Shalika pour les valeurs propres de Hecke des représentations automorphes cuspidales irréductibles ([J-S 1]).

1.5. La stratégie

La stratégie utilisée par Drinfeld et Lafforgue pour définir l'application $\pi \rightarrow \sigma(\pi)$ de \mathcal{A}_r dans \mathcal{G}_r est inspirée des travaux de Shimura, Ihara, Deligne, Langlands, ... Sous sa forme la plus naïve elle peut se décrire comme suit :

On construit un schéma V sur F muni d'une action de l'algèbre de Hecke \mathcal{H} de telle sorte que la cohomologie ℓ -adique à supports compacts

$$H_c^*(\overline{F} \otimes_F V, \overline{\mathbb{Q}}_\ell)$$

soit une représentation du produit de $\mathrm{GL}_r(\mathbb{A})$ et du groupe de Galois Γ_F .

Puis on calcule la trace de cette représentation par la formule des points fixes de Grothendieck-Lefschetz.

Enfin on compare cette formule des points fixes avec la formule des traces d'Arthur-Selberg pour prouver que la représentation

$$\bigoplus_{\pi \in \mathcal{A}_r} \pi \otimes \sigma(\pi)$$

de $\mathrm{GL}_r(\mathbb{A}) \times \Gamma_F$ que l'on cherche est exactement la partie « cuspidale » de $H_c^*(\overline{F} \otimes_F V, \overline{\mathbb{Q}}_\ell)$.

En fait, comme la cohomologie ℓ -adique ci-dessus est automatiquement définie sur \mathbb{Q}_ℓ , il y a une obstruction (de rationalité) à mener à bien un tel programme (cf. [Ka]) et la stratégie doit être légèrement modifiée. Comme l'a proposé Drinfeld, c'est plutôt la représentation

$$\bigoplus_{\pi \in \mathcal{A}_r} \pi \otimes \sigma(\pi)^\vee \otimes \sigma(\pi)$$

du produit $\mathrm{GL}_r(\mathbb{A}) \times \Gamma_F \times \Gamma_F$, où $\sigma(\pi)^\vee$ est la représentation contragrédiente de $\sigma(\pi)$, que l'on peut espérer obtenir comme la partie cuspidale de la cohomologie ℓ -adique à supports compacts d'un schéma sur $F \otimes F$.

2. CHTOUCAS DE DRINFELD

2.1. Le champ des chtoucas

Tous les schémas (ou champs) considérés seront sur \mathbb{F}_q ; on dira donc schéma (ou champ) au lieu de \mathbb{F}_q -schéma (ou \mathbb{F}_q -champ) et on notera simplement $U \times T$ le produit $U \times_{\mathbb{F}_q} T$ de deux tels schémas (ou champs). Pour tout schéma (ou champ) U , on notera $\mathrm{Frob}_U : U \rightarrow U$ son endomorphisme de Frobenius relativement à \mathbb{F}_q : si U est un schéma, Frob_U est donc l'identité sur l'espace topologique sous-jacent à U et l'élévation à la puissance q -ième sur le faisceau structural \mathcal{O}_U . Si \mathcal{M} est un $\mathcal{O}_{U \times X}$ -Module, on notera ${}^\tau \mathcal{M}$ le $\mathcal{O}_{U \times X}$ -Module $(\mathrm{Frob}_U \times \mathrm{Id}_X)^* \mathcal{M}$.

DÉFINITION (Drinfeld). — *Un chtouca à droite (resp. à gauche) $\tilde{\mathcal{E}}$ de rang $r \geq 1$ sur un schéma U est un diagramme dans la catégorie abélienne des $\mathcal{O}_{U \times X}$ -Modules*

$$\mathcal{E} \xrightarrow{j} \mathcal{E}' \xleftarrow{t} {}^\tau \mathcal{E} \quad (\text{resp. } \mathcal{E} \xleftarrow{t} \mathcal{E}' \xrightarrow{j} {}^\tau \mathcal{E})$$

où :

- \mathcal{E} et \mathcal{E}' sont localement libres de rang r ,
- j et t sont injectifs,
- les conoyaux de j et t sont supportés par les graphes $\Gamma_\infty \subset U \times X$ et $\Gamma_o \subset U \times X$ de deux morphismes $\infty : U \rightarrow X$ et $o : U \rightarrow X$, et sont localement libres de rang 1 sur leurs supports.

Les morphismes ∞ et o sont appelés respectivement le pôle et le zéro du chtouca.

En d'autres termes, un chtouca à droite de rang r est la donnée d'un fibré vectoriel \mathcal{E} de rang r , d'une modification élémentaire supérieure $j : \mathcal{E} \hookrightarrow \mathcal{E}'$ de \mathcal{E} ,

d'une modification élémentaire inférieure $j' : \mathcal{E}'' \hookrightarrow \mathcal{E}'$ de \mathcal{E}' et d'un isomorphisme $u : {}^\tau \mathcal{E} \xrightarrow{\sim} \mathcal{E}''$. On a bien sûr une description similaire pour les chtoucas à gauche.

En faisant varier U , on définit de manière évidente le champ Cht^r des chtoucas à droite de rang r , un morphisme $(\infty, o) : \text{Cht}^r \rightarrow X \times X$ et un chtouca à droite de rang r universel sur $\text{Cht}^r \times X$ de pôle et de zéro les deux composantes de ce morphisme. De même, on a le champ ${}^r \text{Cht}$ des chtoucas à gauche de rang r , un morphisme $(\infty, o) : {}^r \text{Cht} \rightarrow X \times X$ et un chtouca à gauche de rang r universel.

On a des morphismes de Frobenius partiels

$$\text{Frob}_\infty : {}^r \text{Cht} \rightarrow \text{Cht}^r \quad \text{et} \quad \text{Frob}_o : \text{Cht}^r \rightarrow {}^r \text{Cht}$$

au dessus de $(\infty, o) \mapsto (\infty, \text{Frob}_X(q_j))$ et $(\infty, o) \mapsto (\text{Frob}_X(\infty), o)$, qui envoient $\tilde{\mathcal{E}}$ sur $(\mathcal{E}' \xrightarrow{j} {}^\tau \mathcal{E} \xleftarrow{{}^\tau t} \mathcal{E}')$ et $(\mathcal{E}' \xleftarrow{{}^\tau t} {}^\tau \mathcal{E} \xrightarrow{j} \mathcal{E}')$ respectivement.

Dans la suite et sauf mention explicite du contraire, les chtoucas seront tous à droite, et on dira simplement « chtouca » pour « chtouca à droite ». Les résultats démontrés pour les chtoucas à droite ont bien entendu des analogues pour les chtoucas à gauche.

PROPOSITION (Drinfeld). — *Le champ Cht^r est algébrique au sens de Deligne-Mumford. Le morphisme $(\infty, o) : \text{Cht}^r \rightarrow X \times X$ est lisse, purement de dimension relative $2r - 2$.*

Le champ Cht^r s'écrit comme réunion disjointe

$$\text{Cht}^r = \coprod_{d \in \mathbb{Z}} \text{Cht}^{r,d}$$

où $\text{Cht}^{r,d}$ classifie les chtoucas de degré

$$d = \deg \mathcal{E} = \deg \mathcal{E}' - 1.$$

Exemple : Pour tout entier d , $\text{Cht}^{1,d}$ peut être défini comme le produit fibré

$$\begin{array}{ccc} \text{Cht}^{1,d} & \longrightarrow & \text{Fib}^{1,d} \\ \downarrow & \square & \downarrow L \\ X \times X & \xrightarrow{J} & \text{Fib}^{1,0} \end{array}$$

où $\text{Fib}^{1,d}$ est le champ algébrique des fibrés en droites de degré d sur X , J est l'application d'Abel-Jacobi qui envoie $(\infty, o) \in (X \times X)(U)$ sur le fibré en droites $\mathcal{O}_{U \times X}(\infty - o)$, et L est l'isogénie de Lang qui envoie un fibré en droites \mathcal{L} sur $U \times X$ sur le fibré en droites $\mathcal{L}^{-1} \otimes_{\mathcal{O}_{U \times X}} {}^\tau \mathcal{L}$.

En particulier, $\text{Cht}^{1,d}$ est de type fini et admet un espace grossier qui est un revêtement fini étale galoisien de $X \times X$. \square

Mais, mis à part le cas $r = 1$, aucun des champs $\text{Cht}^{r,d}$ n'est de type fini.

2.2. Troncatures

Soient k un corps algébriquement clos contenant \mathbb{F}_q et $\tilde{\mathcal{E}}$ un chtouca sur (le spectre de) k . On appelle *sous-objet* de $\tilde{\mathcal{E}}$, et on note simplement

$$\tilde{\mathcal{F}} \subset \tilde{\mathcal{E}},$$

la donnée de deux sous- $\mathcal{O}_{k \otimes X}$ -Modules $\mathcal{F} \subset \mathcal{E}$ et $\mathcal{F}' \subset \mathcal{E}'$ tels que \mathcal{E}/\mathcal{F} et $\mathcal{E}'/\mathcal{F}'$ soient localement libres de même rang et que $j(\mathcal{F}) \subset \mathcal{F}'$ et $t(\tau\mathcal{F}) \subset \mathcal{F}'$; à un tel sous-objet on peut associer son *rang*

$$\text{rg } \tilde{\mathcal{F}} = \text{rg } \mathcal{F} = \text{rg } \mathcal{F}'$$

et, pour chaque nombre réel α , son degré (d'indice α)

$$\text{deg}_\alpha \tilde{\mathcal{F}} = (1 - \alpha) \text{deg } \mathcal{F} + \alpha \text{deg } \mathcal{F}'.$$

Si $0 = \tilde{\mathcal{F}}_0 \subsetneq \tilde{\mathcal{F}}_1 \subsetneq \dots \subsetneq \tilde{\mathcal{F}}_s = \tilde{\mathcal{E}}$ est une filtration de $\tilde{\mathcal{E}}$ par des sous-objets comme ci-dessus, on peut lui associer son polygone (d'indice α) qui est la fonction affine par morceaux

$$p : [0, r] \rightarrow \mathbb{R}$$

avec $p(0) = p(r) = 0$, dont les seules ruptures de pentes interviennent en les entiers $\text{rg } \tilde{\mathcal{F}}_\sigma$, $\sigma = 1, \dots, s - 1$, et qui prend en ces entiers-là la valeur

$$p(\text{rg } \tilde{\mathcal{F}}_\sigma) = \text{deg}_\alpha \tilde{\mathcal{F}}_\sigma - \frac{\text{rg } \tilde{\mathcal{F}}_\sigma}{r} \text{deg}_\alpha \tilde{\mathcal{E}}.$$

PROPOSITION. — *Fixons un nombre réel $\alpha \in [0, 1]$. Alors, parmi tous les polygones attachés aux filtrations de $\tilde{\mathcal{E}}$ comme ci-dessus, il en existe un plus grand que tous les autres, et parmi toutes les filtrations qui définissent ce polygone maximal il en existe une moins fine que tous les autres.*

Le polygone et la filtration dont la proposition ci-dessus assurent l'existence sont appelés respectivement le *polygone de Harder-Narasimhan* et la *filtration de Harder-Narasimhan* d'indice α du chtouca $\tilde{\mathcal{E}}$.

Nous appellerons *paramètre de troncature* toute fonction continue, convexe, affine par morceaux $p : [0, r] \rightarrow \mathbb{R}_{\geq 0}$ avec $p(0) = p(r) = 0$ dont les points de ruptures de pente ont des abscisses entières.

PROPOSITION. — *Étant donnés $\alpha \in [0, 1]$ et un paramètre de troncature p , il existe un unique ouvert $\text{Cht}^{r, \leq \alpha p}$ du champ Cht^r tel qu'un chtouca sur un corps algébriquement clos est dans cet ouvert si et seulement si son polygone de Harder-Narasimhan d'indice α est majoré par p .*

De plus, pour chaque entier d , l'ouvert

$$\text{Cht}^{r,d;\leq\alpha p} := \text{Cht}^{r;\leq\alpha p} \cap \text{Cht}^{r,d}$$

du champ $\text{Cht}^{r,d}$ est de type fini.

On voit donc que, pour chaque entier d , le champ algébrique localement de type fini $\text{Cht}^{r,d}$ est réunion filtrante des ouverts de type fini $\text{Cht}_\alpha^{r,d;\leq p}$.

2.3. Structures de niveau sur les chtoucas

Comme les courbes elliptiques, les chtoucas admettent des structures de niveau. Un *niveau* est ici un sous-schéma fermé fini

$$N = \text{Spec}(\mathcal{O}_N) \subset X.$$

Si \mathcal{E} est un fibré vectoriel sur $U \times X$ pour un schéma (ou un champ) U , on notera simplement \mathcal{E}_N la restriction de \mathcal{E} au fermé $U \times N \subset U \times X$. Si \mathcal{M} est un $\mathcal{O}_{U \times N}$ -Module on notera encore ${}^\tau\mathcal{M}$ le $\mathcal{O}_{U \times N}$ -Module $(\text{Frob}_U \times \text{Id}_N)^*\mathcal{M}$. Pour tout fibré vectoriel \mathcal{E} sur $U \times X$ on a évidemment ${}^\tau(\mathcal{E}_N) \cong ({}^\tau\mathcal{E})_N$.

On considère le champ algébrique

$$\text{Tr}_N^r$$

qui associe au schéma U la catégorie des couples (\mathcal{F}, u) formés d'un fibré vectoriel de rang r sur $U \times N$ et d'un isomorphisme de fibrés vectoriels $u : {}^\tau\mathcal{F} \xrightarrow{\sim} \mathcal{F}$. Si l'on note $\text{GL}_r^N = \text{Res}_{N/\text{Spec}(\mathbb{F}_q)} \text{GL}_r$ le schéma en groupes sur \mathbb{F}_q restriction à la Weil du schéma en groupes linéaire GL_r sur N et par $h \mapsto \tau(h) = \text{Frob}_{\text{GL}_r}(h)$ son endomorphisme de Frobenius (relatif à \mathbb{F}_q), ce champ n'est autre que le quotient

$$\text{Tr}_N^r = [\text{GL}_r^N / {}^\tau\text{Ad}(\text{GL}_r^N)]$$

de GL_r^N par l'action de GL_r^N sur lui-même par la conjugaison tordue

$${}^\tau\text{Ad}(h)(g) = \tau(h)^{-1}gh.$$

Comme l'*isogénie de Lang*

$$L : \text{GL}_r^N \rightarrow \text{GL}_r^N, h \mapsto L(h) = \tau(h)^{-1}h,$$

est un torseur sous le groupe fini $\text{GL}_r(\mathcal{O}_N) = \text{GL}_r^N(\mathbb{F}_q)$ (pour son action par translation à droite sur GL_r^N), Tr_N^r est canoniquement isomorphe au champ classifiant $[\text{Spec}(\mathbb{F}_q)/\text{GL}_r(\mathcal{O}_N)]$.

On ne munira ici de structures de niveau que les chtoucas dont le pôle et le zéro ne rencontrent pas N . Si

$$\tilde{\mathcal{E}} = (\mathcal{E} \xleftarrow{j} \mathcal{E}' \xleftarrow{t} {}^\tau\mathcal{E})$$

est un tel chtouca sur U on remarque que les restrictions

$$j_N : \mathcal{E}_N \rightarrow \mathcal{E}'_N \text{ et } t_N : {}^\tau(\mathcal{E}_N) \cong ({}^\tau\mathcal{E})_N \rightarrow \mathcal{E}'_N$$

des flèches j et t sont des isomorphismes de $\mathcal{O}_{U \times N}$ -Modules. On a donc un morphisme de champs algébriques

$$\text{Cht}^r \times_{X^2} (X - N)^2 \longrightarrow \text{Tr}_N^r, \tilde{\mathcal{E}} \mapsto (\mathcal{E}_N, j_N^{-1} \circ t_N).$$

DÉFINITION (Drinfeld). — Soit $\tilde{\mathcal{E}}$ un chtouca de rang r sur un schéma U dont le pôle et le zéro se factorisent par l'immersion ouverte $X - N \hookrightarrow X$. Une structure (principale) de niveau N sur ce chtouca est la donnée d'un isomorphisme

$$\iota : \mathcal{O}_{U \times N}^r \xrightarrow{\sim} \mathcal{E}_N$$

tel que

$$\iota = (j_N^{-1} \circ t_N) \circ \tau \iota : \mathcal{O}_{U \times N}^r \xrightarrow{\sim} \mathcal{E}_N.$$

On définit de manière évidente le champ Cht_N^r des chtoucas de rang r , dont le pôle et le zéro ne rencontrent pas N , et avec structures de niveau N . On a un carré cartésien

$$\begin{array}{ccc} \text{Cht}_N^r & \longrightarrow & \text{Tr}_N^{r,\tau} \\ \downarrow & \square & \downarrow \\ \text{Cht}^r \times_{X^2} (X - N)^2 & \longrightarrow & \text{Tr}_N^r \end{array}$$

où la flèche horizontale du bas est la flèche de restriction à N définie ci-dessus, où $\text{Tr}_N^{r,\tau}$ est le champ qui associe au schéma U la catégorie des triplets (\mathcal{F}, u, ι) formés d'un objet (\mathcal{F}, u) de $\text{Tr}_N^r(U)$ et d'un isomorphisme de $\mathcal{O}_{U \times N}$ -Modules $\iota : \mathcal{O}_{U \times N}^r \xrightarrow{\sim} \mathcal{F}$ tel que $\iota = u \circ \tau \iota$, et où la flèche verticale de droite est l'oubli de la composante ι . On remarquera que la composante ι de (\mathcal{F}, u, ι) détermine uniquement la composante u et donc que $\text{Tr}_N^{r,\tau}$ est canoniquement isomorphe au champ $[\text{GL}_r^N / \text{GL}_r^N]$ quotient de GL_r^N par l'action de GL_r^N sur lui-même par translation à droite, c'est-à-dire au schéma réduit à un point $\text{Spec}(\mathbb{F}_q)$. Par conséquent, la flèche verticale de droite du carré ci-dessus se décrit encore, soit comme la flèche $[\text{GL}_r^N / \text{GL}_r^N] \rightarrow [\text{GL}_r^N / \tau \text{Ad}(\text{GL}_r^N)]$ induite par l'isogénie de Lang L , soit comme le $\text{GL}_r(\mathcal{O}_N)$ -torseur canonique $\text{Spec}(\mathbb{F}_q) \rightarrow [\text{Spec}(\mathbb{F}_q) / \text{GL}_r(\mathcal{O}_N)]$.

La flèche verticale de gauche est donc un $\text{GL}_r(\mathcal{O}_N)$ -torseur et Cht_N^r est un champ algébrique, lisse purement de dimension $2r - 2$ sur $(X - N)^2$.

3. HOMOMORPHISMES COMPLETS ET CHTOUCAS ITÉRÉS

3.1. Homomorphismes complets

Soient k un corps et V et W deux k -espaces vectoriels de dimension r . On notera $H(V, W)$ le schéma des uplets

$$(u_1, \dots, u_r; \lambda_1, \dots, \lambda_{r-1})$$

où, pour $\rho = 1, \dots, r$, u_ρ est une application linéaire non nulle de $\wedge^\rho V$ dans $\wedge^\rho W$ et où $\lambda_1, \dots, \lambda_{r-1}$ sont des scalaires. Les conditions

- u_1 est un isomorphisme,

- $\lambda_1, \dots, \lambda_{r-1}$ sont inversibles,
- $\wedge^\rho u_1 = \lambda_1^{\rho-1} \lambda_2^{\rho-2} \dots \lambda_{\rho-1} u_\rho$ pour $\rho = 2, \dots, r$,

définissent un sous-schéma localement fermé $\Omega^\circ(V, W)$ de ce schéma affine et la projection $(u_1; \lambda_1, \dots, \lambda_{r-1})$ identifie $\Omega^\circ(V, W)$ au schéma $\text{Isom}(V, W) \times_k \mathbb{G}_{m,k}^{r-1}$. Nous noterons

$$\Omega(V, W) \subset H(V, W)$$

l'adhérence schématique de $\Omega^\circ(V, W)$ dans $H(V, W)$.

Le schéma \mathbb{A}_k^{r-1} est naturellement muni d'un diviseur à croisements normaux, réunion de $r-1$ diviseurs lisses, à savoir les diviseurs $\{\lambda_\rho = 0\}$ pour $\rho = 1, \dots, r-1$. Ce diviseur à croisement normaux définit de la manière habituelle une stratification localement fermée de \mathbb{A}_k^{r-1} , indexée par les parties R de $[r-1] = \{1, \dots, r-1\}$ et de R -ième strate

$$\mathbb{A}_{k,R}^{r-1} = \{(\lambda_1, \dots, \lambda_{r-1}) \mid \lambda_\rho = 0 \Leftrightarrow \rho \in R\} \cong \mathbb{G}_{m,k}^{[r-1]-R}.$$

Par construction, on dispose d'un morphisme

$$(\lambda_1, \dots, \lambda_{r-1}) : \Omega(V, W) \rightarrow \mathbb{A}_k^{r-1}.$$

et on peut relever à $\Omega(V, W)$ le diviseur et la stratification de \mathbb{A}_k^{r-1} que l'on vient d'introduire. En particulier, pour tout sous-ensemble $R \subset [r-1]$, on notera

$$\Omega_R(V, W) \subset \Omega(V, W)$$

l'image réciproque par ce morphisme de la strate localement fermée $\mathbb{A}_{k,R}^{r-1} \subset \mathbb{A}_k^{r-1}$. On vérifie que $\Omega_\emptyset(V, W) = \Omega^\circ(V, W)$.

PROPOSITION. — *Pour chaque $R = \{r_1, r_2, \dots, r_{s-1}\} \subset [r-1]$ où $0 = r_0 < r_1 < r_2 < \dots < r_{s-1} < r_s = r$, $\Omega_R(V, W)$ est isomorphe au schéma des uplets*

$$(V^\bullet, W_\bullet, (v_\sigma)_{\sigma=1, \dots, s-1}; (\lambda_\rho)_{\rho \in [r-1]-R})$$

où :

- les λ_ρ sont des scalaires inversibles,
- $V^\bullet = (V = V^0 \supseteq V^1 \supseteq \dots \supseteq V^s = (0))$ est une filtration décroissante par des sous-espaces vectoriels de codimensions $0 = r_0, r_1, \dots, r_s = r$,
- $W_\bullet = ((0) = W_0 \subsetneq W_1 \subsetneq \dots \subsetneq W_s = W)$ est une filtration croissante par des sous-espaces vectoriels de dimensions $0 = r_0, r_1, \dots, r_s = r$,
- $v_\sigma : V^{\sigma-1}/V^\sigma \xrightarrow{\sim} W_\sigma/W_{\sigma-1}$ est un isomorphisme d'espaces vectoriels.

A fortiori, le morphisme $(\lambda_\rho)_{\rho \in [r-1]-R} : \Omega_R(V, W) \rightarrow \mathbb{G}_{m,k}^{[r-1]-R}$ est lisse de dimension relative r^2 . \square

Remarque : L'écriture $R = \{r_1, r_2, \dots, r_{s-1}\}$ où $0 = r_0 < r_1 < r_2 < \dots < r_{s-1} < r_s = r$ identifie les parties de $[r-1]$ aux décompositions $r = (r_1 - r_0) + (r_2 - r_1) + \dots + (r_s - r_{s-1})$ de r en entiers strictement positifs. \square

COROLLAIRE. — *Le morphisme $(\lambda_1, \dots, \lambda_{r-1})$ est lisse purement de dimension relative r^2 et $\Omega^\circ(V, W) \subset \Omega(V, W)$ est l'ouvert complémentaire d'un diviseur à croisements normaux, réunion des $r-1$ diviseurs lisses $\{\lambda_\rho = 0\}$ pour $\rho = 1, \dots, r-1$.*

Le tore $\mathbb{G}_{m,k}^{r-1} = \{(\mu_1, \dots, \mu_{r-1})\}$ agit librement sur $\Omega(V, W)$ par

$$u_1 \mapsto u_1, \quad u_2 \mapsto \mu_1^{-1}u_2, \quad u_3 \mapsto \mu_1^{-2}\mu_2u_3, \quad \dots, \quad u_r \mapsto \mu_1^{1-r}\mu_2^{2-r} \cdots \mu_{r-1}u_r,$$

et

$$\lambda_1 \mapsto \mu_1\lambda_1, \quad \lambda_2 \mapsto \mu_2\lambda_2, \quad \dots, \quad \lambda_{r-1} \mapsto \mu_{r-1}\lambda_{r-1},$$

et le quotient

$$\widetilde{\text{Hom}}(V, W) := \Omega(V, W) / \mathbb{G}_{m,k}^{r-1}$$

est un schéma quasi-projectif et lisse, qui contient comme ouvert dense $\text{Isom}(V, W)$ avec pour fermé complémentaire une réunion de $r-1$ diviseurs lisses à croisements normaux. C'est par définition le *schéma des homomorphismes complets de V dans W* . Par construction, on a un morphisme représentable, lisse et purement de dimension relative r^2 , de champs algébriques

$$\widetilde{\text{Hom}}(V, W) \rightarrow [\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}],$$

et le diviseur à croisements normaux ci-dessus est l'image réciproque par ce morphisme du diviseur à croisements normaux évident sur le champ algébrique $[\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}]$.

Si $V = W = k^r$ on notera encore $\widetilde{\text{gl}}_{r,k}$ le k -schéma $\widetilde{\text{Hom}}(k^r, k^r)$ des *endomorphismes complets* de k^r . La strate ouverte de $\widetilde{\text{gl}}_{r,k}$ est bien entendu $\text{GL}_{r,k}$.

Remarques : (i) Le morphisme

$$u_1 : \widetilde{\text{Hom}}(V, W) \rightarrow \text{Hom}(V, W) - \{0\}$$

peut aussi être obtenu comme le composé

$$\widetilde{\text{Hom}}(V, W) = H_{r-1} \rightarrow \dots \rightarrow H_1 \rightarrow H_0 = \text{Hom}(V, W) - \{0\}$$

où $H_1 \rightarrow H_0$ est l'éclatement de H_0 le long du fermé des homomorphismes de rang 1 et où, pour $\rho = 2, \dots, r-1$, $H_\rho \rightarrow H_{\rho-1}$ est l'éclatement de $H_{\rho-1}$ le long du transformé strict du fermé de H_0 formé des homomorphismes de rang $\leq \rho$.

(ii) Le groupe multiplicatif agit par homothétie sur les homomorphismes complets. Le quotient $\widetilde{\text{gl}}_{r,k} / \mathbb{G}_{m,k}$ est un schéma projectif, isomorphe à la compactification de De Concini et Procesi de $\text{PGL}_{r,k}$. \square

Pour tout k -schéma U , la catégorie $[\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}](U)$ a pour objets les uplets

$$((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $\mathcal{L}_1, \dots, \mathcal{L}_{r-1}$ sont des \mathcal{O}_U -Modules inversibles et où $\lambda_1, \dots, \lambda_{r-1}$ sont des sections globales de ces fibrés en droites. On peut donc voir un U -homomorphisme complet de V dans W , c'est-à-dire un U -point de $\widetilde{\text{Hom}}(V, W)$, comme un uplet

$$u = (u_1, \dots, u_r; (\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1})) \in \text{ob} [\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}](U)$ et où

$$u_\rho : \bigwedge^\rho V \otimes_k \mathcal{L}_1^{\otimes(\rho-1)} \otimes \mathcal{L}_2^{\otimes(\rho-2)} \otimes \dots \otimes \mathcal{L}_{\rho-1} \rightarrow \bigwedge^\rho W$$

est un homomorphisme partout non nul de \mathcal{O}_U -Modules pour $\rho = 1, \dots, r$. Un homomorphisme complet u est donc bien un homomorphisme u_1 complété par des données supplémentaires et peut être noté commodément $u : V \Rightarrow W$.

Soient maintenant S un k -schéma et \mathcal{V}, \mathcal{W} deux \mathcal{O}_S -modules localement libres de rang constant r . Si U est un S -schéma, on peut considérer plus généralement les uplets

$$u = (u_1, \dots, u_r; (\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1})) \in \text{ob} [\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}](U)$ et où

$$u_\rho : \bigwedge^\rho \mathcal{V}_U \otimes \mathcal{L}_1^{\otimes(\rho-1)} \otimes \mathcal{L}_2^{\otimes(\rho-2)} \otimes \dots \otimes \mathcal{L}_{\rho-1} \rightarrow \bigwedge^\rho \mathcal{W}_U,$$

est un homomorphisme partout non nul de \mathcal{O}_U -Modules pour $\rho = 1, \dots, r$. (On a noté \mathcal{V}_U et \mathcal{W}_U les restrictions de \mathcal{V} et \mathcal{W} à U .) On dira alors qu'un tel uplet est un U -homomorphisme complet de \mathcal{V} dans \mathcal{W} , et on utilisera la notation $u : \mathcal{V} \Rightarrow \mathcal{W}$, si, pour toutes trivialisations locales $\mathcal{V} \cong V \otimes_k \mathcal{O}_S$ et $\mathcal{W} \cong W \otimes_k \mathcal{O}_S$, ce uplet est un U -homomorphisme complet de V dans W .

On définit ainsi un S -schéma $\widetilde{\text{Hom}}(\mathcal{V}, \mathcal{W})$, muni d'un morphisme représentable, lisse et purement de dimension relative r^2 , de champs algébriques

$$\widetilde{\text{Hom}}(\mathcal{V}, \mathcal{W}) \rightarrow S \times_k [\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}]$$

Le S -schéma quasi-projectif $\widetilde{\text{Hom}}(\mathcal{V}, \mathcal{W})$ contient le S -schéma $\text{Isom}(\mathcal{V}, \mathcal{W})$ comme un ouvert dense, et le fermé complémentaire est un diviseur à croisement normaux relatifs sur S , réunion de $r - 1$ diviseurs lisses.

3.2. Chtoucas itérés

DÉFINITION. — Un pré-chtouca itéré de rang r sur un schéma U est la donnée :

- d'un diagramme

$$\mathcal{E} \xrightarrow{j} \mathcal{E}' \xleftarrow{j'} \mathcal{E}''$$

et de deux morphismes $\infty : U \rightarrow X$ et $o : U \rightarrow X$, où \mathcal{E} est un fibré vectoriel de rang r sur $U \times X$, j est une modification élémentaire supérieure de \mathcal{E} le

long du graphe de ∞ et j' est une modification élémentaire inférieure de \mathcal{E}' le long du graphe de o ,

- de \mathcal{O}_U -Modules inversibles $\mathcal{L}_1, \dots, \mathcal{L}_{r-1}$ munis de sections globales $\lambda_1, \dots, \lambda_{r-1}$,
- pour chaque entier $\rho = 1, \dots, r$, d'un homomorphisme de $\mathcal{O}_{U \times X}$ -Modules

$$u_\rho : \wedge^\rho(\tau \mathcal{E}) \otimes \mathcal{L}_1^{\otimes(q-1)(\rho-1)} \otimes \mathcal{L}_2^{\otimes(q-1)(\rho-2)} \otimes \dots \otimes \mathcal{L}_{\rho-1}^{\otimes(q-1)} \rightarrow \wedge^\rho \mathcal{E}'' ,$$

de telle sorte que le uplet

$$u = (u_1, \dots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \dots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : \tau \mathcal{E} \Rightarrow \mathcal{E}''$$

soit un homomorphisme complet.

En faisant varier U , on définit de manière évidente le champ \mathcal{C}^r des pré-chtoucas itérés de rang r . Il n'est pas difficile de vérifier que c'est un champ algébrique (au sens d'Artin), localement de type fini, muni d'un morphisme de champs

$$((\infty, o), ((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))) : \mathcal{C}^r \rightarrow X \times X \times [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}].$$

En particulier, pour tout $R \subset [r-1]$, on dispose de la strate localement fermée $\mathcal{C}_R^r \subset \mathcal{C}^r$ image réciproque de la strate $[\mathbb{A}_R^{r-1}/\mathbb{G}_m^{r-1}] \subset [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$.

D'après la proposition du paragraphe 3.1, si $R = \{r_1, \dots, r_{s-1}\}$ où $0 = r_0 < r_1 < \dots < r_{s-1} < r_s = r$, la donnée d'un U -point de \mathcal{C}_R^r au-dessus d'un U -point $((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$ de $[\mathbb{A}_R^{r-1}/\mathbb{G}_m^{r-1}]$ équivaut aux données suivantes :

- un diagramme

$$\mathcal{E} \xleftarrow{j} \mathcal{E}' \xleftarrow{j'} \mathcal{E}''$$

et deux morphismes $\infty : U \rightarrow X$ et $o : U \rightarrow X$, où \mathcal{E} est un fibré vectoriel de rang r sur $U \times X$, j est une modification élémentaire supérieure de \mathcal{E} le long du graphe de ∞ et j' est une modification élémentaire inférieure de \mathcal{E}' le long du graphe de o ,

- une filtration décroissante

$$\tau \mathcal{E} = \mathcal{F}^0 \supseteq \mathcal{F}^1 \supseteq \dots \supseteq \mathcal{F}^{s-1} \supseteq \mathcal{F}^s = (0)$$

de $\tau \mathcal{E}$ par des sous- $\mathcal{O}_{U \times X}$ -Modules localement facteurs directs, de corangs $0 = r_0, r_1, \dots, r_{s-1}, r_s = r$,

- une filtration croissante

$$(0) = \mathcal{E}''_0 \subsetneq \mathcal{E}''_1 \subsetneq \dots \subsetneq \mathcal{E}''_{s-1} \subsetneq \mathcal{E}''_s = \mathcal{E}''$$

de \mathcal{E}'' par des sous- $\mathcal{O}_{U \times X}$ -Modules localement facteurs directs, de rangs $0 = r_0, r_1, \dots, r_{s-1}, r_s = r$,

- une famille d'isomorphismes de $\mathcal{O}_{U \times X}$ -Modules

$$v_\sigma : (\mathcal{F}^{\sigma-1}/\mathcal{F}^\sigma) \otimes \left(\bigotimes_{\sigma'=1}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \xrightarrow{\sim} (\mathcal{E}''_\sigma/\mathcal{E}''_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right), \quad \sigma = 1, \dots, s.$$

Pour un tel U -point notons

$$\mathcal{E}'_\sigma = j'(\mathcal{E}''_\sigma) \subset \mathcal{E}', \quad \mathcal{E}_\sigma = j^{-1}(\mathcal{E}'_\sigma) \subset \mathcal{E}, \quad \forall \sigma = 0, 1, \dots, s-1, \quad \text{et } \mathcal{E}'_s = \mathcal{E}', \quad \mathcal{E}_s = \mathcal{E},$$

$$\mathcal{G}_1 = \mathcal{E}_1, \quad \mathcal{G}'_1 = \mathcal{E}'_1 \quad \text{et} \quad \mathcal{G}_\sigma = \mathcal{E}_\sigma/\mathcal{E}_{\sigma-1}, \quad \mathcal{G}'_\sigma = \mathcal{F}^{\sigma-1} \cap \tau \mathcal{E}_\sigma, \quad \forall \sigma = 2, \dots, s.$$

On dispose de l'homomorphisme $\mathcal{G}_1 \hookrightarrow \mathcal{G}'_1$ induit par j et du composé

$$\tau \mathcal{G}_1 \hookrightarrow \tau \mathcal{E} \twoheadrightarrow \tau \mathcal{E}/\mathcal{F}^1 \xrightarrow{\sim} \mathcal{E}''_1 \xrightarrow{j'} \mathcal{G}'_1,$$

et, pour chaque $\sigma = 2, \dots, r$, on dispose du composé

$$\begin{aligned} \mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) &\hookrightarrow \mathcal{F}^{\sigma-1} \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \twoheadrightarrow (\mathcal{F}^{\sigma-1}/\mathcal{F}^\sigma) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \\ &\xrightarrow{\sim} (\mathcal{E}''_\sigma/\mathcal{E}''_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \xrightarrow{\bar{j}'} (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \end{aligned}$$

où \bar{j}' est induite par j' , de l'homomorphisme

$$\bar{j}' : \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \rightarrow (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right)$$

induit par j , et du composé

$$\mathcal{G}'_\sigma \hookrightarrow \tau \mathcal{E}_\sigma \twoheadrightarrow \tau \mathcal{E}_\sigma/\tau \mathcal{E}_{\sigma-1} = \tau \mathcal{G}_\sigma.$$

On a donc des diagrammes

$$\tilde{\mathcal{G}}_1 = (\mathcal{G}_1 \hookrightarrow \mathcal{G}'_1 \hookrightarrow \tau \mathcal{G}_1)$$

et

$$\tilde{\mathcal{G}}_\sigma = \left(\begin{array}{c} \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \\ \downarrow \\ (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \hookrightarrow \mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \hookrightarrow \tau \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \end{array} \right)$$

pour $\sigma = 2, \dots, s$.

LEMME. — Pour tout $\rho = 1, \dots, r$, notons u_ρ° la restriction de u_ρ à l'ouvert de $U \times X$ complémentaire des graphes des morphismes ∞ et o , et identifions les restrictions de \mathcal{E} et \mathcal{E}'' à cet ouvert à l'aide de $j'^{-1} \circ j$. Avec ces notations les deux conditions ci-dessous sont équivalentes :

(a) il n'existe aucun point géométrique de U tel que la fibre en ce point d'un des u_ρ° soit nilpotente.

(a') pour tout $\sigma = 1, \dots, s-1$, on a $(\tau j)(\mathcal{F}^\sigma) \cap \tau \mathcal{E}'_\sigma = (0)$ dans $\tau \mathcal{E}'$.

De plus, si ces conditions équivalentes sont vérifiées, le plus grand ouvert de $U \times X$ où toutes les flèches des diagrammes $\tilde{\mathcal{G}}_\sigma$ ci-dessus sont des isomorphismes rencontre chaque fibre de la projection canonique $U \times X \rightarrow U$ suivant un ouvert dense.

La condition (a) a été introduite par Drinfeld en rang $r = 2$.

Suivant Lafforgue considérons les conditions supplémentaires suivantes :

(b) $\mathcal{E}'/\mathcal{E}'_\sigma$ est un $\mathcal{O}_{U \times S}$ -Module localement libre pour $\sigma = 0, 1, \dots, s-1$,

(c) pour tout $\sigma = 1, \dots, s$, l'homomorphisme composé $\mathcal{E}'' \xrightarrow{j} \mathcal{E}' \twoheadrightarrow \mathcal{E}'/j(\mathcal{E})$ est surjectif,

(d) pour tout $\sigma = 1, \dots, s-1$, on a $\mathcal{F}^{\sigma-1} + \tau \mathcal{E}_\sigma = \tau \mathcal{E}$,

La condition (c) assure en particulier que $\bar{j} : \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \rightarrow (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right)$ est un isomorphisme. On peut donc récrire les diagrammes $\tilde{\mathcal{G}}_\sigma$, $\sigma = 2, \dots, s$, sous la forme

$$\tilde{\mathcal{G}}_\sigma = \left(\mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \hookrightarrow \mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \hookrightarrow \tau \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \tau \mathcal{L}_{r_{\sigma'}} \right) \right).$$

LEMME. — (i) Sous les conditions (a), (b), (c) et (d) le diagramme $\tilde{\mathcal{G}}_1$ est un chtouca à droite de rang r_1 et de pôle ∞ , les diagrammes $\tilde{\mathcal{G}}_2, \dots, \tilde{\mathcal{G}}_s$ sont des chtoucas à gauche, le zéro de $\tilde{\mathcal{G}}_\sigma$ est le pôle de $\tilde{\mathcal{G}}_{\sigma+1}$ pour $\sigma = 1, \dots, s-1$ et le zéro de $\tilde{\mathcal{G}}_s$ est o .

(ii) Il existe un unique sous-champ ouvert $\overline{\text{Cht}}^r \subset \mathcal{C}^r$ tel que pour chaque $R = \{r_1, \dots, r_{s-1}\}$ comme ci-dessus la trace de cet ouvert sur \mathcal{C}_R^r soit définie par les conditions (a), (b), (c) et (d).

Le champ $\overline{\text{Cht}}^r$ est par définition le champ des *chtoucas itérés* de rang r . Il contient comme ouvert dense le champ des chtoucas Cht^r . Tout comme un chtouca ordinaire, un chtouca itéré admet un degré, à savoir le degré du fibré vectoriel sous-jacent \mathcal{E} , et on a un découpage de $\overline{\text{Cht}}^r$ en composantes

$$\overline{\text{Cht}}^r = \coprod_{d \in \mathbb{Z}} \overline{\text{Cht}}^{r,d}$$

indexées par ce degré. Bien sûr, pour chaque entier d , $\text{Cht}^{r,d}$ est un ouvert dense de $\overline{\text{Cht}}^{r,d}$.

Pour chaque $R = \{r_1, \dots, r_{s-1}\} \subset [r-1]$ on notera $\overline{\text{Cht}}_R^r$ l'intersection de $\overline{\text{Cht}}^r$ avec la strate \mathcal{C}_R^r . C'est un sous-champ localement fermé de $\overline{\text{Cht}}^r$, qui n'est autre que l'ouvert Cht^r pour $R = \emptyset$, et $\overline{\text{Cht}}^r$ est la réunion disjointes des $\overline{\text{Cht}}_R^r$.

Il résulte du dernier lemme que l'on a un morphisme fini, surjectif et radiciel de champs

$$\overline{\text{Cht}}_R^r \rightarrow \text{Cht}^{r_1} \times_X^{r_2-r_1} \text{Cht} \cdots \times_X^{r_s-r_{s-1}} \text{Cht}.$$

De plus on a un morphisme de Frobenius partiel (au dessus de l'endomorphisme $\text{Id}_X \times \text{Frob}_X$ de $X \times X$)

$$\begin{aligned} & \text{Cht}^{r_1} \times_X^{r_2-r_1} \text{Cht} \cdots \times_X^{r_s-r_{s-1}} \text{Cht} \\ & \rightarrow \text{Cht}^R := \text{Cht}^{r_1} \times_X \text{Cht}^{r_2-r_1} \times_{X, \text{Frob}_X} \cdots \times_{X, \text{Frob}_X} \text{Cht}^{r_s-r_{s-1}} \end{aligned}$$

qui envoie $(\tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2, \dots, \tilde{\mathcal{G}}_s)$ sur

$$(\tilde{\mathcal{G}}_1, \text{Frob}_o(\tilde{\mathcal{G}}_2), \dots, \text{Frob}_o(\tilde{\mathcal{G}}_s))$$

qui est lui aussi fini, surjectif et radiciel. Par composition on a donc un morphisme fini, surjectif et radiciel de champs

$$\overline{\text{Cht}}_R^r \rightarrow \text{Cht}^R.$$

Le but de ce dernier morphisme est un champ de Deligne-Mumford séparé et lisse purement de dimension relative

$$(2r_1 - 2) + (2(r_2 - r_1) - 2) + \cdots + (2(r_s - r_{s-1}) - 2) = 2r - 2s$$

au dessus de $X \times X^{s-1} \times X$ par le morphisme $(\infty = \infty_1, \infty_2, \dots, \infty_s, o_s = o)$. Il se décompose en

$$\text{Cht}^R = \coprod_{d_\bullet \in \mathbb{Z}^s} \text{Cht}^{R, d_\bullet}$$

où

$$\text{Cht}^{R, d_\bullet} := \text{Cht}^{r_1, d_1} \times_X \text{Cht}^{r_2-r_1, d_2} \times_{X, \text{Frob}_X} \cdots \times_{X, \text{Frob}_X} \text{Cht}^{r_s-r_{s-1}, d_s}.$$

On notera

$$\overline{\text{Cht}}_R^r = \coprod_{d_\bullet \in \mathbb{Z}^s} \overline{\text{Cht}}_R^{r, d_\bullet}$$

la décomposition correspondante de la source. On vérifie que

$$\overline{\text{Cht}}_R^{r, d} := \overline{\text{Cht}}_R^r \cap \overline{\text{Cht}}^{r, d} = \coprod_{\substack{d_\bullet \in \mathbb{Z}^s \\ d_1 + \cdots + d_s = d - s + 1}} \overline{\text{Cht}}_R^{r, d_\bullet}$$

(le décalage $-s + 1$ provient des Frobenius partiels).

3.3. Chtoucas itérés et troncature

Soient d un entier et $p : [0, r] \rightarrow \mathbb{R}$ un paramètre de troncature. Pour chaque $\rho = 0, 1, \dots, r$, notons $\tilde{p}(\rho)$ l'unique entier appartenant à l'intervalle de longueur 1

$$]p(\rho) + \frac{\rho}{r}d - 1, p(\rho) + \frac{\rho}{r}d].$$

On définit alors des entiers d_1, \dots, d_s par

$$d_1 = \tilde{p}(r_1) \text{ et } d_\sigma = \tilde{p}(r_\sigma) - \tilde{p}(r_{\sigma-1}) - 1, \quad \forall \sigma = 2, \dots, s,$$

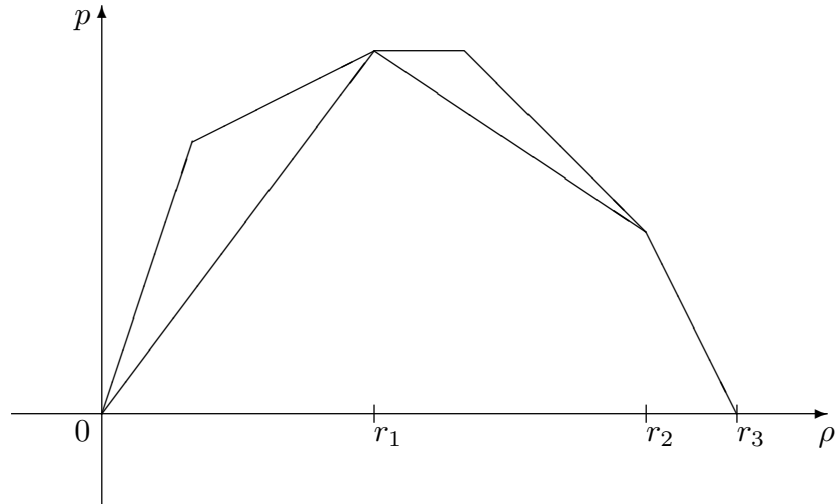
et des paramètres de troncature $p_1 : [0, r_1] \rightarrow \mathbb{R}, \dots, p_s : [0, r_s - r_{s-1}] \rightarrow \mathbb{R}$ en imposant que

$$p_1(\rho_1) = \tilde{p}(\rho_1) - \frac{\rho_1}{r_1}d_1, \quad \forall \rho_1 = 1, \dots, r_1 - 1,$$

et

$$p_\sigma(\rho_\sigma) = \tilde{p}(r_{\sigma-1} + \rho_\sigma) - \tilde{p}(r_{\sigma-1}) - 1 - \frac{\rho_\sigma}{r_\sigma - r_{\sigma-1}}d_\sigma, \quad \forall \rho_\sigma = 1, \dots, r_\sigma - r_{\sigma-1}.$$

À décalage et normalisation près, les p_σ sont essentiellement les restrictions de p aux intervalles $[r_{\sigma-1}, r_\sigma]$ comme dans la figure ci-dessous :



On note

$$\begin{aligned} \text{Cht}^{R,d;\leq p} &= \text{Cht}^{r_1,d_1;\leq p_1} \times_X \text{Cht}^{r_2-r_1,d_2;\leq p_2} \times_{X,\text{Frob}_X} \cdots \\ &\cdots \times_{X,\text{Frob}_X} \text{Cht}^{r_s-r_{s-1},d_s;\leq p_s} \subset \text{Cht}^R. \end{aligned}$$

DÉFINITION. — Un paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ est dit μ -convexe pour un nombre réel $\mu \geq 0$ si l'on a

$$(p(\rho) - p(\rho - 1)) - (p(\rho + 1) - p(\rho)) \geq \mu, \quad \forall 1 \leq \rho \leq r - 1.$$

Une propriété est dite vraie « pour tout paramètre de troncature assez convexe » s'il existe un nombre réel $\mu \geq 0$ tel que la propriété soit vraie pour tout paramètre de troncature qui est μ -convexe.

On remarquera que les paramètres de troncature p_σ définis ci-dessus sont automatiquement $(\mu - 2)$ -convexes dès que p est μ -convexe.

THÉORÈME. — *Pour tout entier d et tout paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ qui est 2-convexe, il existe un ouvert*

$$\overline{\text{Cht}}^{r,d;\leq p} \subset \overline{\text{Cht}}^{r,d}$$

de la composante d -ième du champ des chtoucas itérés ayant les propriétés suivantes :

- pour tout $R \subset [r - 1]$, $\overline{\text{Cht}}_R^{r,d;\leq p} := \overline{\text{Cht}}_R^{r,d} \cap \overline{\text{Cht}}^{r,d;\leq p}$ est l'image réciproque par le morphisme fini, surjectif et radiciel $\overline{\text{Cht}}_R^r \rightarrow \text{Cht}^R$ de l'ouvert $\text{Cht}^{R,d;\leq p}$;
- le morphisme de champs $(\infty, o) : \overline{\text{Cht}}^{r,d;\leq p} = \overline{\text{Cht}}^{r;\leq p} \cap \overline{\text{Cht}}^{r,d} \rightarrow X \times X$ est propre (et donc séparé et de type fini) ;
- si l'on suppose de plus que p est assez convexe relativement au genre de la courbe X , le morphisme de champs (∞, o) ci-dessus est lisse purement de dimension relative $2r - 2$ et le fermé complémentaire dans $\overline{\text{Cht}}^{r,d;\leq p}$ de la strate ouverte $\text{Cht}^{r,d;\leq p} = \overline{\text{Cht}}_{\emptyset}^{r,d;\leq p}$ est un diviseur à croisements normaux relatif sur $X \times X$, qui est réunion de $r - 1$ diviseurs lisses et dont la stratification canonique est celle par les $\overline{\text{Cht}}_R^{r,d;\leq p}$ pour R parcourant les parties non vides de $[r - 1]$.

On posera

$$\overline{\text{Cht}}^{r;\leq p} = \coprod_{d \in \mathbb{Z}} \overline{\text{Cht}}^{r,d;\leq p} \subset \overline{\text{Cht}}^r$$

et

$$\overline{\text{Cht}}_R^{r;\leq p} = \coprod_{d \in \mathbb{Z}} \overline{\text{Cht}}_R^{r,d;\leq p} \subset \overline{\text{Cht}}^{r;\leq p}.$$

3.4. Structures de niveau sur les chtoucas itérés

Dans cet exposé nous nous limiterons au cas où N est réduit et supporté par un point fermé de X rationnel sur le corps fini de base \mathbb{F}_q . On suppose donc $N \cong \text{Spec}(\mathbb{F}_q)$ et on a simplement $\text{GL}_r^N = \text{GL}_r$.

Soit

$$\mathcal{E} \xrightarrow{j} \mathcal{E}' \xleftarrow{j'} \mathcal{E}'' \xleftarrow{u} \tau \mathcal{E},$$

un chtouca itéré sur un schéma U ; $u = (u_1, \dots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \dots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1}))$ est donc un homomorphisme complet de $\tau \mathcal{E}$ dans \mathcal{E}'' . Si le pôle et le zéro de ce chtouca itéré ne rencontrent pas N , les restrictions

$$j_N : \mathcal{E}_N \rightarrow \mathcal{E}'_N \text{ et } j'_N : \mathcal{E}''_N \rightarrow \mathcal{E}'_N$$

des flèches j et j' sont des isomorphismes de \mathcal{O}_U -Modules localement libres de rang r et la restriction

$$u_N = (u_{1,N}, \dots, u_{r,N}; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \dots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : \tau \mathcal{E}_N \implies \mathcal{E}_N$$

est un homomorphisme complet. Le morphisme de restriction à N

$$\text{Cht}^r \times_{X^2} (X - N)^2 \rightarrow \text{Tr}_N^r$$

se prolonge donc en un morphisme

$$\overline{\text{Cht}}^r \times_{X^2} (X - N)^2 \rightarrow \overline{\text{Tr}}_N^r$$

où $\overline{\text{Tr}}_N^r$ est le champ qui associe au schéma U la catégorie des uplets

$$(\mathcal{F}, u_1, \dots, u_r; (\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

formés d'un fibré \mathcal{F} de rang r sur N , de fibrés en droites $\mathcal{L}_1, \dots, \mathcal{L}_{r-1}$ sur U munis de sections globales $\lambda_1, \dots, \lambda_{r-1}$ et d'un homomorphisme complet

$$u = (u_1, \dots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \dots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : {}^\tau\mathcal{F} \implies \mathcal{F}.$$

L'ouvert Tr_N^r de $\overline{\text{Tr}}_N^r$ est l'image réciproque par le morphisme

$$((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1})) : \overline{\text{Tr}}_N^r \rightarrow [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$$

de la strate ouverte $\text{Spec}(\mathbb{F}_q) = [\mathbb{G}_m^{r-1}/\mathbb{G}_m^{r-1}] \subset [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$.

THÉORÈME. — *Considérons le morphisme*

$$\overline{\text{Cht}}^r \times_{X^2} (X - N)^2 \rightarrow \overline{\text{Tr}}_N^r \times (X - N)^2$$

de composante le morphisme précédant et la projection canonique.

Pour tout entier d et tout paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ assez convexe par rapport à X et N , la restriction

$$\overline{\text{Cht}}^{r, d; \leq p} \times_{X^2} (X - N)^2 \rightarrow \overline{\text{Tr}}_N^r \times (X - N)^2$$

de ce morphisme à l'ouvert $\overline{\text{Cht}}^{r, d; \leq p} \subset \overline{\text{Cht}}^r$ est lisse purement de dimension relative $2r - 2$.

Le champ $\overline{\text{Tr}}_N^r$ est le produit fibré

$$\overline{\text{Tr}}_N^r = [\tilde{\text{gl}}_r / {}^\tau\text{Ad}(\text{GL}_r)] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$$

où $\tilde{\text{gl}}_r$ est le schéma des endomorphismes complets du \mathbb{F}_q -espace vectoriel \mathbb{F}_q^r , où le morphisme $[\tilde{\text{gl}}_r / {}^\tau\text{Ad}(\text{GL}_r)] \rightarrow [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$ est induit par le morphisme canonique $\tilde{\text{gl}}_r \rightarrow [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$ et où $\langle q-1 \rangle$ est l'endomorphisme de $[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$ induit par l'élevation à la puissance $q-1$ des coordonnées.

Lafforgue définit alors le champ $\overline{\text{Cht}}_N^{r,d;\leq p}$ de *chtoucas itérés avec structure de niveau* N par le carré cartésien

$$\begin{array}{ccc} \overline{\text{Cht}}_N^{r,d;\leq p} & \longrightarrow & \overline{\text{Tr}}_N^{r,\tau} \\ \downarrow & \square & \downarrow \\ \overline{\text{Cht}}^{r,d;\leq p} \times_{X^2} (X - N)^2 & \longrightarrow & \overline{\text{Tr}}_N^r \end{array}$$

où la flèche verticale de droite

$$\begin{aligned} \overline{\text{Tr}}_N^{r,\tau} &:= [\tilde{\text{gl}}_r^\tau / \text{GL}_r] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}] \\ &\rightarrow [\tilde{\text{gl}}_r / {}^\tau\text{Ad}(\text{GL}_r)] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}] = \overline{\text{Tr}}_N^r \end{aligned}$$

est induite par un morphisme projectif équivariant $\bar{L} : \tilde{\text{gl}}_r^\tau \rightarrow \tilde{\text{gl}}_r$ qui prolonge l'isogénie de Lang $L : \text{GL}_r \rightarrow \text{GL}_r$.

Par construction, le schéma $\tilde{\text{gl}}_r^\tau$ est muni d'un morphisme lisse vers un « champ torique » (voir la sous-section suivante) $\mathcal{J}^{r,\tau}$, qui relève le morphisme $\tilde{\text{gl}}_r \rightarrow [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$. On a donc un morphisme lisse de champs

$$\overline{\text{Tr}}_N^{r,\tau} \rightarrow \mathcal{J}^{r,\tau} \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$$

et il résulte du théorème ci-dessus que $\overline{\text{Cht}}_N^{r,d;\leq p}$ est lisse sur le champ

$$\left(\mathcal{J}^{r,\tau} \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}] \right) \times (X - N)^2.$$

En particulier, comme on sait résoudre les singularités d'un champ torique en raffinant l'éventail qui le définit, on sait aussi résoudre les singularités du champ $\overline{\text{Cht}}_N^{r,d;\leq p}$.

La section suivante est consacré à la construction de $\tilde{\text{gl}}_r^\tau$.

4. COMPACTIFICATION DU CLASSIFIANT DE PGL_r

4.1. Champ des pavages

Pour compactifier l'isogénie de Lang pour GL_r , Lafforgue est amené à construire une famille de champs toriques $(\mathcal{J}^{r,n})_{n \geq 0}$ qu'il appelle les *champs des pavages*.

Une *variété torique* est un triplet formé d'un tore T , d'une immersion ouverte $T \hookrightarrow \bar{T}$ d'image dense de T dans un schéma de type fini normal \bar{T} , et d'une action de T

sur \bar{T} qui prolonge l'action de T sur lui-même par translation. Une telle variété torique est associée à un éventail. Rappelons tout d'abord rapidement cette construction.

Soient Y un \mathbb{Z} -module libre de rang fini et $Y_{\mathbb{R}} = \mathbb{R} \otimes_{\mathbb{Z}} Y$. Un *cône convexe polyédral rationnel* est une partie localement fermée σ de Y de la forme

$$\sigma = \mathbb{R}_{>0} y_1 + \cdots + \mathbb{R}_{>0} y_n \subset Y_{\mathbb{R}}$$

pour des vecteurs $y_1, \dots, y_n \in Y \subset Y_{\mathbb{R}}$. L'adhérence $\bar{\sigma}$ de σ est le cône convexe fermé

$$\sigma = \mathbb{R}_{\geq 0} y_1 + \cdots + \mathbb{R}_{\geq 0} y_n \subset Y_{\mathbb{R}}.$$

On définit les faces de σ comme les cônes convexes polyédraux rationnels τ de la forme $\bar{\tau} = H \cap \bar{\sigma}$ pour un hyperplan H de $Y_{\mathbb{R}}$ ne rencontrant pas σ . Un *éventail* est une famille finie E de cônes convexes polyédraux rationnels deux à deux disjoints, appelés les *cellules* de E et ayant les propriétés suivantes :

- toute face d'une cellule de E est encore une cellule de E ,
- tout couple (σ_1, σ_2) de cellules de E admet une face commune $\tau \in E$ telle que $\bar{\sigma}_1 \cap \bar{\sigma}_2 = \bar{\tau}$.
- le cône convexe polyédral rationnel $\{0\}$ est une cellule de E .

Si $\sigma = \mathbb{R}_{>0} y_1 + \cdots + \mathbb{R}_{>0} y_n \subset Y_{\mathbb{R}}$ est un cône convexe polyédral rationnel ne contenant aucune droite vectorielle de $Y_{\mathbb{R}}$, la collection formée de σ et de toutes ses faces est un éventail. On associe à cet éventail particulier la variété torique affine

$$T = Y \otimes_{\mathbb{Z}} \mathbb{G}_m = \text{Spec}(\mathbb{F}_q[Y^{\vee}]) \hookrightarrow \bar{T}(\sigma) = \text{Spec}(\mathbb{F}_q[Y^{\vee} \cap \sigma^{\vee}])$$

où $Y^{\vee} := \text{Hom}_{\mathbb{Z}}(Y, \mathbb{Z})$ et

$$\sigma^{\vee} = \{y^{\vee} : Y_{\mathbb{R}} \rightarrow \mathbb{R} \mid y^{\vee}(y_1), \dots, y^{\vee}(y_n) \geq 0\} \subset Y_{\mathbb{R}}^{\vee}$$

est le cône fermé dual de σ . On vérifie que $Y^{\vee} \cap \sigma^{\vee}$ est un monoïde à engendrement fini, saturé et contenant un système de générateurs du groupe Y^{\vee} . L'exemple type est bien entendu le plongement torique affine standard $\mathbb{G}_m^{r-1} \hookrightarrow \mathbb{A}^{r-1}$ qui est obtenu par la construction ci-dessus pour $Y = \mathbb{Z}^{r-1}$ et $\sigma = (\mathbb{R}_{>0})^{r-1} \subset \mathbb{R}^{r-1} = Y_{\mathbb{R}}$.

Si τ est une face de σ , on a $\sigma^{\vee} \subset \tau^{\vee}$ et l'immersion ouverte $T \hookrightarrow \bar{T}(\sigma)$ se factorise par l'immersion ouverte $T \hookrightarrow \bar{T}(\tau)$ en

$$T \hookrightarrow \bar{T}(\tau) \hookrightarrow \bar{T}(\sigma).$$

En particulier, si σ_1 et σ_2 sont deux cellules de E et si $\bar{\tau} = \bar{\sigma}_1 \cap \bar{\sigma}_2$, on a une donnée de recollement

$$\bar{T}(\sigma_1) \hookrightarrow \bar{T}(\tau) \hookrightarrow \bar{T}(\sigma_2)$$

pour les variétés toriques affines $T \hookrightarrow \bar{T}(\sigma_1)$ et $T \hookrightarrow \bar{T}(\sigma_2)$.

Le variété torique $T \hookrightarrow \bar{T}$ associée à l'éventail E est obtenue en recollant les variétés toriques affines $T \hookrightarrow \bar{T}(\sigma)$ pour $\sigma \in E$ à l'aide des données de recollements ci-dessus.

Les orbites pour l'action de T sur \overline{T} sont en nombre fini et forment une stratification en parties localement fermées de \overline{T} . On peut indexer ces orbites par les cellules de E : l'orbite \overline{T}_σ est contenue dans la carte affine $\overline{T}(\sigma)$ et est en fait l'unique orbite fermée pour l'action de T sur cette variété torique affine. L'orbite \overline{T}_σ est dans l'adhérence de l'orbite \overline{T}_τ si et seulement si τ est une face de σ . En particulier, $\overline{T}_{\{0\}} = T$ est l'unique orbite ouverte dans \overline{T} . Chaque orbite \overline{T}_σ a par construction un point marqué \bar{t}_σ .

Le *champ torique* associé à l'éventail E est le champ algébrique quotient

$$\overline{\mathcal{T}} = [\overline{T}/T].$$

Ce champ algébrique, qui est normal, de type fini et de dimension 0, n'a qu'un nombre fini de points, à savoir les points définis par les $\bar{t}_\sigma \in \overline{T}$. La gerbe résiduelle d'un tel point est le classifiant du fixateur de \bar{t}_σ dans T .

Soient maintenant k un corps et n un entier ≥ 0 . Considérons l'espace affine $\mathbb{R}^{r,n} = \{x = (x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_0 + x_1 + \dots + x_n = r\}$ et son réseau $\mathbb{Z}^{r,n} = \{x = (x_0, \dots, x_n) \in \mathbb{Z}^{n+1} \mid x_0 + x_1 + \dots + x_n = r\}$. On a le simplexe

$$S_{\mathbb{R}}^{r,n} = \{x = (x_0, \dots, x_n) \in (\mathbb{R}_{\geq 0})^{n+1} \mid x_0 + x_1 + \dots + x_n = r\} \subset \mathbb{R}^{r,n}$$

et l'ensemble de ses points entiers

$$S^{r,n} = \mathbb{Z}^{r,n} \cap S_{\mathbb{R}}^{r,n}.$$

On appelle *pavé entier* toute partie P de $S_{\mathbb{R}}^{r,n}$ d'intérieur P° non vide et de la forme

$$P = \{x \in \mathbb{R}^{r,n} \mid \sum_{j \in J} x_j \geq d_J, \forall J\}$$

pour (d_J) une famille d'entiers indexés par les parties J de $\{0, 1, \dots, n\}$ qui est convexe au sens où

$$d_\emptyset = 0, \quad d_{\{0,1,\dots,n\}} = r \quad \text{et} \quad d_{J'} + d_{J''} \leq d_{J' \cup J''} + d_{J' \cap J''}, \quad \forall J', J''.$$

On appelle *pavage entier* de $S_{\mathbb{R}}^{r,n}$ toute famille \mathbf{P} de pavés entiers telle que

$$S_{\mathbb{R}}^{r,n} = \bigcup_{P \in \mathbf{P}} P$$

et que $P_1^\circ \cap P_2^\circ = \emptyset$ pour tous $P_1 \neq P_2$ dans \mathbf{P} .

Une application $S^{r,n} \rightarrow \mathbb{R}$ est dite affine si elle est la restriction d'une application affine $\mathbb{R}^{r,n} \rightarrow \mathbb{R}$. Soient $Y_{\mathbb{R}}^{r,n}$ le \mathbb{R} -espace vectoriel quotient de $\mathbb{R}^{S^{r,n}} = \{S^{r,n} \rightarrow \mathbb{R}\}$ par le sous-espace des applications affines, et $Y^{r,n}$ l'image dans ce quotient du réseau $\mathbb{Z}^{S^{r,n}} \subset \mathbb{R}^{S^{r,n}}$.

Pour tout pavage entier \mathbf{P} de $S_{\mathbb{R}}^{r,n}$, notons $\sigma_{\mathbf{P}} \subset Y_{\mathbb{R}}$ l'ensemble des classes y de fonctions $S^{r,n} \rightarrow \mathbb{R}$ qui, pour chaque pavé $P \in \mathbf{P}$, admettent un représentant $y_P : S^{r,n} \rightarrow \mathbb{R}$ à valeurs positives tel que

$$P = \{x \in S_{\mathbb{R}}^{r,n} \mid y_P(x) = 0\}.$$

Le pavage entier \mathbf{P} est dit *admissible* si $\sigma_{\mathbf{P}}$ est non vide.

LEMME (Lafforgue). — Pour chaque pavage entier admissible \mathbf{P} de $S_{\mathbb{R}}^{r,n}$, $\sigma_{\mathbf{P}}$ est un cône convexe polyédral rationnel dans $Y_{\mathbb{R}}$. De plus, $\sigma_{\mathbf{Q}}$ est une face de $\sigma_{\mathbf{P}}$ si et seulement le pavage entier admissible \mathbf{Q} est plus grossier que \mathbf{P} .

La famille des $\sigma_{\mathbf{P}}$, pour \mathbf{P} parcourant l'ensemble des pavages entiers admissibles de $S_{\mathbb{R}}^{r,n}$, est un éventail.

On appellera *champ des pavages* le k -champ torique

$$\mathcal{T}^{r,n} = [\bar{T}^{r,n}/T^{r,n}]$$

associé à cet éventail. On vérifie que l'on a une suite exacte

$$1 \rightarrow \mathbb{G}_{m,k} \rightarrow \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k} \rightarrow \mathbb{G}_{m,k}^{S^{r,n}} \rightarrow T^{r,n} \rightarrow 1$$

où la deuxième flèche envoie z sur $(z, \dots, z; z^r)$ et la troisième est donnée par

$$(u_0, u_1, \dots, u_n; z) \mapsto (u_0^{i_0} u_1^{i_1} \cdots u_n^{i_n} z^{-1})_{i \in S^{r,n}}.$$

Par construction, le k -schéma $\bar{T}^{r,n}$ est normal de type fini et l'action de $T^{r,n}$ sur $\bar{T}^{r,n}$ n'a qu'un nombre fini d'orbites. Ces orbites sont indexées par les pavages entiers admissibles de $S^{r,n}$ et chacune a un point marqué $\bar{t}_{\mathbf{P}}$. Le champ des pavages $\mathcal{T}^{r,n}$ est donc algébrique normal de type fini et n'a qu'un nombre fini de points définis par les points $\bar{t}_{\mathbf{P}}$. Le point défini par $\bar{t}_{\mathbf{P}}$ est dans l'adhérence du point défini par $\bar{t}_{\mathbf{Q}}$ si le pavage entier admissible \mathbf{P} raffine le pavage entier admissible \mathbf{Q} . Le point ouvert correspond au pavage le plus grossier, c'est-à-dire au pavage trivial (un seul pavé).

Exemples : (i) Pour $n = 0$, $S_{\mathbb{R}}^{r,n} = S^{r,n} = \{0\}$ et le champ des pavages est simplement $\text{Spec}(k)$. Pour $n = 1$ on peut identifier $S_{\mathbb{R}}^{r,n}$ à l'intervalle $[0, r] \subset \mathbb{R}$ et les pavés entiers sont précisément les intervalles $[i, j]$ où $i < j$ sont des entiers compris entre 0 et r . Un pavage entier est donc une décomposition

$$[0, r] = [0, r_1] \cup [r_1, r_1 + r_2] \cup \cdots \cup [r_1 + \cdots + r_{s-1}, r_1 + \cdots + r_{s-1} + r_s]$$

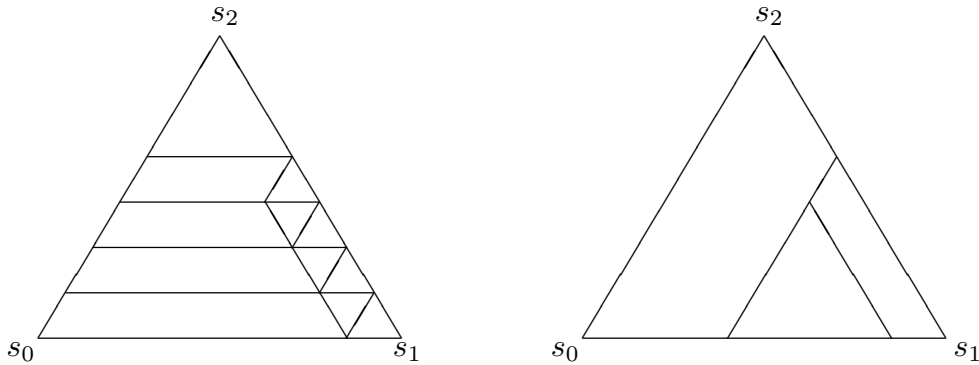
de l'intervalle $[0, r]$ où $r_1 + \cdots + r_s = r$ est une décomposition de r en entiers strictement positifs. Le champ des pavages n'est autre dans ce cas que le champ torique $[\mathbb{A}_k^{r-1}/\mathbb{G}_{m,k}^{r-1}]$.

(ii) Pour $n = 2$ on peut identifier $S_{\mathbb{R}}^{r,n}$ à un triangle équilatéral dans \mathbb{R}^2 dont les trois côtés sont de longueur r . On numérote les sommets $s_0 = s_3, s_1, s_2$ de ce triangle et, pour $i = 0, 1, 2$, on désigne par x_i la distance au sommet s_i d'un point du côté $s_i s_{i+1}$. Par tout point de $S_{\mathbb{R}}^{r,n}$ passent trois droites D_{12}, D_{20} et D_{01} parallèles aux côtés du triangle $s_1 s_2, s_2 s_0$ et $s_0 s_1$, et on peut repérer un tel point par les coordonnées x_0, x_1 et x_2 des intersections $D_{12} \cap s_0 s_1, D_{20} \cap s_1 s_2$ et $D_{01} \cap s_2 s_0$. Par définition, $S^{r,n} \subset S_{\mathbb{R}}^{r,n}$ est formé des points à coordonnées (x_0, x_1, x_2) entières.

Appelons droite entière toute droite qui est parallèle à l'un des trois côté du triangle $S_{\mathbb{R}}^{r,n}$ et qui passe par un point de $S^{r,n}$; appelons segment entier tout segment d'une droite entière dont les deux extrémités sont des points de $S^{r,n}$.

Si l'on trace toutes les droites entières, on obtient un pavage constitué de petits triangles équilatéraux de côtés de longueur 1 qui est entier admissible. Il est plus fin que tout autre pavage entier : tout pavé entier est une réunion d'une famille de ces petits triangles équilatéraux et est un hexagone, éventuellement dégénéré, dont les côtés sont des segments entiers. La variété torique $T^{r,n} \hookrightarrow \overline{T}^{r,n}$ est donc affine (Cette dernière propriété n'est plus satisfaite dès que $n \geq 3$).

Des deux pavages entiers ci-dessous, seul celui de gauche (qui joue un rôle dans la compactification de l'isogénie de Lang) est admissible.



(iii) Si $r = 2$, le champ torique $\mathcal{T}^{r,n}$ est lisse, mais ce n'est plus le cas dès que $n \geq 2$ et $r \geq 3$ □

Pour chaque entier $0 \leq m < n$ et chaque application injective $\{0, 1, \dots, m\} \hookrightarrow \{0, 1, \dots, n\}$, on a une identification induite de $S_{\mathbb{R}}^{r,m}$ à une face de $S_{\mathbb{R}}^{r,n}$ et tout pavage entier admissible de $S_{\mathbb{R}}^{r,n}$ induit par restriction un pavage entier admissible de $S_{\mathbb{R}}^{r,m}$. Dualement, on en déduit des projections $T^{r,n} \rightarrow T^{r,m}$, $\overline{T}^{r,n} \rightarrow \overline{T}^{r,m}$ et $\mathcal{T}^{r,n} \rightarrow \mathcal{T}^{r,m}$.

4.2. Compactifications des $\mathrm{PGL}_r^{n+1} / \mathrm{PGL}_r$

Considérons un espace vectoriel V de dimension r sur un corps k , que l'on verra aussi comme un k -schéma affine isomorphe à \mathbb{A}_k^r , et notons $G = \mathrm{GL}(V)$ le k -schéma des automorphismes linéaires de V . Le k -schéma en groupes G^{n+1} agit de manière évidente sur V^{n+1} et donc sur $\bigwedge^r V^{n+1}$.

On peut décomposer la représentation $\bigwedge^r V^{n+1}$ de G^{n+1} en la somme directe de représentations irréductibles

$$\bigwedge^r V^{n+1} = \bigoplus_{i \in S^{r,n}} \bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \dots \otimes \bigwedge^{i_n} V$$

où $S^{r,n}$ est l'ensemble des points entiers du simplexe introduit dans la sous-section précédente. Le tore $\mathbb{G}_{m,k}^{S^{r,n}}$ agit sur $\bigwedge^r V^{n+1}$ par

$$(t = (t_i)_{i \in S^{r,n}}, x = \bigoplus_{i \in S^{r,n}} x_i) \mapsto \bigoplus_{i \in S^{r,n}} t_i x_i,$$

et cette action commute à celle de G^{n+1} . On en déduit une action du k -schéma en groupes produit $G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}}$ sur l'espace projectif $\mathbb{P}(\bigwedge^r V^{n+1})$ des droites de $\bigwedge^r V^{n+1}$.

On envoie le produit $G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}$ dans $G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}}$ par

$$(g; u_0, u_1, \dots, u_n; z) \mapsto (u_0^{-1}g, u_1^{-1}g, \dots, u_n^{-1}g; (u_0^{i_0} u_1^{i_1} \cdots u_n^{i_n} z^{-1})_{i \in S^{r,n}}).$$

L'homomorphisme ainsi défini n'est pas injectif (il admet pour noyau $\mathbb{G}_{m,k}$ plongé par $z \mapsto (z \text{Id}_V; z, \dots, z, z^r)$), mais on notera quand même $(G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}})/(G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k})$ le quotient de $G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}}$ par l'image de cet homomorphisme.

Le plongement diagonal $V \hookrightarrow V^{n+1}$ induit un plongement

$$\bigwedge^r V \hookrightarrow \bigwedge^r V^{n+1}$$

et définit un point $\omega \in \mathbb{P}(\bigwedge^r V^{n+1})$ dont le fixateur est précisément l'image de $G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}$ dans $G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}}$. On a donc un plongement localement fermé de k -schémas

$$(G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}})/(G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}) \hookrightarrow \mathbb{P}\left(\bigwedge^r V^{n+1}\right), (g, t) \mapsto (g, t) \cdot \omega.$$

Par construction l'image de ce plongement est contenue dans l'ouvert

$$\left[\prod_{i \in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right) - \{0\} \right) \right] / \mathbb{G}_{m,k} \subset \mathbb{P}\left(\bigwedge^r V^{n+1}\right).$$

des droites de $\bigwedge^r V^{n+1}$ dont la projection sur chaque facteur direct $\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V$ n'est pas nulle.

Considérons alors le morphisme

$$(G^{n+1} \times \mathbb{G}_{m,k}^{S^{r,n}})/(G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}) \hookrightarrow \overline{T}^{r,n} \times_k \left[\prod_{i \in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right) - \{0\} \right) \right] / \mathbb{G}_{m,k}$$

dont les composantes sont le composé de la projection

$$(G^{n+1} \times_k \mathbb{G}_{m,k}^{S^{r,n}})/(G \times \mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}) \twoheadrightarrow \mathbb{G}_{m,k}^{S^{r,n}}/(\mathbb{G}_{m,k}^{n+1} \times \mathbb{G}_{m,k}) = T^{r,n}$$

et de l'immersion ouverte

$$T^{r,n} \hookrightarrow \overline{T}^{r,n},$$

et l'immersion localement fermée ci-dessus. Ce morphisme est une immersion localement fermée. Notons

$$\Omega^n(V) \subset \left(\left[\prod_{i \in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right) - \{0\} \right) \right] / \mathbb{G}_{m,k} \right) \times_k \bar{T}^{r,n}$$

l'adhérence schématique de son image. et

$$\bar{\Omega}^n(V) \rightarrow \prod_{i \in S^{r,n}} \mathbb{P} \left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right)$$

le quotient de $\Omega^n(V)$ par l'action libre du tore $\mathbb{G}_{m,k}^{S^{r,n}} / \mathbb{G}_{m,k}$ (où $\mathbb{G}_{m,k}$ est plongé diagonalement dans $\mathbb{G}_{m,k}^{S^{r,n}}$).

THÉORÈME. — (i) *La première projection*

$$\Omega^n(V) \rightarrow \left[\prod_{i \in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right) - \{0\} \right) \right] / \mathbb{G}_{m,k}$$

et son quotient par les actions libres du tore $\mathbb{G}_{m,k}^{S^{r,n}} / \mathbb{G}_{m,k}$

$$\bar{\Omega}^n(V) \rightarrow \prod_{i \in S^{r,n}} \mathbb{P} \left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V \right)$$

sont projectives.

(ii) *La seconde projection*

$$\Omega^n(V) \rightarrow \bar{T}^{r,n},$$

qui est équivariante relativement au morphisme de tores $\mathbb{G}_{m,k}^{S^{r,n}} / \mathbb{G}_{m,k} \rightarrow T^{r,n}$, est lisse purement de dimension relative nr^2 ; le morphisme de champs qu'on en déduit par passage au quotient

$$\bar{\Omega}^n(V) = \Omega^n(V) / (\mathbb{G}_{m,k}^{S^{r,n}} / \mathbb{G}_{m,k}) \rightarrow [\bar{T}^{r,n} / T^{r,n}] = \mathcal{T}^{r,n}$$

est lisse purement de dimension relative $n(r^2 - 1)$.

Plus généralement, pour tout entier $0 \leq m < n$ et toute application injective $\{0, 1, \dots, m\} \hookrightarrow \{0, 1, \dots, n\}$, l'inclusion $S^{r,m} \hookrightarrow S^{r,n}$ et la projection $G^{n+1} \rightarrow G^{m+1}$ correspondantes induisent un épimorphisme de tores $T^{r,n} \rightarrow T^{r,m}$, un prolongement torique $\bar{T}^{r,n} \rightarrow \bar{T}^{r,m}$ de cet épimorphisme et un morphisme de schémas $\Omega^n(V) \rightarrow \Omega^m(V)$ dit de face au-dessus de ce morphisme torique, et les morphismes

$$\Omega^n(V) \rightarrow \Omega^m(V) \times_{\bar{T}^{r,m}} \bar{T}^{r,n} \quad \text{et} \quad \bar{\Omega}^n(V) \rightarrow \bar{\Omega}^m(V) \times_{\mathcal{T}^{r,m}} \mathcal{T}^{r,n}$$

que l'on en déduit, sont lisses purement de dimensions relatives $(n - m)r^2$ et $(n - m)(r^2 - 1)$ respectivement.

Le k -schéma $\bar{\Omega}^n(V)$ est donc une compactification projective et toroïdale de \bar{G}^{n+1} / \bar{G} où on a noté $\bar{G} = \text{PGL}(V)$. Les $\bar{\Omega}^n(V)$ et les morphismes de face $\bar{\Omega}^n(V) \rightarrow \bar{\Omega}^m(V)$ sont sous-jacents à un k -schéma simplicial $\Omega^\bullet(V)$ qui contient comme ouvert dense le k -schéma simplicial classifiant $B(\bar{G}) = \bar{G}^{\bullet+1} / \bar{G}$.

Remarques : (i) Pour $n = 0$, $\Omega^n(V)$ n'est autre que $\text{Spec}(k)$. Pour $n = 1$, le carré commutatif

$$\begin{array}{ccc}
 (\text{GL}(V)^2 \times_k \mathbb{G}_{m,k}^{r+1}) / (\text{GL}(V) \times_k \mathbb{G}_{m,k}^2 \times_k \mathbb{G}_{m,k}) & \xrightarrow{\sim} & \text{GL}(V) \times_k \mathbb{G}_{m,k}^{r-1} \\
 \downarrow & + & \downarrow \\
 \mathbb{G}_{m,k}^{r+1} / (\mathbb{G}_{m,k}^2 \times_k \mathbb{G}_{m,k}) & \xrightarrow{\sim} & \mathbb{G}_{m,k}^{r-1}
 \end{array}$$

où les deux flèches horizontales

$$(g_0, g_1, (\mu_\rho)_{\rho=0, \dots, r}) \mapsto \left(\frac{\mu_1}{\mu_0} g_0 g_1^{-1}, \frac{\mu_0 \mu_2}{\mu_1^2}, \dots, \frac{\mu_{r-2} \mu_r}{\mu_{r-1}^2} \right)$$

et

$$(\mu_\rho)_{\rho=0, 1, \dots, r} \mapsto \left(\frac{\mu_0 \mu_2}{\mu_1^2}, \dots, \frac{\mu_{r-2} \mu_r}{\mu_{r-1}^2} \right)$$

sont des isomorphismes, se prolonge en un carré commutatif

$$\begin{array}{ccc}
 \Omega^1(V) & \xrightarrow{\sim} & \Omega(V, V) \\
 \downarrow & + & \downarrow \\
 \overline{T}^{r,1} & \xrightarrow{\sim} & \mathbb{A}_k^{r-1}
 \end{array}$$

où, de nouveau, les deux flèches horizontales sont des isomorphismes. (La première flèche verticale est celle du théorème ci-dessus et la seconde $\Omega(V, V) \rightarrow \mathbb{A}_k^{r-1}$ a été construite dans la section 3.) En particulier, on peut identifier le k -schéma $\widetilde{\text{End}}(V)$ des endomorphismes complets de V à un quotient de $\Omega^1(V)$ par un tore $\mathbb{G}_{m,k}^r$ et, en divisant en plus par les homothéties, on en déduit un isomorphisme

$$\overline{\Omega}^1(V) \xrightarrow{\sim} \widetilde{\text{End}}(V) / \mathbb{G}_{m,k}$$

qui échange les projections

$$\overline{\Omega}^1(V) \rightarrow \mathcal{T}^{r,n} \text{ et } \widetilde{\text{End}}(V) / \mathbb{G}_{m,k} \rightarrow [\mathbb{A}_k^{r-1} / \mathbb{G}_{m,k}^{r-1}].$$

Le k -schéma projectif $\overline{\Omega}^1(V)$ est donc la compactification de De Concini et Procesi de $\text{PGL}(V)$.

(ii) Le morphisme $\bar{\Omega}^n(V) \rightarrow (\bar{\Omega}^1(V))^{n+1}$ produit des morphismes induits par les applications injectives $\{0, 1\} \hookrightarrow \{0, 1, \dots, n\}$ données par $\{0, 1\} \mapsto \{0, 1\}$, $\{0, 1\} \mapsto \{1, 2\}, \dots, \{0, 1\} \mapsto \{n, 0\}$, envoie \bar{G}^{n+1}/\bar{G} sur le fermé d'équation $g_0 g_1 \cdots g_n = 1$ dans $(\bar{G}^2/\bar{G})^{n+1} \cong \bar{G}^{n+1}$. Ainsi, $\bar{\Omega}^n(V)$ réalise une « compactification » du morphisme de multiplication de n éléments dans \bar{G} . \square

Rappelons que le k -schéma $\bar{T}^{r,n}$ est stratifié par les orbites de $T^{r,n}$ et que ces orbites sont indexées par les pavages entiers admissibles \mathbf{P} de $S_{\mathbb{R}}^{r,n}$ et possèdent chacune un point marqué $\bar{t}_{\mathbf{P}} \in \bar{T}^{r,n}$. L'image réciproque par la projection $\Omega^n(V) \rightarrow \bar{T}^{r,n}$ de cette stratification est évidemment une stratification de $\Omega^n(V)$ par des parties localement fermées $\Omega_{\mathbf{P}}^n(V)$, indexées aussi par les pavages entiers admissibles \mathbf{P} de $S_{\mathbb{R}}^{r,n}$. L'inclusion dans la strate $\Omega_{\mathbf{P}}^n(V)$ de la fibre $\Omega^n(V)_{\bar{t}_{\mathbf{P}}}$ de $\Omega^n(V) \rightarrow \bar{T}^{r,n}$ en $\bar{t}_{\mathbf{P}}$ induit un isomorphisme de schémas

$$(\Omega^n(V)_{\bar{t}_{\mathbf{P}}} \times_k \mathbb{G}_{\mathbf{m},k}^{S_{r,n}}) / (\mathbb{G}_{\mathbf{m},k}^{S_{r,n}})_{\bar{t}_{\mathbf{P}}} \xrightarrow{\sim} \Omega_{\mathbf{P}}^n(V)$$

où $(\mathbb{G}_{\mathbf{m},k}^{S_{r,n}})_{\bar{t}_{\mathbf{P}}}$ est le stabilisateur de $\bar{t}_{\mathbf{P}}$ dans $\mathbb{G}_{\mathbf{m},k}^{S_{r,n}}$.

En particulier, si \mathbf{P} est le pavage trivial, $\Omega^n(V)_{\bar{t}_{\mathbf{P}}} = G^{n+1}/G$ et $\Omega_{\mathbf{P}}^n$ est l'ouvert $(G^{n+1} \times \mathbb{G}_{\mathbf{m},k}^{S_{r,n}}) / (G \times \mathbb{G}_{\mathbf{m},k}^{n+1} \times \mathbb{G}_{\mathbf{m},k})$ de $\Omega^n(V)$.

Lafforgue appelle *k-schéma des graphes recollés dans V^{n+1}* associé à un pavage entier admissible \mathbf{P} de $S_{\mathbb{R}}^{r,n}$, et note $\text{Gr}_{\mathbf{P}}^n(V)$, le schéma de modules des familles $(W_P)_{P \in \mathbf{P}}$ de sous-espaces de dimension r dans V^{n+1} indexés par les pavés P de \mathbf{P} telles que :

- pour tout pavé $P \in \mathbf{P}$ et tout $J \subset \{0, 1, \dots, n\}$, on a

$$\dim(W_P \cap V^J) = \min_{i \in P \cap S_{r,n}} \left\{ \sum_{j \in J} i_j \right\},$$

- pour tous pavés $P', P'' \in \mathbf{P}$ ayant en commun un bord d'équation $\sum_{j \in J} x_j = d_J$ pour un certain $J \subset \{0, 1, \dots, n\}$ et

$$d_J = \min_{i' \in P' \cap S_{r,n}} \left\{ \sum_{j \in J} i'_j \right\} = \max_{i'' \in P'' \cap S_{r,n}} \left\{ \sum_{j \in J} i''_j \right\},$$

on a

$$i_J^{-1}(W_{P'}) = \text{pr}_J(W_{P''}) \text{ et } \text{pr}_{J^c}(W_{P'}) = i_{J^c}^{-1}(W_{P''})$$

où J^c est le complémentaire de J dans $\{0, 1, \dots, n\}$ et $i_J : V^J \hookrightarrow V^{n+1}$, $i_{J^c} : V^{J^c} \hookrightarrow V^{n+1}$ et $\text{pr}_J : V^{n+1} \twoheadrightarrow V^J$, $\text{pr}_{J^c} : V^{n+1} \twoheadrightarrow V^{J^c}$ sont les inclusions et projections canoniques.

THÉORÈME. — *Pour chaque pavage entier admissible \mathbf{P} de $S_{\mathbb{R}}^{r,n}$, les coordonnées de Plücker induisent un isomorphisme du k -schéma des graphes recollés $\mathrm{Gr}_{\mathbf{P}}^n(V)$ sur la fibre $\Omega^n(V)_{\bar{t}_{\mathbf{P}}}$.*

4.3. Compactification de l'isogénie de Lang

Le corps de base est de nouveau \mathbb{F}_q . On notera encore simplement τ l'endomorphisme de Frobenius d'un schéma.

Rappelons que l'on a défini un éventail $(\sigma_{\mathbf{P}})_{\mathbf{P}}$ dans l'espace vectoriel réel $Y_{\mathbb{R}}^{r,2}$ quotient de $\mathbb{R}^{S^{r,2}} = \{S^{r,2} \rightarrow \mathbb{R}\}$ par le sous-espace des applications affines et que l'on a noté $T^{r,2} \hookrightarrow \bar{T}^{r,2}$ la variété torique correspondante.

Considérons maintenant le sous-espace de $\mathbb{R}^{S^{r,2}}$ formé des applications

$$y : S^{r,2} = \{(i_0, i_1, i_2) \in \mathbb{N}^3 \mid i_0 + i_1 + i_2 = r\} \rightarrow \mathbb{R}$$

telles que

$$y(0, i_1, i_2) = qy(i_1, 0, i_2), \quad \forall (0, i_1, i_2) \in S^{r,2},$$

et son image $Y_{\mathbb{R}}^{r,\tau}$ dans $Y_{\mathbb{R}}^{r,2}$. On remarquera que la donnée d'une telle application y équivaut à la donnée de sa restriction à

$$S^{r,\tau} = \{(i_0, i_1, i_2) \in \mathbb{N}^3 \mid i_0 + i_1 + i_2 = r \text{ et } i_0 \neq 0\} \subset S^{r,2}$$

Appelons *pavage entier q -admissible* tout pavage \mathbf{P} de $S_{\mathbb{R}}^{r,2}$ tel que le cône convexe polyédral rationnel $\sigma_{\mathbf{P}} \subset Y_{\mathbb{R}}^{r,2}$ rencontre le sous-espace $Y_{\mathbb{R}}^{r,\tau}$. La famille des intersections $\sigma_{\mathbf{P}}^{\tau} := \sigma_{\mathbf{P}} \cap Y_{\mathbb{R}}^{r,\tau}$ pour \mathbf{P} parcourant l'ensemble des pavages entiers q -admissibles est un éventail dans $Y_{\mathbb{R}}^{r,\tau}$ et définit donc une variété torique normale

$$T^{r,\tau} \hookrightarrow \bar{T}^{r,\tau}.$$

De la même façon que l'on a une suite exacte

$$1 \rightarrow (\mathbb{G}_m^3 \times \mathbb{G}_m) / \mathbb{G}_m \rightarrow \mathbb{G}_m^{S^{r,2}} \rightarrow T^{r,2} \rightarrow 1$$

on a une suite exacte

$$1 \rightarrow \mathbb{G}_m \rightarrow \mathbb{G}_m^{S^{r,\tau}} \rightarrow T^{r,\tau} \rightarrow 1$$

où la deuxième flèche envoie u sur $(u^{i_0+qi_1})_{i \in S^{r,\tau}}$. On vérifie que le plongement

$$\mathbb{G}_m^{S^{r,\tau}} \hookrightarrow \mathbb{G}_m^{S^{r,2}}$$

qui envoie $(t_i)_{i \in S^{r,\tau}}$ sur $(t_i)_{i \in S^{r,2}}$, où $t_{(0,i_1,i_2)} = t_{(i_1,0,i_2)}^q$ si $i_1 \neq 0$ et $t_{(0,0,r)} = 1$, induit un plongement $T^{r,\tau} \hookrightarrow T^{r,2}$ qui se prolonge en une immersion fermée

$$\bar{T}^{r,\tau} \hookrightarrow \bar{T}^{r,2}.$$

On identifiera dans la suite $\bar{T}^{r,\tau}$ à l'image de cette immersion fermée.

On a vu que les trois applications strictement croissantes $\{0, 1\} \hookrightarrow \{0, 1, 2\}$, $(0, 1) \mapsto (1, 2), (0, 2), (0, 1)$, induisent trois morphismes toriques $p_0, p_1, p_2 : \overline{T}^{r,2} \rightarrow \overline{T}^{r,1}$ et trois morphismes de face $\pi_0, \pi_1, \pi_2 : \Omega^{r,2} \rightarrow \Omega^{r,1}$ au-dessus de ceux-ci. Par construction, la relation

$$p_0 = \tau \circ p_1$$

est vérifiée sur le fermé $\overline{T}^{r,\tau}$ de $\overline{T}^{r,2}$.

Soit alors

$$\Omega^{r,\tau} \subset \Omega^{r,2} \times_{\overline{T}^{r,2}} \overline{T}^{r,\tau}$$

le sous-schéma fermé défini par l'équation $\pi_0 = \tau \circ \pi_1$ dans $\Omega^{r,1}$. L'action évidente du sous-tore $\mathbb{G}_m^{S^{r,\tau}} \subset \mathbb{G}_m^{S^{r,2}}$ sur $\Omega^{r,2} \times_{\overline{T}^{r,2}} \overline{T}^{r,\tau}$ stabilise ce fermé. On notera π_1^τ, π_2^τ les restrictions des morphismes π_1, π_2 à $\Omega^{r,\tau}$. Elles relèvent les restrictions $p_1^\tau, p_2^\tau : \overline{T}^{r,\tau} \rightarrow \overline{T}^{r,1}$ de p_1, p_2 et sont $\mathbb{G}_m^{S^{r,\tau}}$ -équivariantes.

THÉORÈME. — *La seconde projection $\Omega^{r,\tau} \rightarrow \overline{T}^{r,\tau}$ est $\mathbb{G}_m^{S^{r,\tau}}$ -équivariante et lisse purement de dimension relative r^2 .*

La stratification par les $T^{r,\tau}$ -orbites de $\overline{T}^{r,\tau}$ induit par image réciproque une stratification de $\Omega^{r,\tau}$ dont les strates $\Omega_{\mathbf{P}}^{r,\tau}$ sont indexées par les pavages entiers q -admissibles. Un tel pavage \mathbf{P} induit des pavages entiers convexes admissibles $\mathbf{Q}_0 = \mathbf{Q}_1$ et \mathbf{Q}_2 sur les côtés $\{x_0 = 0\}, \{x_1 = 0\}$ et $\{x_2 = 0\}$ de $S_{\mathbb{R}}^{r,2}$ identifiés à $S_{\mathbb{R}}^{r,1}$ et on a des morphismes de face $p_i : \Omega_{\mathbf{P}}^{r,\tau} \rightarrow \Omega_{\mathbf{Q}_i}^{r,1}$ pour $i = 0, 1, 2$. Ces strates et ces morphismes de face admettent des descriptions en termes des schémas de graphes recollés $\text{Gr}_{\mathbf{P}}^{r,2}$ et $\text{Gr}_{\mathbf{Q}}^{r,1}$.

En particulier, pour le pavage trivial, on vérifie que la strate ouverte $\Omega_{\mathbf{P}}^{r,\tau}$ est isomorphe au fermé

$$(\{(g_0, g_1, g_2) \mid g_1 g_2^{-1} = \tau(g_0 g_2^{-1})\} \times \mathbb{G}_m^{S^{r,\tau}}) / (\text{GL}_r \times \mathbb{G}_m) \subset (\text{GL}_r^3 \times \mathbb{G}_m^{S^{r,\tau}}) / (\text{GL}_r \times \mathbb{G}_m)$$

où le quotient est pris pour le plongement $(g, u) \mapsto ((u^{-1}g, u^{-q}g, g), (u^{i_0+qi_1})_{i \in S^{r,\tau}})$. On peut identifier ce fermé au schéma

$$\text{GL}_r \times (\mathbb{G}_m^{S^{r,\tau}} / \mathbb{G}_m)$$

par le morphisme $((g_0, g_1, g_2), t) \mapsto (g_2 g_0^{-1}, t)$. Si l'on identifie $\Omega^{r,1}$ à $\text{GL}_r \times (\mathbb{G}_m^{S^{r,1}} / \mathbb{G}_m)$ par le morphisme $((g_0, g_1), t) \mapsto (g_1 g_0^{-1}, t)$, les projections $p_1, p_2 : \Omega_{\mathbf{P}}^{r,\tau} \rightarrow \Omega_{\mathbf{Q}}^{r,1}$, où $\mathbf{Q} = \mathbf{Q}_0 = \mathbf{Q}_1 = \mathbf{Q}_2$ est le pavage trivial de $S_{\mathbb{R}}^{r,1}$, sont induites par les morphismes $\text{GL}_r \rightarrow \text{GL}_r$ qui envoient g sur g et sur $\tau(g)^{-1}g$ respectivement et les morphismes $\mathbb{G}_m^{S^{r,\tau}} \rightarrow \mathbb{G}_m^{S^{r,1}}$ qui envoient t sur $(t_{(i_0,0,i_1)})_{i \in S^{r,1}}$ (avec $t_{(0,0,r)} = 1$) et sur $(t_{(i_0,i_1,0)})_{i \in S^{r,1}}$ (avec $t_{(0,r,0)} = t_{r,0,0}^q$) respectivement.

Soit $S^{r,\tau,\circ} = S^{r,\tau} - \{(1, 0, r - 1)\}$. L'inclusion $\mathbb{G}_m^{S^{r,\tau,\circ}} \subset \mathbb{G}_m^{S^{r,\tau}}$ est une section de l'épimorphisme $\mathbb{G}_m^{S^{r,\tau}} \rightarrow T^{r,\tau}$.

THÉORÈME. — *Le quotient $\widetilde{\mathrm{gl}}_r^\tau$ de $\Omega^{r,\tau}$ par l'action libre du tore $\mathbb{G}_m^{S^{r,\tau,0}}$ contient GL_r comme un ouvert dense. Il est lisse sur le champ torique $\mathcal{T}^{r,\tau} = [\overline{T}^{\mathrm{m},\tau}/T^{r,\tau}]$ et il est muni de deux morphismes projectifs sur le schéma des homomorphismes complets $\widetilde{\mathrm{gl}}_r$ qui s'insèrent dans les deux carrés commutatifs*

$$\begin{array}{ccc} \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r^\tau \\ \mathrm{Id} \downarrow & + & \downarrow \\ \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r \end{array} \quad \text{et} \quad \begin{array}{ccc} \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r^\tau \\ L \downarrow & + & \downarrow \\ \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r \end{array}$$

où $L : g \rightarrow \tau(g)^{-1}g$ est l'isogénie de Lang.

Remarque : En passant au quotient par l'action libre du tore $\mathbb{G}_m^{r,\tau}$ tout entier, on obtient un schéma projectif $\overline{\Omega}^{r,\tau}$, muni d'un morphisme vers la compactification de De Concini et Procesi de PGL_r qui prolonge l'isogénie de Lang $g \mapsto \tau(g)^{-1}g$ de ce schéma en groupes. \square

5. ALLURE DES NOMBRES DE LEFSCHETZ

5.1. Correspondances de Hecke et endomorphismes de Frobenius partiels

On fixe dans toute cette section un niveau $N = \mathrm{Spec}(\mathcal{O}_N) \hookrightarrow X$.

Tout élément $g \in \mathrm{GL}_r(\mathbb{A})$ définit une correspondance dite de Hecke

$$c = (c_1, c_2) : \mathrm{Cht}_N^r(g) \rightarrow \mathrm{Cht}_N^r \times_{(X-N)^2} \mathrm{Cht}_N^r$$

où $\mathrm{Cht}_N^r(g)$ est un champ de Deligne-Mumford et c_1, c_2 sont des morphismes représentables étales. Cette correspondance ne dépend que de la double classe $K_N g K_N$ où K_N est le sous-groupe d'indice fini $\mathrm{Ker}(\mathrm{GL}_r(\mathcal{O}) \rightarrow \mathrm{GL}_r(\mathcal{O}_N))$ de $K = \mathrm{GL}_r(\mathcal{O})$. Si S_g est la réunion du support $|N| \subset |X|$ de N et de l'ensemble fini des places x de X telles que $g_x \notin F_x^\times K_{N,x} \subset \mathrm{GL}_r(F_x)$, c_1 et c_2 sont finies au-dessus de l'ouvert $(X - S_g)^2$ de $(X - N)^2$.

Si $a \in \mathbb{A}^\times$, la double classe $F^\times a \mathrm{Ker}(\mathcal{O}^\times \rightarrow \mathcal{O}_N^\times)$ dans $F^\times \backslash \mathbb{A}^\times / \mathrm{Ker}(\mathcal{O}^\times \rightarrow \mathcal{O}_N^\times)$ est la classe d'isomorphie d'un \mathcal{O}_X -Module inversible \mathcal{L} muni d'une trivialisatoin de sa restriction à N . Si $g \in aK_N \subset \mathrm{GL}_r(\mathbb{A})$, on a $\mathrm{Cht}_N^r(g) = \mathrm{Cht}_N^r$, c_1 est l'identité et c_2 est l'automorphisme de Cht_N^r qui envoie un chtouca $\widetilde{\mathcal{E}}$ muni d'une structure de niveau N sur le produit tensoriel $\mathcal{L} \otimes \widetilde{\mathcal{E}}$ muni de la structure de niveau N évidente. On notera simplement a cet automorphisme.

Soit Λ_X le schéma intersection de tous les ouverts de $X \times X$ complémentaires des images réciproques de la diagonales par les endomorphismes $\mathrm{Frob}_X^n \times \mathrm{Id}_X$ et $\mathrm{Id}_X \times \mathrm{Frob}_X^n$ de $X \times X$ pour tous les entiers $n \geq 0$. Au-dessus de Λ_X , il n'y a plus de différence entre chtoucas à gauche et chtoucas à droite : un chtouca à

droite $\mathcal{E} \xrightarrow{j} \mathcal{E}' \xleftarrow{t} \tau \mathcal{E}$ dont le couple pôle-zéro (∞, o) se factorise par $\Lambda_X \subset X \times X$ définit un chtouca à gauche $\mathcal{E} \xleftarrow{t'} \mathcal{E}'' \xrightarrow{j'} \tau \mathcal{E}$ de pôle $\infty' = \infty$ et de zéro $o' = o$, où $\mathcal{E}'' = \mathcal{E} \times_{j, \mathcal{E}', t} \tau \mathcal{E}$ et t', j' sont les deux projections, et vice versa. Au-dessus de Λ_X , les morphismes de Frobenius partiels introduits en 2.1 peuvent donc être vus comme des endomorphismes, dits encore de Frobenius partiels,

$$\text{Frob}_\infty \text{ et } \text{Frob}_o : \Lambda_X \times_{X^2} \text{Cht}_N^r \rightarrow \Lambda_X \times_{X^2} \text{Cht}_N^r$$

au dessus de $\text{Frob}_X \times \text{Id}_X$ et $\text{Id}_X \times \text{Frob}_X$ respectivement.

On fixe dans la suite un élément $a \in \mathbb{A}^\times$ de degré $\deg(a) > 0$ et on considère le quotient

$$\text{Cht}_N^r / a^\mathbb{Z} \cong \prod_{d=1}^{r \deg(a)} \text{Cht}_N^{r,d}.$$

Les correspondances de Hecke et les endomorphismes de Frobenius partiels que nous venons d'introduire passent au quotient par $a^\mathbb{Z}$.

Pour tout paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ et tout $\alpha \in [0, 1]$, $a^\mathbb{Z}$ stabilise l'ouvert $\text{Cht}_N^{r; \leq \alpha p}$ de Cht_N^r et l'ouvert quotient

$$\text{Cht}_N^{r; \leq \alpha p} / a^\mathbb{Z} \subset \text{Cht}_N^r / a^\mathbb{Z}$$

est de type fini. *Par contre, cet ouvert n'est stabilisé ni par les endomorphismes de Frobenius partiels $\text{Frob}_\infty, \text{Frob}_o$, ni par les correspondances de Hecke (sauf celles qui sont associées à des $g \in \mathbb{A}^\times K$).* C'est la difficulté majeure qu'a dû surmonter Lafforgue.

5.2. Nombres de Lefschetz

Si ∞ et o sont deux points fermés de X le sous-schéma fini $\infty \times o \subset X \times X$ a exactement $\delta(\infty, o)$ points fermés, où $\delta(\infty, o)$ est le plus grand commun diviseur de $\deg(\infty)$ et $\deg(o)$. Le corps résiduel $\kappa(\xi)$ de chaque $\xi \in \infty \times o$ est une extension composée de $\kappa(\infty)$ et $\kappa(o)$, et son degré $\deg(\xi)$ sur \mathbb{F}_q est le plus petit commun multiple

$$\mu(\infty, o) = \frac{\deg(\infty) \deg(o)}{\delta(\infty, o)}$$

de $\deg(\infty)$ et $\deg(o)$. Pour chaque $\xi \in \infty \times o$, les suites de points fermés de $X \times X$,

$$n \mapsto (\text{Frob}_X^n \times \text{Id}_X)(\xi) \text{ et } n \mapsto (\text{Id}_X \times \text{Frob}_X^n)(\xi)$$

sont en fait à valeurs dans $\infty \times o$ et sont périodiques de périodes $\deg(\infty)$ et $\deg(o)$ respectivement.

Fixons $g \in \text{GL}_r(\mathbb{A})$, deux points fermés distincts ∞ et o de $X - S_g$, un entier t , deux entiers $n_\infty, n_o \geq 1$ tels que $\deg(o)n_o - \deg(\infty)n_\infty = t$, un point fermé ξ de $\infty \times o$, un paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ et un nombre réel $\alpha \in [0, 1]$.

On note $\text{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$ la fibre en ξ de la projection canonique $\text{Cht}_N^r/a^{\mathbb{Z}} \rightarrow (X-N)^2$. C'est un champ algébrique de Deligne-Mumford lisse purement de dimension relative $2r - 2$ sur le corps fini $\kappa(\xi)$. On note

$$c_\xi = (c_{1,\xi}, c_{2,\xi}) : \text{Cht}_{N,\xi}^r(g)/a^{\mathbb{Z}} \rightarrow \text{Cht}_{N,\xi}^r/a^{\mathbb{Z}} \times_{\kappa(\xi)} \text{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$$

la fibre en ξ de la correspondance de Hecke définie par g . Les deux composantes $c_{1,\xi}$ et $c_{2,\xi}$ sont représentables finies étales.

Puisque ∞ et o sont distincts, le point fermé ξ est dans $\Lambda_X \subset X^2$. Les puissances $\text{Frob}_\infty^{\text{deg}(\infty)}$ et $\text{Frob}_o^{\text{deg}(o)}$ des endomorphismes de Frobenius partiels sont donc définies sur la fibre $\text{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$ et elles la stabilisent.

DÉFINITION. — *Le nombre de Lefschetz*

$$\text{Lef}_\xi(g \times \text{Frob}_\infty^{\text{deg}(\infty)n_\infty} \times \text{Frob}_o^{\text{deg}(o)n_o}, \text{Cht}_N^{r; \leq \alpha p} / a^{\mathbb{Z}})$$

est la somme

$$\sum_y \frac{1}{|\text{Aut}(y)|}$$

où y parcourt l'ensemble des classes d'isomorphie de points dans $\text{Cht}_{N,\xi}^r(g)/a^{\mathbb{Z}}$ munis d'un isomorphisme

$$\begin{aligned} c_{1,\xi}(y) &\cong (\text{Frob}_\infty^{\text{deg}(\infty)n_\infty} \circ \text{Frob}_o^{\text{deg}(o)n_o})(c_{2,\xi}(y)) \\ &\cong (\text{Frob}_o^{\text{deg}(o)n_o} \circ \text{Frob}_\infty^{\text{deg}(\infty)n_\infty})(c_{2,\xi}(y)) \end{aligned}$$

et tels que

$$c_{1,\xi}(y) \in \text{Cht}_{N,\xi}^{r; \leq \alpha p} / a^{\mathbb{Z}} \subset \text{Cht}_{N,\xi}^r / a^{\mathbb{Z}},$$

et où $\text{Aut}(y)$ est le groupe fini des automorphismes du point fixe y .

En utilisant

- la description adélique de Drinfeld des points de $\text{Cht}_{N,\xi}^r$ ([Dr 5], [La 1]),
- le cas particulier du « Lemme Fondamental » prouvé par Drinfeld ([Lau 1]),
- la décomposition spectrale de Langlands ([Lan 2], [Mo-Wa]) et la formule des traces d'Arthur-Selberg ([Ar], [La 1]),
- la formule de Cauchy pour calculer les intégrales qui apparaissent dans la formule des traces,

Lafforgue démontre alors :

THÉORÈME. — *Si $\text{deg}(\infty)$ et $\text{deg}(o)$ sont assez grands en fonction de la double classe $K_N g K_N$ et de l'entier t et si p est assez convexe en fonction du niveau N et de la double classe $K_N g K_N$, la moyenne de chacune des deux suites périodiques*

$$k \mapsto \text{Lef}_{(\text{Frob}_X^k \times \text{Id}_X)(\xi)}(g \times \text{Frob}_\infty^{\text{deg}(\infty)n_\infty} \times \text{Frob}_o^{\text{deg}(o)n_o}, \text{Cht}_N^{r; \leq \alpha p} / a^{\mathbb{Z}})$$

et

$$k \mapsto \text{Lef}_{(\text{Id}_X \times \text{Frob}_X^k)(\xi)}(g \times \text{Frob}_\infty^{\deg(\infty)n_\infty} \times \text{Frob}_o^{\deg(o)n_o}, \text{Cht}_N^{r; \leq \alpha p} / a^{\mathbb{Z}}),$$

de période le p.g.c.d. de $\deg(o)$, $\deg(\infty)$ et $r!$, est égale à l'expression spectrale suivante

$$\begin{aligned} & \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \text{Tr}_\pi(1_{K_N g K_N}) q^{(\deg(\infty)n_\infty + \deg(o)n_o) \frac{r-1}{2}} S_\infty^{(-n_\infty)}(\pi) S_o^{(n_o)}(\pi) \\ & + \sum_{\substack{1 \leq r' < r \\ 1 \leq r'' < r}} \sum_{\substack{\pi' \in \mathcal{A}_{r'}(K_N) \\ \pi'' \in \mathcal{A}_{r''}(K_N)}} \text{Tr}_{\pi', \pi''}^{\leq \alpha p}(1_{K_N g K_N}, \deg(\infty)n_\infty) S_\infty^{(-n_\infty)}(\pi') S_o^{(n_o)}(\pi'') \end{aligned}$$

où

- $\mathcal{A}_r(K_N) \subset \mathcal{A}_r$ est un système de représentants des classes d'isomorphie de représentations automorphes cuspidales irréductibles de $\text{GL}_r(\mathbb{A})$ qui admettent au moins un vecteur non nul fixe sous K_N ,
- $1_{K_N g K_N}$ est la fonction caractéristique de $K_N g K_N \subset \text{GL}_r(\mathbb{A})$ et $\text{Tr}_\pi(1_{K_N g K_N})$ est la trace de l'opérateur $\pi(1_{K_N g K_N})$,
- pour $x = \infty, o$ et ν un entier positif, $S_x^{(\nu)}(\pi) = z_1(\pi_x)^\nu + \dots + z_r(\pi_x)^\nu$ est la somme des ν -ièmes puissances des valeurs propres de Hecke de la composante locale non ramifiée π_x de π ,
- $\nu \mapsto \text{Tr}_{\pi', \pi''}^{\leq \alpha p}(1_{K_N g K_N}, \nu)$ est une fonction complexe de l'entier ν , qui ne dépend pas des places $o, \infty \in |X| - S_g$, qui est de la forme

$$\sum_{z \in \mathbb{C}^\times} P_z(\nu) z^\nu$$

pour une famille à support fini de polynômes $P_z(u) \in \mathbb{C}[u]$, $z \in \mathbb{C}^\times$, et qui est identiquement nulle sauf pour un nombre fini de couples (π', π'') . \square

6. REPRÉSENTATIONS GALOISIENNES

6.1. Représentations de Galois r -négligeables

Rappelons que F est le corps des fonctions rationnelles sur la courbe X et que l'on a fixé une clôture séparable \overline{F} de F . On désignera par $\overline{\mathbb{F}}_q$ la clôture algébrique de \mathbb{F}_q dans \overline{F} . On identifiera le groupe de Galois $\Gamma_{\mathbb{F}_q} := \text{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ à $\widehat{\mathbb{Z}}$ en choisissant pour générateur topologique de $\Gamma_{\mathbb{F}_q}$ l'élément de Frobenius géométrique (l'inverse de l'élévation à la puissance q -ième dans $\overline{\mathbb{F}}_q$).

Notons E le corps des fractions de $F \otimes F$, c'est-à-dire le corps des fonctions rationnelles sur la surface $X \times X$, et fixons une clôture algébrique \overline{E} de E et un

plongement $\overline{F} \otimes_{\mathbb{F}_q} \overline{F} \hookrightarrow \overline{E}$ au dessus du plongement $F \otimes F \hookrightarrow E$. On a donc défini un point géométrique $\overline{\delta} : \text{Spec}(\overline{E}) \rightarrow X \times X$ localisé en le point générique $\delta = \text{Spec}(E)$ de $X \times X$ et tel que ses images par les deux projections canoniques de $X \times X$ se factorisent par le point géométrique $\overline{\eta} : \text{Spec}(\overline{F}) \rightarrow X$ localisé au point générique η de X .

Pour tout ouvert non vide U de X , le groupe fondamental de Grothendieck $\pi_1(U, \overline{\eta})$ est le quotient de $\Gamma_F = \text{Gal}(\overline{F}/F)$ qui classe les extensions finies de F dans \overline{F} qui sont non ramifiées en toutes les places $x \in |U|$. Il admet comme quotient le groupe de Galois $\Gamma_{\mathbb{F}_q} = \widetilde{\mathbb{Z}}$. De même, pour tout ouvert non vide V de $X \times X$, le groupe fondamental de Grothendieck $\pi_1(V, \overline{\delta})$ est un quotient du groupe de Galois $\text{Gal}(\overline{E}/E)$ et admet $\widetilde{\mathbb{Z}}$ comme quotient.

Soit V un ouvert non vide de $X \times X$ de la forme $U_1 \times U_2$ pour des ouverts U_1 et U_2 de X . Cet ouvert est stable par les endomorphismes $\text{Frob}_X \times \text{Id}_X$ et $\text{Id}_X \times \text{Frob}_X$ de $X \times X$ et les foncteurs «images réciproques» par ces endomorphismes sont des équivalences, quasi-inverses l'une de l'autre, de la catégorie des revêtements finis étales de V sur elle-même (le foncteur image réciproque par $\text{Frob}_V = \text{Frob}_{U_1} \times \text{Frob}_{U_2}$ est canoniquement isomorphe à l'identité). On peut donc considérer la catégorie des revêtements finis étales V' de V munis d'un isomorphisme $(\text{Frob}_{U_1} \times \text{Id}_{U_2})^* V' \xrightarrow{\sim} V'$ ou, ce qui revient au même, d'un isomorphisme $(\text{Id}_{U_1} \times \text{Frob}_{U_2})^* V' \xrightarrow{\sim} V'$. Le foncteur fibre en $\overline{\delta}$ identifie cette catégorie à la catégorie des ensembles finis munis de l'action d'un groupe pro-fini $\widetilde{\pi}_1(U_1 \times U_2, \overline{\delta})$ qui s'insère dans un diagramme commutatif

$$\begin{array}{ccccccc}
 1 & \longrightarrow & \pi_1(U_1 \times U_2, \overline{\delta}) & \longrightarrow & \widetilde{\pi}_1(U_1 \times U_2, \overline{\delta}) & \longrightarrow & \widehat{\mathbb{Z}} \longrightarrow 0 \\
 & & \downarrow & & \downarrow & & \parallel \\
 0 & \longrightarrow & \widehat{\mathbb{Z}} & \longrightarrow & \widehat{\mathbb{Z}}^2 & \longrightarrow & \widehat{\mathbb{Z}} \rightarrow 0
 \end{array}$$

à lignes exactes et à flèches verticales surjectives. Tout revêtement fini étale V' de $V = U_1 \times U_2$ de la forme $V' = U'_1 \times U'_2$ où chaque U'_i est un revêtement fini étale de U_i , est canoniquement muni d'un isomorphisme $(\text{Frob}_{U_1} \times \text{Id}_{U_2})^* V' \xrightarrow{\sim} V'$. On a donc un homomorphisme de $\widetilde{\pi}_1(U_1 \times U_2, \overline{\delta})$ dans $\pi_1(U_1, \overline{\eta}) \times \pi_1(U_2, \overline{\eta})$ qui prolonge l'homomorphisme $\pi_1(U_1 \times U_2, \overline{\delta}) \rightarrow \pi_1(U_1, \overline{\eta}) \times \pi_1(U_2, \overline{\eta})$ défini par les deux projections canoniques de $U_1 \times U_2$, et qui induit l'identité entre les quotients $\widehat{\mathbb{Z}}^2$ de sa source et de son but.

PROPOSITION (Drinfeld). — *L'homomorphisme défini ci-dessus*

$$\widetilde{\pi}_1(U_1 \times U_2, \overline{\delta}) \rightarrow \pi_1(U_1, \overline{\eta}) \times \pi_1(U_2, \overline{\eta}),$$

est un isomorphisme.

On a fixé dans la section 1 un nombre premier ℓ distinct de la caractéristique de \mathbb{F}_q et une clôture algébrique $\overline{\mathbb{Q}}_\ell$ de \mathbb{Q}_ℓ .

Rappelons que, pour tout ouvert non vide U de X (resp. V de $X \times X$), le foncteur fibre en $\overline{\eta}$ (resp. $\overline{\delta}$) est une équivalence de la catégorie des systèmes locaux ℓ -adiques (c'est-à-dire des $\overline{\mathbb{Q}}_\ell$ -faisceaux lisses) sur U (resp. V) sur la catégorie des

représentations ℓ -adiques de $\pi_1(U, \bar{\eta})$ (resp. $\pi_1(V, \bar{\delta})$). Pour tout système local ℓ -adique L sur U (resp. M sur V) et tout point fermé x de U (resp. y de V), la classe de conjugaison dans $\pi_1(U, \bar{\eta})$ (resp. $\pi_1(V, \bar{\delta})$) des éléments de Frobenius géométriques en x (resp. y) définit donc une classe de conjugaison d'endomorphismes de $L_{\bar{\eta}}$ (resp. $M_{\bar{\delta}}$). On note $\text{Tr}(\text{Frob}_x, L)$ (resp. $\text{Tr}(\text{Frob}_y, M)$) sa trace et $\det(1 - T \text{Frob}_x, L)$ (resp. $\det(1 - T \text{Frob}_y, M)$) son polynôme caractéristique.

Soient U un ouvert non vide de X et M un système local ℓ -adique *irréductible* sur $U \times U$. On considère les propriétés suivantes :

- (i) M est de la forme $L_1 \boxtimes L_2$ pour des systèmes locaux ℓ -adiques (irréductibles) L_1 sur L_2 sur U et, en particulier, il existe des fonctions $\Delta_1, \Delta_2 : |U \times U| \rightarrow 1 + T\overline{\mathbb{Q}}_\ell[T]$ telles que

$$\det(1 - T \text{Frob}_\xi, M) = \Delta_1(\infty) * \Delta_2(o)$$

pour tout $\xi \in |U \times U|$ de projections $\infty, o \in |U|$, où l'opération $*$ est définie par

$$\prod_{i_1=1}^{r_1} (1 - \alpha_{1,i_1} T) * \prod_{i_2=1}^{r_2} (1 - \alpha_{2,i_2} T) = \prod_{i_1, i_2} (1 - \alpha_{1,i_1} \alpha_{2,i_2} T).$$

- (ii) Il existe un isomorphisme $(\text{Frob}_U \times \text{Id}_U)^* M \xrightarrow{\sim} M$ et, en particulier, la fonction $\xi \mapsto \det(1 - T \text{Frob}_\xi, M)$ est constante sur $|\infty \times o|$ pour chaque couple (∞, o) de points fermés de U .

La propriété (i) implique évidemment la propriété (ii) et il devrait résulter de la proposition de Drinfeld que les propriétés (i) et (ii) sont en fait équivalentes.

Lafforgue démontre un résultat partiel dans cette direction.

PROPOSITION. — *On suppose de plus que M est ponctuellement pur au sens de [De] et qu'il existe des fonctions $\Delta_1, \Delta_2 : |U \times U| \rightarrow 1 + T\overline{\mathbb{Q}}_\ell[T]$ telles que*

$$\det(1 - T \text{Frob}_\xi, M) = \Delta_1(\infty) * \Delta_2(o)$$

pour tout $\xi \in |U \times U|$ dont les projections $\infty, o \in |U|$ sont distinctes.

Alors, la propriété (ii) implique la propriété (i).

DÉFINITION. — *Si U est un ouvert de X , un système local ℓ -adique sur $U \times U$ est dit r -négligeable si tous ses sous-quotients irréductibles sont des facteurs directs de systèmes locaux de la forme $L_1 \boxtimes L_2$ où L_1 et L_2 sont des systèmes locaux ℓ -adiques irréductibles de rangs $\leq r - 1$ sur U .*

Si λ est une unité ℓ -adique, on notera par $\overline{\mathbb{Q}}_\ell^{(\lambda)}$ le $\overline{\mathbb{Q}}_\ell$ -faisceau sur $\text{Spec}(\mathbb{F}_q)$ défini par le caractère $\widehat{\mathbb{Z}} \rightarrow \overline{\mathbb{Q}}_\ell^\times, 1 \mapsto \lambda$, du groupe de Galois $\Gamma_{\mathbb{F}_q}$. Si F est un faisceau ℓ -adique sur un (\mathbb{F}_q) -schéma S , on notera suivant Deligne $F^{(\lambda)}$ le produit tensoriel de F par l'image réciproque sur S de $\overline{\mathbb{Q}}_\ell^{(\lambda)}$. En particulier, $F^{(q^{-1})}$ n'est autre que la torsion à la Tate $F(1)$ de F . Rappelons que tout système local ℓ -adique L sur un schéma S normal de type fini peut s'écrire sous la forme $L = M^{(\lambda)}$ où λ est une unité ℓ -adique et M est un système local ℓ -adique dont le déterminant (c'est-à-dire la puissance extérieure maximale) est *d'ordre fini*.

Remarque : Comme $\pi_1(X \times X, \bar{\delta})$ diffère du produit $\pi_1(X, \bar{\eta}) \times \pi_1(X, \bar{\eta})$ par une copie de $\widehat{\mathbb{Z}}$, il arrive que, pour deux systèmes locaux ℓ -adiques irréductibles L_1 et L_2 sur X , le système local $L_1 \boxtimes L_2$ sur $X \times X$ ne soit pas irréductible. C'est le cas si on a $L_1 \cong L_1^{(\lambda)}$ et $L_2 \cong L_2^{(\lambda^{-1})}$ pour une racine de l'unité $\lambda \neq 1$ d'ordre divisant à la fois le rang de L_1 et celui de L_2 . Cependant un tel système local $L_1 \boxtimes L_2$ est toujours semi-simple et ses facteurs irréductibles sont images les uns des autres par les puissances de $\text{Frob}_X \times \text{Id}_X$.

□

6.2. La récurrence

Soient $N \subset X$ un niveau, $p : [0, r] \rightarrow \mathbb{R}$ un paramètre de troncature assez convexe par rapport au genre de la courbe X et à N , et d un entier. On a vu dans la section 3 comment Lafforgue construit un morphisme de champs algébriques

$$\overline{\text{Cht}}_N^{r,d;\leq p} \rightarrow \overline{\text{Tr}}_N^{r,\tau} \times (X - N)^2$$

lisse purement de dimension relative $2r - 2$, dont le composé avec la projection sur $(X - N)^2$ est propre, et dont la restriction au-dessus de l'ouvert $\text{Spec}(\mathbb{F}_q) = \text{Tr}_N^{r,\tau} \subset \overline{\text{Tr}}_N^{r,\tau}$ est le morphisme pôle-zéro $\text{Cht}_N^{r,d;\leq p} \rightarrow (X - N)^2$.

Par construction (cf. 3.4), le champ $\widetilde{\text{Tr}}_N^{r,\tau}$ est lisse sur $\mathcal{T}^{r,\tau} \times_{[\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}], \langle q-1 \rangle} [\mathbb{A}^{r-1}/\mathbb{G}_m^{r-1}]$ où $\mathcal{T}^{r,\tau}$ est un champ torique. On sait donc construire des résolutions des singularités $\widetilde{\text{Tr}}_N^{r,\tau} \rightarrow \overline{\text{Tr}}_N^{r,\tau}$. Fixons-en une et notons

$$\widetilde{\text{Cht}}_N^{r,d;\leq p} = \overline{\text{Cht}}_N^{r,d;\leq p} \times_{\overline{\text{Tr}}_N^{r,\tau}} \widetilde{\text{Tr}}_N^{r,\tau}.$$

Le morphisme de champs

$$\widetilde{\text{Cht}}_N^{r,d;\leq p} \rightarrow (X - N)^2$$

prolonge le morphisme pôle-zéro $\text{Cht}_N^{r,d;\leq p} \rightarrow (X - N)^2$ et est encore propre. Il est de plus lisse purement de dimension relative $2r - 2$, et le fermé complémentaire de l'ouvert $\text{Cht}_N^{r,d;\leq p} \subset \widetilde{\text{Cht}}_N^{r,d;\leq p}$ est un diviseur à croisements normaux relatif sur $(X - N)^2$ qui est réunion de diviseurs lisses sur $(X - N)^2$.

En prenant la réunion disjointe sur les entiers d des champs $\widetilde{\text{Cht}}_N^{r,d;\leq p}$ et en divisant par l'action de $a^{\mathbb{Z}}$, on obtient un morphisme de champs

$$\widetilde{\text{Cht}}_N^{r;\leq p} / a^{\mathbb{Z}} \rightarrow (X - N)^2$$

qui prolonge le morphisme pôle-zéro $\text{Cht}_N^{r;\leq p} / a^{\mathbb{Z}} \rightarrow (X - N)^2$, qui est propre et lisse purement de dimension relative $2r - 2$, et pour lequel le fermé complémentaire de l'ouvert $\text{Cht}_N^{r;\leq p} / a^{\mathbb{Z}} \subset \widetilde{\text{Cht}}_N^{r;\leq p} / a^{\mathbb{Z}}$ est un diviseur à croisements normaux relatif sur $(X - N)^2$ qui est réunion de diviseurs lisses sur $(X - N)^2$.

Le théorème suivant est une description partielle de la cohomologie ℓ -adique des variétés de chtoucas compactifiées.

THÉORÈME. — *Pour tout niveau N et tout paramètre de troncature p assez convexe par rapport au genre de la courbe X et à N , les images directes supérieures du faisceau constant de valeur $\overline{\mathbb{Q}}_\ell$ par le morphisme canonique $\widetilde{\text{Cht}}_N^{r; \leq p} / a^{\mathbb{Z}} \rightarrow (X - N)^2$ sont toutes des systèmes locaux ℓ -adiques $(r + 1)$ -négligeables.*

Lafforgue prouve le théorème principal de la section 1 et le théorème ci-dessus par une récurrence simultanée sur r . Plus précisément, pour tout entier $r' \geq 1$, il considère les propriétés suivantes :

$P_1(r')$ Quel que soit $\pi' \in \mathcal{A}_{r'}$, il existe $\sigma(\pi') \in \mathcal{G}_{r'}$ tel que l'on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma(\pi')_x, T)) = L(\pi'_x, T)$$

pour presque toute place $x \notin S_{\sigma(\pi')} \cup S_{\pi'}$.

$P_2(r')$ Quel que soit $\sigma' \in \mathcal{G}_{r'}$, il existe $\pi(\sigma') \in \mathcal{A}_{r'}$ tel que l'on ait l'égalité de facteurs L locaux

$$L(\pi(\sigma')_x, T) = \iota(L(\sigma'_x, T))$$

pour presque toute place $x \notin S_{\pi(\sigma')} \cup S_{\sigma'}$.

$P_3(r')$ Quels que soient $\pi' \in \mathcal{A}_{r'}$ et $x \in |X| - S_{\pi'}$, les valeurs propres de Hecke $z_1(\pi'), \dots, z_{r'}(\pi')$ sont toutes de valeur absolue 1.

$P_4(r')$ Le théorème ci-dessus est démontré en rang r' .

Et il prouve :

THÉORÈME. — *Fixons un entier $r \geq 1$. Alors les propriétés $P_1(r')$, $P_3(r')$ et $P_4(r')$ pour $r' = 1, \dots, r - 1$ impliquent ces mêmes propriétés pour $r' = r$.*

Comme on l'a déjà signalé en 1.4, les propriétés $P_1(1), \dots, P_1(r-1)$ impliquent les propriétés $P_2(1), \dots, P_2(r-1), P_2(r)$, et en particulier la correspondance de Langlands en tout rang $r' < r$. Le théorème ci-dessus implique donc bien le théorème principal de 1.3. Ces mêmes propriétés impliquent aussi des propriétés renforcées où l'on impose que $S_{\sigma(\pi')} = S_{\pi'}$ et les égalités de facteurs L locaux pour toutes les places $x \notin S_{\pi'}$.

Dans la suite de cette section nous allons commencer la démonstration du théorème précédent en nous concentrant sur les aspects galoisiens.

6.3. Les strates du bord sont r -négligeables

Dans toute cette sous-section on fixe un paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ assez convexe en fonction du genre de la courbe X et de N .

Pour tout entier ν on note

$$H_N^{r; \leq p, \nu}, \widetilde{H}_N^{r; \leq p, \nu} \text{ et } \widetilde{H}_{N, \partial}^{r; \leq p, \nu}$$

les ν -ièmes images directes supérieures à supports compacts du faisceau constant $\overline{\mathbb{Q}}_\ell$ par les morphismes de champs

$$\mathrm{Cht}_N^{r; \leq p} / a^{\mathbb{Z}} \rightarrow (X - N)^2, \quad \widetilde{\mathrm{Cht}}_N^{r; \leq p} / a^{\mathbb{Z}} \rightarrow (X - N)^2$$

et

$$\left(\widetilde{\mathrm{Cht}}_N^{r; \leq p} - \mathrm{Cht}_N^{r; \leq p} \right) / a^{\mathbb{Z}} \rightarrow (X - N)^2.$$

Ce sont des systèmes locaux ℓ -adiques sur $(X - N)^2$, qui sont nuls pour $\nu \notin \{0, 1, \dots, 4r - 4\}$ (et même pour $\nu \notin \{0, 1, \dots, 4r - 6\}$ dans le cas de $\widetilde{H}_{N, \partial}^{r; \leq p, \nu}$), et on a une suite exacte longue

$$\dots \rightarrow H_N^{r; \leq p, \nu} \rightarrow \widetilde{H}_N^{r; \leq p, \nu} \rightarrow \widetilde{H}_{N, \partial}^{r; \leq p, \nu} \rightarrow H_N^{r; \leq p, \nu+1} \rightarrow \dots$$

Le diviseur à croisements normaux relatif $\left(\widetilde{\mathrm{Cht}}_N^{r; \leq p} - \mathrm{Cht}_N^{r; \leq p} \right) / a^{\mathbb{Z}}$ sur $(X - N)^2$ est réunion d'une famille finie de diviseurs lisses. Il est donc muni d'une stratification canonique. Une *strate fermée du bord* est une intersection d'une sous-famille non vide de cette famille. La relation d'inclusion définit une relation d'ordre sur les strates fermées du bord. Une *strate ouverte du bord* est le complémentaire dans une strate fermée du bord de la réunion des strates fermées du bord strictement plus petite. Par exemple, pour $N = \emptyset$, les strates ouvertes du bord sont les $\overline{\mathrm{Cht}}_R^{r; \leq p} / a^{\mathbb{Z}}$ pour $\emptyset \neq R = \{r_1, \dots, r_{s-1}\} \subset [r - 1]$ définies en 3.3, et les strates fermées du bord sont les adhérences des strates ouvertes du bord.

Toute strate du bord (fermée ou ouverte) est lisse sur $(X - N)^2$, et les images directes supérieures à supports compacts de $\overline{\mathbb{Q}}_\ell$ par la restriction à une telle strate du morphisme pôle-zéro sont toutes des systèmes locaux ℓ -adiques sur $(X - N)^2$, qu'il est commode d'appeler *images directes supérieures de la strate*.

PROPOSITION. — *Supposons que les propriétés $P_1(1) \dots, P_1(r-1)$ soient vérifiées pour ce niveau N . Alors, les systèmes locaux ℓ -adiques*

- *images directes supérieures des strates fermées du bord,*
- *images directes supérieures des strates ouvertes du bord*
- $\widetilde{H}_{N, \partial}^{r; \leq p, \nu}$, $0 \leq \nu \leq 4r - 6$,

sont tous r -négligeables.

Esquissons la preuve de cette proposition pour $N = \emptyset$. On procède par récurrence sur r . Compte tenu de la suite exacte longue ci-dessus, les hypothèses de la proposition et de récurrence assurent que les images directes supérieures par les morphismes

$$\mathrm{Cht}^{r', d'; \leq p'} \rightarrow X \times X$$

sont toutes $(r' + 1)$ -négligeables quels que soient les entiers $1 \leq r' < r$ et d' et le paramètre de troncature p' assez convexe par rapport au genre de X .

Chaque strate $\overline{\text{Cht}}_R^{r,d;\leq p}$, $\emptyset \neq R = \{r_1, \dots, r_{s-1}\} \subset [r-1]$, est finie et radicielle sur un champ $\text{Cht}^{R,d;\leq p}$ et lui est donc équivalente du point de vue de la cohomologie ℓ -adique. Par hypothèse de récurrence, tous les sous-quotients irréductibles des images directes supérieures par le morphisme

$$(\infty = \infty_1, o_1 = \infty_2, \dots, o_{s-1} = \infty_s, o_s = o) : \text{Cht}^{R,d;\leq p} \rightarrow X \times X^{s-1} \times X$$

sont des facteurs directs de faisceaux de la forme

$$L'_1 \boxtimes \left(\boxtimes_{j=1}^{s-1} (L''_j \otimes L'_{j+1}) \right) \boxtimes L''_s$$

pour des systèmes locaux ℓ -adiques irréductibles L'_1, L''_1 de rangs $\leq r_1 < r$, L'_2, L''_2 de rangs $\leq r_2 - r_1 < r$, ... et L'_s, L''_s de rangs $\leq r - r_{s-1} < r$ sur la courbe X . Par suite, les images directes supérieures par le morphisme $(\infty, o) : \overline{\text{Cht}}_R^{r,d;\leq p} \rightarrow X \times X$ se décomposent en systèmes locaux ℓ -adiques qui sont des facteurs directs de faisceaux de la forme

$$(L'_1 \boxtimes L''_s) \otimes H^\nu \left(\overline{\mathbb{F}}_q \otimes_{\mathbb{F}_q} X^{s-1}, \boxtimes_{j=1}^{s-1} (L''_j \otimes L'_{j+1}) \right),$$

et tous leurs sous-quotients irréductibles sont des facteurs directs de systèmes locaux de la forme $(L'_1 \boxtimes L''_s)^{(\lambda)}$ où λ est une valeur propre de Frob_q agissant sur $H^\nu(\overline{\mathbb{F}}_q \otimes_{\mathbb{F}_q} X^{s-1}, \boxtimes_{j=1}^{s-1} (L''_j \otimes L'_{j+1}))$; la proposition s'en suit.

Le cas général est un peu plus difficile mais similaire en utilisant le fait essentiel que les $\overline{\text{Cht}}_N^{r;\leq p}/a^{\mathbb{Z}}$ et $\widetilde{\text{Cht}}_N^{r;\leq p}/a^{\mathbb{Z}}$ sont construits à partir des $\overline{\text{Cht}}^{r;\leq p}/a^{\mathbb{Z}}$ en formant un produit fibré au dessus du champ $\overline{\text{Tr}}_N^r$. On a besoin d'une formule de comptage supplémentaire où interviennent les fonctions d'Euler-Poincaré introduites par Kottwitz (cf. [Lau 1]) et un argument inspiré de [Ha-Ta].

6.4. Séparation de la partie cuspidale de la cohomologie

Soit S un schéma de type fini sur \mathbb{F}_q . Un système local ℓ -adique *virtuel* M sur $U \times U$ est une combinaison linéaire formelle $M = \sum_L m_M(L)[L]$ où L parcourt un ensemble de représentants des classes d'isomorphie de systèmes locaux ℓ -adiques irréductibles sur S (fixé une fois pour toute) et où $L \mapsto m_M(L)$ est une fonction sur cet ensemble à valeurs *rationnelles* et à support fini. Le coefficient $m_M(L) \in \mathbb{Q}$ est appelé la *multiplicité* de L dans M . On dit que L *intervient dans* M ou encore que L *est un constituant* de M si $m_M(L) \neq 0$. Un système local ℓ -adique M définit un système local ℓ -adique virtuel $[M] : m_M(L)$ est la multiplicité de L dans n'importe quelle filtration de Jordan-Hölder de M .

On dira qu'un système local ℓ -adique virtuel $M = \sum_L m_M(L)[L]$ sur S est effectif si $m_M(L) \geq 0$ quel que soit L . On dira qu'il est *pur de poids* un entier w si tous les L qui interviennent dans M sont ponctuellement purs de poids w . On dira qu'il est *mixte de poids contenus dans* un ensemble W d'entiers si chaque L qui intervient dans M est ponctuellement pur de poids appartenant à W .

Soit maintenant U un ouvert non vide de X . Un système local ℓ -adique *virtuel* M sur $U \times U$ est dit r -négligeable si tous les L qui interviennent dans M sont r -négligeables. Si $M = \sum_L m_M(L)[L]$ est un système local ℓ -adique virtuel sur $U \times U$, on appellera *morceau non r -négligeable* de M et on notera $M_{\text{cusp}} = \sum_L m_{M_{\text{cusp}}}(L)[L]$ le système local ℓ -adique virtuel sur $U \times U$ défini par $m_{M_{\text{cusp}}}(L) = 0$ si L est r -négligeable et $m_{M_{\text{cusp}}}(L) = m_M(L)$ sinon.

En confrontant la formule des points fixes de Grothendieck-Lefschetz ([Gr])

$$\text{Lef}_\xi \left(\text{Frob}_{\text{Cht}_N^{r; \leq p} / a^z}^{\deg(\xi)n} \right) = \sum_\nu (-1)^\nu \text{Tr}(\text{Frob}_\xi^n, H_N^{r; \leq p, \nu})$$

et le cas particulier $g = 1$ et $\deg(\infty)n_\infty = \deg(o)n_o = \deg(\xi)n$ de la formule pour la moyenne de nombres de Lefschetz énoncée en 5.2, on voit que sans changer la valeur de l'expression

$$\sum_{\substack{1 \leq r' < r \\ 1 \leq r'' < r}} \sum_{\substack{\pi' \in \mathcal{A}_{r'}(K_N) \\ \pi'' \in \mathcal{A}_{r''}(K_N)}} \text{Tr}_{\pi', \pi''}^{\leq p}(1_{K_N}, \deg(\xi)n) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi') S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi'')$$

on peut y remplacer les fonctions $\nu \mapsto \text{Tr}_{\pi', \pi''}^{\leq p}(1_{K_N}, \nu)$ par des fonctions de la forme $\sum_{z \in \mathbb{C}^\times} c_z z^\nu$ pour une famille à support fini de scalaires $c_z \in \mathbb{Q}$, $z \in \mathbb{C}^\times$.

L'hypothèse de récurrence permet de remplacer les π' et π'' qui interviennent dans cette même expression par des systèmes locaux ℓ -adiques irréductibles et purs de poids 0 sur $X - N$. Compte tenu du théorème de pureté de Deligne ([De]) on en déduit :

PROPOSITION. — *Si $p : [0, r] \rightarrow \mathbb{R}$ est un paramètre de troncature assez convexe en fonction du genre de la courbe X et de N , il existe un système local ℓ -adique virtuel mixte de poids entiers $\tilde{H}_{N, \text{cusp}}^{r; \leq p}$ sur $(X - N)^2$ tel que la différence*

$$\tilde{H}_{N, \text{cusp}}^{r; \leq p} - \frac{1}{r!} \sum_{k=1}^r \sum_\nu (-1)^\nu [(\text{Frob}_X^k \times \text{Id}_X)^* H_N^{r; \leq p, \nu}]$$

soit r -négligeable et que, pour tout couple (∞, o) de points fermés distincts de $X - N$, tout point fermé $\xi \in \infty \times o$ et tout entier $n \geq 1$, on ait

$$\text{Tr}(\text{Frob}_\xi^n, \tilde{H}_{N, \text{cusp}}^{r; \leq p}) = \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a) = 1}} q^{(r-1)\deg(\xi)n} \text{Tr}_\pi(1_{K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi).$$

L'encadrement de Jacquet et Shalika ([J-S 1])

$$q^{-\frac{1}{2}} < |z_i(\pi_x)|^{\frac{1}{\deg(x)}} < q^{\frac{1}{2}}, \quad \forall i = 1, \dots, r, \quad \forall x \in |X - N|,$$

pour les valeurs propres de Hecke de tout $\pi \in \mathcal{A}_r(K_N)$ assure que $\tilde{H}_{N, \text{cusp}}^{r; \leq p}$ est mixte de poids contenus dans $\{2r - 3, 2r - 2, 2r - 1\}$ et qu'il a des constituants de poids $2r - 3$ si et seulement si il en a de poids $2r - 1$.

Par des arguments de fonctions L pour les système locaux ℓ -adiques sur la surface $(X - N)^2$, Lafforgue déduit de la proposition ci-dessus, du théorème de pureté de Deligne et des résultats de 6.3 :

COROLLAIRE. — (i) *Le système local ℓ -adique virtuel $\widetilde{H}_{N,\text{cusp}}^{r;\leq p}$ sur $(X - N)^2$ est effectif, pur de poids $2r - 2$ et n'admet aucun constituant r -négligeable.*

(ii) *Les systèmes locaux ℓ -adiques $\widetilde{H}_N^{r;\leq p,\nu}$, $\nu \neq 2r - 2$, sont tous r -négligeables et $\widetilde{H}_{N,\text{cusp}}^{r;\leq p}$ est exactement le morceau non r -négligeable du système local ℓ -adique virtuel*

$$\frac{1}{r!} \sum_{k=1}^{r!} [(\text{Frob}_X^k \times \text{Id}_X)^* \widetilde{H}_N^{r;\leq p, 2r-2}].$$

Pour tous paramètres de troncatures $p \leq q$ qui sont assez convexes par rapport au genre de X et au niveau N , on a un carré commutatif

$$\begin{array}{ccc} H_N^{r;\leq q,*} & \longrightarrow & \widetilde{H}_N^{r;\leq q,*} \\ \uparrow & + & \downarrow \\ H_N^{r;\leq p,*} & \longrightarrow & \widetilde{H}_N^{r;\leq p,*} \end{array}$$

où les deux flèches horizontales et la flèche verticale montante de gauche sont les prolongements par zéro pour les inclusions de $\text{Cht}_N^{r;\leq p} \subset \widetilde{\text{Cht}}_N^{r;\leq p}$, $\text{Cht}_N^{r;\leq q} \subset \widetilde{\text{Cht}}_N^{r;\leq q}$ et $\text{Cht}_N^{r;\leq p} \subset \text{Cht}_N^{r;\leq q}$ et la flèche verticale descendante de droite est induite par la correspondance dans $\widetilde{\text{Cht}}_N^{r;\leq p} \times \widetilde{\text{Cht}}_N^{r;\leq q}$ qui est l'adhérence du graphe de l'inclusion $\text{Cht}_N^{r;\leq p} \subset \text{Cht}_N^{r;\leq q}$. La proposition de 6.3, le corollaire ci-dessus et l'existence de ce carré commutatif impliquent facilement le lemme suivant qui est crucial pour la suite.

LEMME. — *Pour tous paramètres de troncatures $p \leq q$ qui sont assez convexes par rapport au genre de X et au niveau N , les noyau et conoyau de l'homomorphisme de prolongement par zéro*

$$H_N^{r;\leq p, 2r-2} \rightarrow H_N^{r;\leq q, 2r-2}$$

sont r -négligeables.

En utilisant ce lemme, Lafforgue prouve que le ind-système local ℓ -adique «de rang infini» sur $(X - N)^2$

$$H_N^{r; 2r-2} = \varinjlim_p H_N^{r;\leq p, 2r-2},$$

où p parcourt l'ensemble des paramètres de troncature qui sont assez convexes par rapport au genre de la courbe X et à N , a la propriété suivante :

LEMME. — *Il existe une unique filtration finie*

$$F^\bullet = ((0) = F^0 \subseteq F^1 \subsetneq F^2 \subsetneq \dots \subsetneq F^{2u+1} \subsetneq F^{2u} \subsetneq \dots \subsetneq F^T = H_N^{r; 2r-2}$$

telle que :

- pour tout entier $u \geq 0$ tel que $2u+1 \leq T$, F^{2u+1}/F^{2u} est la somme de tous les sous-systèmes locaux ℓ -adiques (de rang fini) r -négligeables de $H_N^{r; 2r-2}/F^{2u}$,
- pour tout entier $u \geq 0$ tel que $2u + 2 \leq T$, F^{2u+2}/F^{2u+1} est la somme de tous les sous-systèmes locaux ℓ -adiques (de rang fini) $H_N^{r; 2r-2}/F^{2u+1}$ dont aucun sous-quotient n'est r -négligeable,
- si p est un paramètre de troncature assez convexe par rapport au genre de X et au niveau N et si on note $F^{\leq p, \bullet}$ la filtration sur $H_N^{r; \leq p, 2r-2}$ induite par F^\bullet , alors, pour tout entier $u \geq 0$, le plongement

$$F^{\leq p, 2u+2}/F^{\leq p, 2u+1} \hookrightarrow F^{2u+2}/F^{2u+1}$$

est un isomorphisme.

Considérons alors le système local ℓ -adique (de rang fini) sur $(X - N)^2$

$$H_{N, \text{cusp}}^r = \bigoplus_{u \geq 0} F^{2u+2}/F^{2u+1}.$$

Rassemblant les résultats de cette sous-section, on obtient finalement :

PROPOSITION. — (i) Pour tout paramètre de troncature p assez convexe par rapport au genre de X et au niveau N , on a l'égalité de systèmes locaux ℓ -adiques virtuels

$$\tilde{H}_{N, \text{cusp}}^{r; \leq p} = \frac{1}{r!} \sum_{k=1}^{r!} [(\text{Frob}_{X-N}^k \times \text{Id}_{X-N})^* H_{N, \text{cusp}}^r].$$

(ii) Le système local ℓ -adique $H_{N, \text{cusp}}^r$ est ponctuellement pur de poids $2r - 2$.

(iii) Pour tout couple (∞, o) de points fermés distincts de $X - N$, tout point fermé ξ de $\infty \times o$ et tout entier n , on a

$$\begin{aligned} & \frac{1}{r!} \sum_{k=1}^{r!} \text{Tr}(\text{Frob}_\xi^n, (\text{Frob}_{X-N}^k \times \text{Id}_{X-N})^* H_{N, \text{cusp}}^r) \\ &= q^{(r-1) \deg(\xi)n} \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \text{Tr}_\pi(1_{K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi). \end{aligned}$$

7. UNE VARIANTE D'UN THÉORÈME DE PINK

7.1. Le cas des schémas

Soient S un schéma de type fini et lisse (sur \mathbb{F}_q) et $p : X \rightarrow S$ un morphisme de schémas propre et lisse purement de dimension relative d . On se donne un diviseur

$Y \subset X$ à croisements normaux relatif sur S qui est réunion $Y = Y_1 \cup \dots \cup Y_m$ de m diviseurs lisses sur S . Pour toute partie I de $\{1, \dots, m\}$ on note $Y_I = \bigcap_{i \in I} Y_i$ et $X_I = Y_I - \bigcup_{J \supsetneq I} Y_J$. On a donc $X_\emptyset = X - Y \subset Y_\emptyset = X$ et X est la réunion disjointe de ses parties localement fermées X_I . De plus, pour chaque $I \subset \{1, \dots, m\}$, la restriction p_I de p à Y_I est propre et lisse purement de dimension relative $d - |I|$, et X_I est un ouvert dense de Y_I .

On considère un endomorphisme $f : S \rightarrow S$ et un morphisme fini de schémas

$$\Gamma_\emptyset \rightarrow Z_\emptyset := X_\emptyset \times_{f,S} X_\emptyset \subset X_\emptyset \times X_\emptyset$$

où « $\times_{f,S}$ » est une notation abrégée pour « $\times_{p_\emptyset \circ f, S, p_\emptyset}$ ». On suppose que la première projection $\text{pr}_{\emptyset,1} : \Gamma_\emptyset \rightarrow X_\emptyset$ est étale. Le S -schéma est donc lisse purement de dimension relative d sur S . En particulier il est normal.

Alors, si s est un point fermé de S et si $n \geq 1$ est un entier tel que

$$\text{Frob}_S^n(f(s)) = f(\text{Frob}_S^n(s)) = s,$$

la fibre $\Gamma_{\emptyset,s} \rightarrow Z_{\emptyset,s} \subset X_{\emptyset,f(s)} \times X_{\emptyset,s}$ en s de Γ_\emptyset coupe transversalement la fibre $X_{\emptyset,f(s)} \rightarrow X_{\emptyset,s} \times X_{\emptyset,f(s)}$ en $(s, f(s))$ du graphe $\delta_\emptyset^n = (\text{Frob}_{X_\emptyset}^n, \text{Id}_{X_\emptyset}) : X_\emptyset \rightarrow X_\emptyset \times X_\emptyset$. On note

$$\text{Lef}_s(\Gamma_\emptyset \times \text{Frob}_{X_\emptyset}^n, X_\emptyset)$$

le nombre des points d'intersection.

On note $\Gamma \rightarrow Z := X \times_{f,S} X \subset X \times X$ le morphisme fini de schémas obtenu par normalisation de Z dans Γ_\emptyset . On se propose de donner une interprétation cohomologique de $\text{Lef}_s(\Gamma_\emptyset \times \text{Frob}_X^n, X_\emptyset)$ sous une hypothèse de « stabilité de l'ouvert $X_\emptyset \subset X$ au voisinage des points fixes de Γ ».

7.2. Stabilité au voisinage des points fixes

On note $\text{pr}_1, \text{pr}_2 : \Gamma \rightarrow X$ les deux projections canoniques de Γ et, pour chaque entier $n \geq 0$, δ^n le graphe $(\text{Frob}_X^n, \text{Id}_X) : X \rightarrow X \times X$.

DÉFINITION. — *On dira que Γ stabilise X_\emptyset au voisinage de ses points fixes s'il existe un ouvert $U \subset X \times X$ contenant toutes les intersections de l'image de Γ dans $Z \subset X \times X$ avec les images des graphes de Frobenius $\delta^n(X) \subset X \times X$, $n \geq 0$, et tel que*

$$\text{pr}_1(\text{pr}_2^{-1}(X_\emptyset) \cap U_\Gamma) \subset X_\emptyset$$

où $U_\Gamma \subset \Gamma$ est la trace de U sur Γ .

LEMME. — *Pour que Γ stabilise X_\emptyset au voisinage de ses points fixes, il faut et il suffit que Γ satisfasse au critère valuatif suivant :*

(*) *Pour tout anneau de valuation discrète A de corps des fractions K et de corps résiduel k et pour tout point $\gamma : \text{Spec}(A) \rightarrow \Gamma$ tel que $\text{pr}_1(\gamma(\text{Spec}(K))) \in Y(K)$ et $\text{pr}_2(\gamma(\text{Spec}(K))) \in X_\emptyset(K)$, l'image de $\gamma(\text{Spec}(k)) \in \Gamma(k)$ dans $Z \subset X \times X$ n'est dans l'image d'aucun des graphes de Frobenius δ^n , $n \in \mathbb{N}$.*

Comme l'a fait Pink dans [Pi], Lafforgue introduit l'éclatement $\pi : \tilde{Z} \rightarrow Z$ de $Z = X \times_{f,S} X$ le long du fermé réunion des $Y_i \times_{f,S} Y_i$, $1 \leq i \leq m$. Cet éclatement est aussi le produit fibré au-dessus de Z des éclatements $\tilde{Z}_i \rightarrow Z$, $1 \leq i \leq m$, de Z le long des $Y_i \times_{f,S} Y_i$. Il est donc muni de diviseurs E_i , $1 \leq i \leq m$, qui sont les images réciproques dans \tilde{Z} des diviseurs exceptionnels dans les \tilde{Z}_i . Pour toute partie non vide $I \subset \{1, \dots, m\}$, on vérifie que

$$E_I := \bigcap_{i \in I} E_i \subset \bigcap_{i \in I} \pi^{-1}(Y_i \times_{f,S} Y_i) = \pi^{-1}(Y_I \times_{f,S} Y_I),$$

que E_I est le produit fibré sur Z des diviseurs exceptionnels dans les \tilde{Z}_i pour $i \in I$ et des \tilde{Z}_i pour $i \notin I$, que E_I est lisse sur S purement de dimension relative $2d - |I|$ et que la restriction de $\pi_I : E_I \rightarrow Y_I \times_{f,S} Y_I$ au-dessus de l'ouvert $X_I \times_{f,S} X_I$ est un fibré projectif de rang $|I|$. Les E_i sont donc des diviseurs lisses sur S et leur réunion $E = E_1 \cup \dots \cup E_m$ est un diviseur à croisements normaux relatif sur S .

Par construction $\pi : \tilde{Z} \rightarrow Z$ est un isomorphisme au dessus de Z_\emptyset . On notera $\tilde{\Gamma} \rightarrow \tilde{Z}$ la normalisation de Γ_\emptyset dans \tilde{Z} . Bien entendu π «envoie» $\tilde{\Gamma}$ dans Γ .

Pour tout point fermé $s \in S$ la fibre $\pi_s : \tilde{Z}_s \rightarrow Z_s$ en s de π est l'éclatement de Z_s le long de la réunion des fermés $Y_{i,s} \times_{f,f(s)} Y_{i,f(s)}$ de $Z_s = X_s \times_{f,f(s)} X_{f(s)}$. Pour tout point fermé $s \in S$ et tout entier $n \geq 0$ tel que $\text{Frob}_S^n(f(s)) = f(\text{Frob}_S^n(s)) = s$, la fibre $\delta_{(s,f(s))}^n : X_{f(s)} \rightarrow Z_s \subset X_s \times X_{f(s)}$ en $(s, f(s))$ du graphe de Frobenius se relève canoniquement en un morphisme

$$\tilde{\delta}_{(s,f(s))}^n : X_{f(s)} \rightarrow \tilde{Z}_s$$

puisque, pour chaque $i = 1, \dots, m$, l'image réciproque par $\delta_{(s,f(s))}^n$ du fermé $Y_{i,s} \times_{f,f(s)} Y_{i,f(s)}$ $Y_{i,f(s)} \subset Z_s$ est le diviseur $Y_{i,f(s)} \subset X_{f(s)}$.

La proposition suivante prolonge un résultat de Pink ([Pi]).

PROPOSITION. — *Supposons que Γ stabilise X_\emptyset au voisinage de ses points fixes. Alors, quitte à remplacer f par*

$$f^{(n_0)} := \text{Frob}_S^{n_0} \circ f = f \circ \text{Frob}_S^{n_0}$$

et Γ_\emptyset par

$$\Gamma_\emptyset^{(n_0)} := (X_\emptyset \times X_\emptyset) \times_{\text{Frob}_{X_\emptyset}^{n_0} \times \text{Id}_{X_\emptyset, X_\emptyset \times X_\emptyset}} \Gamma_\emptyset.$$

pour un entier $n_0 \geq 0$ assez grand, la normalisation $\tilde{\Gamma}$ vérifie la propriété suivante : pour tout point fermé $s \in S$ et tout entier $n > 0$ tel que $\text{Frob}_S^n(f(s)) = f(\text{Frob}_S^n(s)) = s$, l'image de $\tilde{\delta}_{f,f(s)}^n$ dans \tilde{Z}_s ne rencontre pas l'image de $\tilde{\Gamma}_s$ en dehors de $Z_{\emptyset,s} \subset \tilde{Z}_s$.

7.3. Formule des points fixes

Rappelons que l'on a fixé un nombre premier ℓ distinct de la caractéristique de \mathbb{F}_q . Considérons la cohomologie ℓ -adique de X et Z au-dessus de S

$$Rp_*\mathbb{Q}_\ell \text{ et } Rq_*\mathbb{Q}_\ell \in D_c^b(S, \mathbb{Q}_\ell)$$

où $q : X \times_{f,S} X \rightarrow S$ est la projection canonique, et plus généralement les cohomologies ℓ -adique des Y_I au-dessus de S

$$Rp_{I,*}\mathbb{Q}_\ell \text{ et } Rq_{I,*}\mathbb{Q}_\ell \in D_c^b(S, \mathbb{Q}_\ell)$$

où $q_I : Y_I \times_{f,S} Y_I \rightarrow S$ est la projection canonique. La formation de ces cohomologies commute aux changements de base $S' \rightarrow S$. En particulier, la fibre en un point géométrique \bar{s} de S de $Rp_{I,*}\mathbb{Q}_\ell$ est la cohomologie ℓ -adique $R\Gamma(Y_{I,\bar{s}}, \mathbb{Q}_\ell)$ de la fibre de p_I en ce point.

Nos hypothèses assurent que les faisceaux de cohomologie $R^j p_{I,*}\mathbb{Q}_\ell$ (resp. $R^j q_{I,*}\mathbb{Q}_\ell$) sont tous lisses sur S et qu'ils sont nuls pour les j qui ne sont pas dans l'intervalle $[0, 2(d - |I|)]$ (resp. $[0, 4(d - |I|)]$).

Soient $I \subset \{0, 1, \dots, m\}$, n un entier ≥ 1 et $u_I \in H^0(S, R^{2(d-|I|)} q_{I,*}\mathbb{Q}_\ell)(d - |I|)$ une classe de cohomologie. Alors

- l'endomorphisme de Frobenius $\text{Frob}_{Y_I}^n$ qui relève Frob_S^n , induit un homomorphisme

$$\text{Frob}_{Y_I}^n : (\text{Frob}_S^n)^* Rp_{I,*}\mathbb{Q}_\ell \rightarrow Rp_{I,*}\mathbb{Q}_\ell,$$

- d'après la formule de Künneth et la dualité de Poincaré, u_I peut être vu comme un homomorphisme

$$u_I : f^* Rp_{I,*}\mathbb{Q}_\ell \rightarrow Rp_{I,*}\mathbb{Q}_\ell,$$

- les homomorphismes composés

$$\text{Frob}_{Y_I}^n \circ (\text{Frob}_S^n)^*(u_I) : (\text{Frob}_S^n)^* f^* Rp_{I,*}\mathbb{Q}_\ell \rightarrow Rp_{I,*}\mathbb{Q}_\ell$$

et

$$u_I \circ f^*(\text{Frob}_{Y_I}^n) : f^*(\text{Frob}_S^n)^* Rp_{I,*}\mathbb{Q}_\ell \rightarrow Rp_{I,*}\mathbb{Q}_\ell$$

coïncident.

On notera

$$u_I \times \text{Frob}_{Y_I}^n : (\text{Frob}_S^n)^* f^* Rp_{I,*}\mathbb{Q}_\ell = f^*(\text{Frob}_S^n)^* Rp_{I,*}\mathbb{Q}_\ell \rightarrow Rp_{I,*}\mathbb{Q}_\ell$$

ces homomorphismes composés. Pour tout point géométrique \bar{s} de S qui vérifie $f(\text{Frob}_S^n(\bar{s})) = \text{Frob}_S^n(f(\bar{s})) = \bar{s}$, $u_I \times \text{Frob}_{Y_I}^n$ induit un endomorphisme de la cohomologie $R\Gamma(Y_{I,\bar{s}}, \mathbb{Q}_\ell)$ de la fibre en \bar{s} de $p_I : Y_I \rightarrow S$ dont on notera

$$\text{tr}(u_I \times \text{Frob}_{Y_I}^n, R\Gamma(Y_{I,\bar{s}}, \mathbb{Q}_\ell)) \in \mathbb{Q}_\ell$$

la trace.

Le «cycle de codimension d » Γ sur $X \times_{f,S} X$ admet une classe de cohomologie

$$\text{cl}(\Gamma) \in H^0(S, R^{2d} q_* \mathbb{Q}_\ell)(d).$$

De même, si on note $\tilde{q} = q \circ \pi : \tilde{Z} \rightarrow S$, $\tilde{\Gamma}$ admet une classe de cohomologie

$$\text{cl}(\tilde{\Gamma}) \in H^0(S, R^{2d} \tilde{q}_* \mathbb{Q}_\ell)(d).$$

Pour chaque $I \subset \{1, \dots, m\}$ non vide, on a un homomorphisme de faisceaux

$$R^{2d} \tilde{q}_* \mathbb{Q}_\ell(d) \rightarrow R^{2d} \tilde{q}_{I,*} \mathbb{Q}_\ell(d) \rightarrow R^{2(d-|I|)} q_{I,*} \mathbb{Q}_\ell(d-|I|)$$

où $\tilde{q}_I : E_I \rightarrow S$ est la projection canonique, où la première flèche est l'homomorphisme de restriction au fermé $E_I \subset \tilde{Z}$ et où la seconde flèche est duale de l'homomorphisme

$$\pi_I^* : R^{2(d-|I|)} q_{I,*} \mathbb{Q}_\ell(d-|I|) \rightarrow R^{2(d-|I|)} \tilde{q}_{I,*} \mathbb{Q}_\ell(d-|I|).$$

On note $\text{cl}(\Gamma)_I$ l'image de $\text{cl}(\tilde{\Gamma})$ par cet homomorphisme composé.

Plus généralement on peut refaire cette construction après avoir remplacé f par $f^{(n_0)}$ et Γ_\emptyset par $\Gamma_\emptyset^{(n_0)}$ pour un entier $n_0 \geq 0$ (cf. la proposition de 7.2). Pour chaque $I \subset \{1, \dots, m\}$ non vide, on obtient une section globale $\text{cl}(\Gamma^{(n_0)})_I$ de $R^{2(d-|I|)} q_{I,*}^{(n_0)} \mathbb{Q}_\ell(d-|I|)$, où $q_I^{(n_0)} : Y_I \times_{f^{(n_0)}, S} Y_I \rightarrow S$ est la projection canonique. L'image directe de cette section globale par le morphisme radiciel $\text{Frob}_{Y_I}^{n_0} \times \text{Id}_{Y_I} : Y_I \times_{f^{(n_0)}, S} Y_I \rightarrow Y_I \times_{f, S} Y_I$ est une section globale notée $\text{cl}(\Gamma)_I^{(n_0)}$ de $R^{2(d-|I|)} q_{I,*} \mathbb{Q}_\ell(d-|I|)$.

Avec ces notations, Lafforgue prouve la généralisation suivante de la formule des points fixes de Grothendieck-Lefschetz ([Gr]).

THÉORÈME. — *Supposons que Γ stabilise l'ouvert X_\emptyset au voisinage de ses points fixes et soit $n_0 \geq 0$ un entier vérifiant la conclusion de la proposition de 7.3.*

Alors, pour tout point fermé s de S et tout entier $n > n_0$ tels que $\text{Frob}_S^n(f(s)) = f(\text{Frob}_S^n(s)) = s$, on a

$$\begin{aligned} \text{Lef}_s(\Gamma_\emptyset \times \text{Frob}_{X_\emptyset}^n, X_\emptyset) &= \text{tr}(\text{cl}(\Gamma) \times \text{Frob}_X^n, R\Gamma(X_{\bar{s}}, \mathbb{Q}_\ell)) \\ &+ \sum_{\substack{I \subset \{1, \dots, m\} \\ I \neq \emptyset}} (-1)^{|I|} \text{tr}(\text{cl}(\Gamma)_I^{(n_0)} \times \text{Frob}_{Y_I}^n, R\Gamma(Y_{I, \bar{s}}, \mathbb{Q}_\ell)). \end{aligned}$$

7.4. Extension à certains champs algébriques

Comme le montre Lafforgue, les résultats de cette section valent encore après avoir remplacé le S -schéma $p : X \rightarrow S$, le diviseur $Y = Y_1 \cup \dots \cup Y_m$ et le morphisme de schémas $\Gamma_\emptyset \rightarrow Z_\emptyset$ par un S -champ algébrique $p : \mathcal{X} \rightarrow S$, un diviseur \mathcal{Y} dans \mathcal{X} et un morphisme de champs algébriques $\Gamma_\emptyset \rightarrow \mathcal{Z}_\emptyset$ qui vérifient les propriétés suivantes :

- le morphisme $p : \mathcal{X} \rightarrow S$ est de type fini,
- au voisinage de chacun de ses points le champ \mathcal{X} admet un recouvrement fini et plat par un espace algébrique muni de l'action d'un groupe fini qui agit transitivement sur les fibres de ce recouvrement,

- le morphisme p satisfait au critère valuatif de propreté,
- le morphisme p est lisse purement de dimension relative d ,
- \mathcal{Y} est une réunion de diviseurs $\mathcal{Y}_1, \dots, \mathcal{Y}_m$ qui sont lisses sur S et est à croisement normaux relatif sur S ,
- \mathcal{X}_\emptyset est un S -champ algébrique de Deligne-Mumford, le morphisme $\Gamma_\emptyset \rightarrow \mathcal{Z}_\emptyset = \mathcal{X}_\emptyset \times_{f,S} \mathcal{X}_\emptyset$ est représentable fini et la première projection de Γ_\emptyset sur \mathcal{X}_\emptyset est étale.

8. FIN DE LA RÉCURRENCE

8.1. Ce qu'il reste à démontrer

Soient $N \subset X$ un niveau. Pour tout $g \in \mathrm{GL}_r(\mathbb{A})$ la correspondance de Hecke

$$c = (c_1, c_2) : \mathrm{Cht}_N^r(g)/a^{\mathbb{Z}} \rightarrow (\mathrm{Cht}_N^r/a^{\mathbb{Z}}) \times_{(X-N)^2} (\mathrm{Cht}_N^r/a^{\mathbb{Z}})$$

envoie $\mathrm{Cht}_N^{r; \leq p}/a^{\mathbb{Z}}$ dans $\mathrm{Cht}_N^{r; \leq q}/a^{\mathbb{Z}}$ dès que $q - p$ est assez convexe en fonction de $K_N g K_N$. Par conséquent, le ind-système local ℓ -adique $H_N^{r; 2r-2}$ défini en 6.4 est muni d'une action de la $\overline{\mathbb{Q}}_\ell$ -algèbre de Hecke $\mathcal{H}(K_N) = \mathcal{C}_c(\mathrm{GL}_r(\mathbb{A})//K_N)$ des fonctions à valeurs dans $\overline{\mathbb{Q}}_\ell$, K_N -invariantes à gauche et à droite et à supports compacts sur $\mathrm{GL}_r(\mathbb{A})$.

De même, les morphismes de Frobenius partiels induisent un isomorphisme

$$(\mathrm{Frob}_{X-N} \times \mathrm{Id}_{X-N})^* H_N^{r; 2r-2} \xrightarrow{\sim} H_N^{r; 2r-2}$$

qui commute à l'action de $\mathcal{H}(K_N)$.

Comme la construction faite en 6.4 du système local ℓ -adique $H_{N, \mathrm{cusp}}^r$ à partir de $H_N^{r; 2r-2}$ est canonique, $H_{N, \mathrm{cusp}}^r$ est lui aussi muni d'une action de l'algèbre de Hecke $\mathcal{H}(K_N)$ et d'un isomorphisme $(\mathrm{Frob}_{X-N} \times \mathrm{Id}_{X-N})^* H_{N, \mathrm{cusp}}^r \xrightarrow{\sim} H_{N, \mathrm{cusp}}^r$. Dans l'énoncé de la dernière proposition de la section 6.4, on peut donc remplacer la moyenne $\frac{1}{r!} \sum_{k=1}^{r!} [(\mathrm{Frob}_{X-N}^k \times \mathrm{Id}_{X-N})^* H_{N, \mathrm{cusp}}^r]$ par $[H_{N, \mathrm{cusp}}^r]$.

Dans cette dernière section nous allons montrer :

PROPOSITION. — *Pour tout $g \in \mathrm{GL}_r(\mathbb{A})$, tout couple (∞, o) de points fermés distincts de $X - S_g$ (où $S_g \subset |N|$ est l'ensemble fini de points fermés de X défini en 5.1), tout point fermé ξ de $\infty \times o$ et tout entier n , on a*

$$\mathrm{Tr}(1_{K_N g K_N} \times \mathrm{Frob}_\xi^n, H_{N, \mathrm{cusp}}^r) \\ = q^{(r-1) \deg(\xi)n} \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N g K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi).$$

Cette proposition généralise la partie (iii) de la dernière proposition de la sous-section 6.4 et des arguments standard permettent d'en déduire les propriétés $P_1(r)$, $P_2(r)$ et $P_3(r)$.

8.2. Trace des opérateurs de Hecke sur $\tilde{H}_{N,\text{cusp}}^{r;\leq p}$

LEMME. — Soient U un ouvert non vide de X et M un système local ℓ -adique sur $U \times U$ muni d'un endomorphisme $h : M \rightarrow M$. Alors, un fois fixé un ensemble de représentants des classes d'isomorphie de systèmes locaux ℓ -adiques irréductibles sur $U \times U$, il existe une unique application à support fini $L \mapsto \lambda_L$ de cet ensemble dans $\overline{\mathbb{Q}}_\ell$ telle que

$$\text{Tr}(h \circ \gamma, M_{\bar{\delta}}) = \sum_L \lambda_L \text{Tr}(\gamma, L_{\bar{\delta}})$$

pour tout $\gamma \in \pi_1(U \times U, \bar{\delta})$.

Si $[M] = \sum_L m_M(L)[L]$ est le système local ℓ -adique virtuel sur $U \times U$ associé à M et si $M_{\text{cusp}} = \sum_L m_{M_{\text{cusp}}}(L)[L]$ est son morceau non r -négligeable (cf. 6.4), on notera

$$\text{Tr}_{M_{\text{cusp}}}(h \circ \gamma) = \sum_{\substack{L \\ m_M(L) \neq 0}} \frac{m_{M_{\text{cusp}}}(L)}{m_M(L)} \lambda_L \text{Tr}(\gamma, L_{\bar{\delta}})$$

pour tout $\gamma \in \pi_1(U \times U, \bar{\delta})$.

Bien entendu, on va appliquer ceci au morceau non r -négligeable $\tilde{H}_{N,\text{cusp}}^{r;\leq p}$ de la cohomologie ℓ -adique $\tilde{H}_N^{r;\leq p,*}$ et à l'endomorphisme de $\tilde{H}_N^{r;\leq p,2r-2}$ induit par une correspondance de Hecke.

Plus précisément, soit $p : [0, r] \rightarrow \mathbb{R}$ un paramètre de troncature assez convexe par rapport au genre de la courbe X et à N et soit $g \in \text{GL}_r(\mathbb{A})$. D'une part, on a déjà vu que la correspondance de Hecke $c : \text{Cht}_N^r(g) \rightarrow \text{Cht}_N^r \times_{(X-N)^2} \text{Cht}_N^r$ envoie $\text{Cht}_N^{r;\leq p}$ dans $\text{Cht}_N^{r;\leq q}$ pour un paramètre de troncature $q \geq p$ convenable et qu'elle induit un morphisme entre les cohomologies à supports compacts $h : H_N^{r;\leq p,*} \rightarrow H_N^{r;\leq q,*}$. D'autre part, cette correspondance induit par restriction un cycle

$$c^{\leq p} : \text{Cht}_N^{r;\leq p}(g) \rightarrow \text{Cht}_N^{r;\leq p} \times_{(X-N)^2} \text{Cht}_N^{r;\leq p},$$

puis par normalisation une correspondance

$$\tilde{c}^{\leq p} : \widetilde{\text{Cht}}_N^{r;\leq p}(g) \rightarrow \widetilde{\text{Cht}}_N^{r;\leq p} \times_{(X-N)^2} \widetilde{\text{Cht}}_N^{r;\leq p},$$

et donc un endomorphisme noté \tilde{h} de la cohomologie ℓ -adique $\tilde{H}_N^{r;\leq p,*}$, d'où une trace

$$\text{Tr}_{\tilde{H}_{N,\text{cusp}}^{r;\leq p}}(\tilde{h} \circ \gamma)$$

pour tout $\gamma \in \pi_1((X-N)^2, \bar{\delta})$.

Quitte à agrandir q , on peut supposer que q est assez convexe par rapport au genre de la courbe X et à N . On alors un diagramme commutatif

$$\begin{array}{ccccc}
 H_N^{r; \leq q, *} & \longrightarrow & & & \widetilde{H}_N^{r; \leq q, *} \\
 h \uparrow & & & + & \downarrow \\
 H_N^{r; \leq p, *} & \longrightarrow & \widetilde{H}_N^{r; \leq p, *} & \xrightarrow{\widetilde{h}} & \widetilde{H}_N^{r; \leq p, *}
 \end{array}$$

où les flèches h et \widetilde{h} sont celles définies ci-dessus et les autres flèches sont les mêmes que dans le carré commutatif de 6.4. Par un argument facile utilisant les résultats de 6.4, Lafforgue en déduit :

PROPOSITION. — *Pour tout paramètre de troncature p assez convexe en fonction de N , tout point fermé ξ de $(X - N)^2$ et tout entier n , on a*

$$\text{Tr}(1_{K_N g K_N} \times \text{Frob}_\xi^n, H_{N, \text{cusp}}^r) = \text{Tr}_{\widetilde{H}_{N, \text{cusp}}^{r; \leq p}}(\widetilde{h} \circ \text{Frob}_\xi^n).$$

8.3. Application de la variante du théorème de Pink

Fixons un paramètre de troncature $p : [0, r] \rightarrow \mathbb{R}$ assez convexe par rapport au genre de la courbe X et à N , et un élément $g \in \text{GL}_r(\mathbb{A})$. On a défini un ensemble fini $S_g \supset |N|$ de points fermés de X .

On cherche à appliquer les résultats généraux du chapitre 7 à

$$S = (X - S_g)^2, \quad f = \text{Id}_S : S \rightarrow S,$$

$$\begin{aligned}
 \mathcal{X}_\emptyset &= (X - S_g)^2 \times_{(X-N)^2} (\text{Cht}_N^{r; \leq p} / a^{\mathbb{Z}}) \subset (X - S_g)^2 \times_{(X-N)^2} (\widetilde{\text{Cht}}_N^{r; \leq p} / a^{\mathbb{Z}}) = \mathcal{X}, \\
 (c_1, c_2) : (X - S_g)^2 \times_{(X-N)^2} (\text{Cht}_N^r(g) / a^{\mathbb{Z}}) &= \Gamma_\emptyset \rightarrow \mathcal{X}_\emptyset \times_S \mathcal{X}_\emptyset,
 \end{aligned}$$

et

$$\Gamma \rightarrow \mathcal{X} \times_S \mathcal{X}$$

la normalisation de $\mathcal{X}_\emptyset \times_S \mathcal{X}_\emptyset$ dans Γ_\emptyset . Le fermé complémentaire $\mathcal{Y} = \mathcal{X} - \mathcal{X}_\emptyset$ est un diviseur à croisements normaux relatif sur $(X - S_g)^2$, qui est réunion $\mathcal{Y} = \mathcal{Y}_1 \cup \dots \cup \mathcal{Y}_m$ de diviseurs lisses sur $(X - S_g)^2$.

Pour chaque entier $n_0 \geq 0$, on dispose des classes de cohomologie suivantes :

- $\text{cl}(\Gamma)$ dans la cohomologie de $\mathcal{X} \times_S \mathcal{X}$ au-dessus de $(X - S_g)^2$,
- $\text{cl}(\Gamma)_I^{(n_0)}$ dans la cohomologie de $\mathcal{Y}_I \times_S \mathcal{Y}_I$ au-dessus de $(X - S_g)^2$ pour tout $I \subset \{1, \dots, m\}$ non vide.

Pour tout ouvert U de X notons Λ_U le schéma intersection de tous les ouverts de $U \times U$ complémentaires des images réciproques de la diagonales par les endomorphismes $\text{Frob}_U^n \times \text{Id}_U$ et $\text{Id}_U \times \text{Frob}_U^n$ de $U \times U$ pour tous les entiers $n \geq 0$, c'est-à-dire l'intersection de $\Lambda_X \subset X \times X$ avec $U \times U$.

Nous avons besoin du théorème suivant de Lafforgue, qui prolonge en rang arbitraire un résultat établi par Drinfeld en rang 2 ([Dr 7]).

THÉORÈME. — *Supposons que p est assez convexe et $U \subset X - S_g$ est assez petit relativement à N et à $K_N g K_N$. Il existe alors un ouvert de Zariski $V \subset U \times U$, contenant Λ_U , tel que, après avoir remplacé $S = (X - S_g)^2$ par V et $\mathcal{X}_\emptyset \subset \mathcal{X}$ et $\Gamma_\emptyset \subset \Gamma$ par leurs restrictions à cet ouvert de S , la correspondance Γ stabilise l'ouvert \mathcal{X}_\emptyset au voisinage de ses points fixes.*

En d'autres termes, l'hypothèse du théorème de la sous-section 7.3 est vérifiée au-dessus de $V \supset \Lambda_U$. Par conséquent, si $n_0 \geq 0$ est l'entier de loc. cit., pour tout point géométrique $\bar{\xi}$ localisé en un point fermé ξ de Λ_U et tout entier n tel que $\deg(\xi)n > n_0$, on a la formule de points fixes

$$\begin{aligned} \text{Lef}_\xi(\Gamma_\emptyset \times \text{Frob}_{\mathcal{X}_\emptyset}^{\deg(\xi)n}, \mathcal{X}_\emptyset) &= \text{tr}(\text{cl}(\Gamma) \times \text{Frob}_{\mathcal{X}}^{\deg(\xi)n}, R\Gamma(\mathcal{X}_{\bar{\xi}}, \mathbb{Q}_\ell)) \\ &+ \sum_{I \neq \emptyset} (-1)^{|I|} \text{tr}(\text{cl}(\Gamma)_I^{(n_0)} \times \text{Frob}_{\mathcal{Y}_I}^{\deg(\xi)n}, R\Gamma(\mathcal{Y}_{I, \bar{\xi}}, \mathbb{Q}_\ell)). \end{aligned}$$

On remarquera que, pour chaque $I \subset \{1, \dots, m\}$ non vide, il existe une famille $(m_I(L))_L$ de scalaires dans $\overline{\mathbb{Q}}_\ell$, indexée par un ensemble de représentants des systèmes locaux ℓ -adiques irréductibles sur $(X - S_g)^2$, à support fini, telle que

$$\text{tr}(\text{cl}(\Gamma)_I^{(n_0)} \times \text{Frob}_{\mathcal{Y}_I}^{\deg(\xi)n}, R\Gamma(\mathcal{Y}_{I, \bar{\xi}}, \mathbb{Q}_\ell)) = \sum_L m_I(L) \text{Tr}(\text{Frob}_\xi^n, L),$$

et que $m_I(L) = 0$ pour tout L qui n'est pas r -négligeable d'après 6.3.

Dans la suite de cette sous-section, on identifie $\overline{\mathbb{Q}}_\ell$ à \mathbb{C} par l'isomorphisme fixé en 1.3.

Dans la formule de points fixes ci-dessus, on peut remplacer $\bar{\xi}$ par $(\text{Frob}_X^k \times \text{Id}_X)(\bar{\xi})$ pour $k = 1, \dots, r!$ et faire la moyenne des $r!$ expressions obtenues. Le premier membre de cette moyenne est égal à

$$\begin{aligned} &\sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \text{Tr}_\pi(1_{K_N g K_N}) q^{(r-1) \deg(\xi)n} S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi) \\ &+ \sum_{\substack{1 \leq r' < r \\ 1 \leq r'' < r}} \sum_{\substack{\pi' \in \mathcal{A}_{r'}(K_N) \\ \pi'' \in \mathcal{A}_{r''}(K_N)}} \text{Tr}_{\pi', \pi''}^{\leq \alpha p}(1_{K_N g K_N}, \deg(\xi)n) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi') S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi'') \end{aligned}$$

d'après 5.2. Compte tenu de la proposition de la sous-section précédente et de l'hypothèse de récurrence $P_1(r')$ pour tout $r' < r$ (cf. 6.2), on a donc une expression pour la différence

$$\begin{aligned} &\text{Tr}(1_{K_N g K_N} \times \text{Frob}_\xi^n, H_{N, \text{cusp}}^r) \\ &- q^{(r-1) \deg(\xi)n} \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \text{Tr}_\pi(1_{K_N g K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi) \end{aligned}$$

de la forme

$$\sum_L P_L(n \deg(\xi)) \operatorname{Tr}(\operatorname{Frob}_\xi^n, L)$$

où $(\nu \mapsto P_L(\nu))_L$ est une famille à support fini de fonctions complexes de l'entier ν de la forme

$$P_L(\nu) = \sum_z P_{L,z}(\nu) z^\nu$$

pour une famille à support fini de polynômes $P_{L,z}(u) \in \mathbb{C}[u]$, $z \in \mathbb{C}^\times$, qui sont identiquement nuls quand L n'est pas r -négligeable.

En développant les polynômes $P_{L,z}(u)$ et en utilisant l'indépendance linéaire des fonctions $\nu \mapsto \nu^k z^\nu$ de l'entiers ν pour $k \in \mathbb{Z}$ et $z \in \mathbb{C}^\times$, on voit tout d'abord que l'on ne perd rien en remplaçant chaque $P_{L,z}(u)$ par son terme constant $P_{L,z}(0)$.

En utilisant la correspondance de Langlands et la conjecture de Ramanujan-Petersson déjà connues en les rangs $< r$ et un dernier argument de fonctions L , Lafforgue conclut alors la preuve de la proposition de 8.1 et la récurrence.

BIBLIOGRAPHIE

Travaux de L. Lafforgue

- [La 1] L. LAFFORGUE – *Chtoucas de Drinfeld et conjecture de Ramanujan-Petersson*, *Astérisque*, 243, (1997).
- [La 2] L. LAFFORGUE – Sur la conjecture de Ramanujan-Petersson pour les corps de fonctions. I : Étude géométrique, *C. R. Acad. Sci. Paris* 322, (1996), 605-608.
- [La 3] L. LAFFORGUE – Sur la conjecture de Ramanujan-Petersson pour les corps de fonctions. II : Étude spectrale, *C. R. Acad. Sci. Paris* 322, (1996), 707-710.
- [La 4] L. LAFFORGUE – Sur la dégénérescence des chtoucas de Drinfeld, *C. R. Acad. Sci. Paris* 323, (1996), 491-494.
- [La 5] L. LAFFORGUE – Compactification de l'isogénie de Lang et dégénérescence des structures de niveau simple des chtoucas de Drinfeld, *C. R. Acad. Sci. Paris* 325, (1997), 1309-1312.
- [La 6] L. LAFFORGUE – Chtoucas de Drinfeld et applications, dans *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, *Doc. Math.*, (1998), 563-570.
- [La 7] L. LAFFORGUE – Une compactification des champs classifiant les chtoucas de Drinfeld, *J. Amer. Math. Soc.* 11, (1998), 1001-1036.
- [La 8] L. LAFFORGUE – Pavages des simplexes, schémas de graphes recollés et compactification des $\operatorname{PGL}_r^{n+1} / \operatorname{PGL}_r$, *Invent. Math.* 136, (1999), 233-271.
- [La 9] L. LAFFORGUE – Compactification des $\operatorname{PGL}_r^{n+1} / \operatorname{PGL}_r$ et restriction des scalaires à la Weil, à paraître dans le volume du Cinquantenaire des Annales de l'Institut Fourier.

- [La 10] L. LAFFORGUE – La correspondance de Langlands sur les corps de fonctions, manuscrit (en cours de frappe), juin 1999.

Articles de V.G. Drinfeld sur les chtoucas

- [Dr 1] V. G. DRINFELD – Commutative subrings of certain noncommutative rings, *Funct. Anal. and its Appl.* 11, (1977), 9-12.
- [Dr 2] V. G. DRINFELD – Proof of the global Langlands conjecture for $GL(2)$ over a function field, *Funct. Anal. and its Appl.* 11, (1977), 223-225.
- [Dr 3] V. G. DRINFELD – A proof of Petersson’s conjecture for function fields, *Uspehi Mat. Nauk* 32, (1977), 209-210.
- [Dr 4] V. G. DRINFELD – Langlands’ conjecture for $GL(2)$ over functional fields, dans *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, Acad. Sci. Fennica, (1980), 565-574.
- [Dr 5] V. G. DRINFELD – Moduli varieties of F -sheaves, *Funct. Anal. and its Appl.* 21, (1987), 107-122.
- [Dr 6] V. G. DRINFELD – Proof of the Petersson conjecture for $GL(2)$ over a global field of characteristic p , *Funct. Anal. and its Appl.* 22, (1988), 28-43.
- [Dr 7] V. G. DRINFELD – Cohomology of compactified moduli varieties of F -sheaves of rank 2, *J. of Soviet Math.* 46, (1989), 1789-1821.
- [Dr 8] V. G. DRINFELD – On my paper : “Cohomology of compactified moduli varieties of F -sheaves of rank 2”, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI), Anal. Teor. Chisel i Teor. Funktsii.* 9 168, (1988), 45-47.

Autres publications sur la correspondance de Langlands

- [Ca] H. CARAYOL – Variétés de Drinfeld compactes, d’après Laumon, Rapoport et Stuhler, dans *Séminaire Bourbaki 1991/92*, Astérisque 206, (1992), 369-409.
- [D-H] P. DELIGNE, D. HUSEMOLLER – Survey of Drinfeld modules, dans *Current trends in arithmetical algebraic geometry (Arcata, 1985)*, Contemp. Math. 67, (1987), 25-91.
- [Dr 9] V.G. DRINFELD – Elliptic modules, *Math. USSR Sbornik* 23, (1974), 561-592.
- [Dr 10] V.G. DRINFELD – Elliptic modules II, *Math. USSR Sbornik* 31, (1977), 159-170.
- [Dr 11] V. G. DRINFELD – Number of two-dimensional irreducible representations of the fundamental group of a curve over a finite field, *Functional Anal. and its Appl.* 15, (1982), 294-295.
- [F-K] Y. FLICKER, D. KAZHDAN – Geometric Ramanujan conjecture and Drinfeld reciprocity law, dans *Number theory, trace formulas and discrete groups*, Academic Press, (1989), 201-218.

- [H-K] G. HARDER, D. KAZHDAN – Automorphic forms on GL_2 over function fields (after V. G. Drinfeld), *Proc. Sym. Pure Math.* 33, Part 2, (1979), 357-379.
- [Ka] D. KAZHDAN – An introduction to Drinfeld’s “shtuka”, *Proc. Sym. Pure Math.* 33, Part 2, (1979), 347-356.
- [Lau 1] G. LAUMON – *Cohomology of Drinfeld modular varieties, Part I (Geometry, counting of points and local harmonic analysis)*, Cambridge University Press, (1995).
- [Lau 2] G. LAUMON – *Cohomology of Drinfeld modular varieties, Part II (Automorphic forms, trace formulas and Langlands correspondence)*, Cambridge University Press, (1996).
- [Lau 3] G. LAUMON – Drinfeld Shtukas, dans *CIME Session “Vector bundles on Curves. New Directions” (Cetraro, juin 1995)*, Lecture Notes in Mathematics 1649, (1996), 50-109.
- [Lau 4] G. LAUMON – The Langlands Correspondence for Function Fields following Laurent Lafforgue, dans *Current Developments in Mathematics Conference (Harvard University, novembre 1999)*.
- [L-R-S] G. LAUMON, M. RAPOPORT et U. STUHLER – D-elliptic sheaves and the Langlands correspondence, *Inv. Math.* 113, (1993), 217-338.

Sur la correspondance de Langlands géométrique

- [Dr 12] V. G. DRINFELD – Two-dimensional l -adic representations of the fundamental group of a curve over a finite field and automorphic forms on $GL(2)$, *Amer. J. Math.* 105, (1983), 85-114.
- [Dr 13] V. G. DRINFELD – Two-dimensional l -adic representations of the Galois group of a global field of characteristic p and automorphic forms on $GL(2)$, *J. of Soviet Math.* 36, (1987), 93-105.
- [F-G-K-V] E. FRENKEL, D. GAITSGORY, D. KAZHDAN, K. VILONEN – Geometric realization of Whittaker functions and the Langlands conjecture, *J. Amer. Math. Soc.* 11, (1998), 451-484.
- [Lau 5] G. LAUMON – Correspondance de Langlands géométrique pour les corps de fonctions, *Duke Math. J.* 54, (1987), 309-359.
- [Lau 6] G. LAUMON – Faisceaux automorphes pour GL_n : la première construction de Drinfeld, prépublication électronique, réf. alg-geom/9511004.

Autres références de géométrie algébrique

- [De] P. DELIGNE – La conjecture de Weil. II, *Publ. Math. IHES* 52, (1980), 137-252.
- [DC-P] C. DE CONCINI, C. PROCESI – Complete symmetric varieties, dans *CIME Session “Invariant Theory” (Montecatini)*, Lecture Notes in Math. 996, Springer-Verlag, (1982), 1-44.

- [Fa] G. FALTINGS – Explicit resolution of local singularities of moduli-spaces, *Journal für die reine und angewandte Mathematik* 483, (1997), 183-196.
- [Ge] A. GENESTIER – A toroidal resolution for the bad reduction of some Shimura varieties, <http://xxx.lanl.gov/abs/math/9912054>, (1999).
- [Gr] A. GROTHENDIECK – Formule de Lefschetz et rationalité des fonctions L , Sémin. Bourbaki 1964/65, exp. n° 279, collection hors série Astérisque **9** (1995), 41-55, et dans *Dix exposés sur la cohomologie des schémas*, North Holland, (1968), 31-45.
- [H-N] G. HARDER, M.S. NARASIMHAN – On the cohomology groups of moduli spaces of vector bundles on curves, *Math. Ann.* 212, (1975), 215-248.
- [H-T] M. HARRIS, R. TAYLOR – On the geometry and cohomology of some simple Shimura varieties, Institut de Mathématiques de Jussieu, Prépublication 227, (1999).
- [Lak] D. LAKSOV – Completed quadrics and linear maps, dans *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*, Proc. Sympos. Pure Math., 46, Part 2, Amer. Math. Soc., (1987), 371-387.
- [Lau 7] G. LAUMON – Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil, *Publ. Math. I.H.E.S.* 65, (1987), 131-210.
- [L-M] G. LAUMON, L. MORET-BAILLY – *Champs algébriques*, Springer-Verlag, (1999).
- [Pi] R. PINK – On the calculation of local terms in the Lefschetz-Verdier trace formula and its application to a conjecture of Deligne, *Ann. Math.* 135, (1992), 483-525.
- [SD] B. SAINT-DONAT – Affine Embeddings, dans *Toroidal Embeddings I*, Lecture Notes in Math. 339, Springer-Verlag, (1973), 1-40.
- [Se] J.-P. SERRE – *Abelian ℓ -Adic Representations and Elliptic Curves*, Addison-Wesley, (1988).

Autres références automorphes

- [Ar] J. ARTHUR – A trace formula for reductive groups I : terms associated to classes in $G(\mathbb{Q})$, *Duke Math. J.* 45, (1978), 911-953.
- [C-PS] J. W. COGDELL, I. I. PIATETSKI-SHAPIRO – Converse theorems for GL_n , *Publ. Math. IHES* 79, (1994), 157-214.
- [G-J] R. GODEMENT, H. JACQUET – Zeta functions of simple algebras, Lecture Notes in Math. 260, Springer-Verlag, (1972).
- [J-PS-S] H. JACQUET, I. I. PIATETSKI-SHAPIRO, J.A. SHALIKA – Rankin-Selberg convolutions, *Amer. J. Math.* 105, (1983), 367-464.
- [J-S 1] H. JACQUET, J.A. SHALIKA – On Euler products and the classification of automorphic representations I, *Amer. J. Math.* 103, (1981), 499-558.

- [J-S 2] H. JACQUET, J.A. SHALIKA – On Euler products and the classification of automorphic representations II, *Amer. J. Math.* 103, (1981), 777-815.
- [Lan 1] R.P. LANGLANDS – Problems in the theory of automorphic forms, dans *Lectures in modern analysis and applications III*, Lecture Notes in Math. 170, Springer-Verlag, (1970), 18-61.
- [Lan 2] R.P. LANGLANDS – *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. 544, Springer-Verlag, (1976).
- [M-W] C. MOEGLIN, J.-L. WALDSPURGER – *Décomposition spectrale et séries d'Eisenstein, Une paraphrase de l'Écriture*, Birkhäuser, (1994).
- [PS 1] I.I. PIATETSKI-SHAPIRO – Zeta functions of $GL(n)$, prépublication de L'université du Maryland, (1976).
- [PS 2] I.I. PIATETSKI-SHAPIRO – Multiplicity one theorem, *Proc. Sym. Pure Math.* 33, Part 1, (1979), 209-212.
- [Sh] J.A. SHALIKA – The multiplicity one theorem for GL_n , *Ann. of Math.* 100, (1974), 171-193.

Gérard LAUMON

Université de Paris-Sud
et CNRS, UMR 8628
Mathématiques, Bât. 425
F-91405 Orsay Cedex (France)
Gerard.Laumon@math.u-psud.fr

MOTIVIC MEASURES

by **Eduard LOOIJENGA**

1. INTRODUCTION

An n -jet of an arc in an algebraic variety is a one parameter Taylor series of length n in that variety. To be precise, if the variety X is defined over the algebraically closed field k , then it is a $k[[t]]/(t^{n+1})$ -valued point of X . The set of such n -jets are the closed points of a variety $\mathcal{L}_n(X)$ also defined over k and the arc space of X , $\mathcal{L}(X)$, is the projective limit of these. Probably Nash [23] was the first to study arc spaces in a systematic fashion (the paper in question was written in 1968). He concentrated on arcs based at a given $p \in X$ and observed that to each irreducible component of this ‘provariety’ there corresponds in an injective manner an irreducible component of the preimage of p in any resolution of X . He asked the (still unanswered) question how to identify these components on a given resolution. The renewed interest in arc spaces has a different origin, however. Batyrev [4] proved that two connected projective complex manifolds with trivial canonical bundle which are birationally equivalent must have the same Betti numbers. This he showed by first lifting the data to a situation over a discrete valuation ring with finite residue field and then exploiting a p -adic integration technique. (Such a p -adic integration approach to problems in complex algebraic geometry had also been used by Denef and Loeser [12] in their work on topological zeta functions attached to singular points of complex varieties.) When Kontsevich learned of Batyrev’s result he saw how this proof could be made to work in a complex setting using arc spaces. The new proof also gave more: equality of Hodge numbers, and even an isomorphism of Hodge structures with rational coefficients. The underlying technique, now going under the name of *motivic integration*, has led to an avalanche of applications. These include new (so-called *stringy*) invariants of singularities, a complex analogue of the Igusa zeta function, a motivic version of the Thom-Sebastiani property and the motivic McKay correspondence. Some of these were covered in a recent talk by Reid [25] in this seminar.

The idea is simple if we keep in mind an analogous, more classical situation. Consider the case of a complete discrete valuation ring (R, m) with finite residue field F . There is a Haar measure on the Boolean algebra consisting of the cosets of powers of m that takes the value 1 on R (so it is also a probability measure). This induces one on a suitable Boolean algebra of subsets of the set of R -valued points of any scheme that is flat of pure dimension and of finite type over $\text{Spec}(R)$. Associated to this measure is a function that essentially counts the number of ‘points’ in each reduction modulo m^k : the *Igusa zeta function*, introduced by Weil, and intensively studied by Igusa, Denef and Loeser (and reported on by Denef in this seminar [11]). A missing case was that of equal characteristic zero: $\mathcal{O} = k[[t]]$, $k \supset \mathbb{Q}$. The proposal of Kontsevich is to give \mathcal{O} a measure that takes values in a Grothendieck ring of k -varieties in which the class of the affine line, \mathbb{L} , is invertible: the value on the ideal (t^n) is then simply \mathbb{L}^{-n} (or \mathbb{L}^{1-n} , which is sometimes more convenient). If \mathcal{X} is a suitable \mathcal{O} -scheme, then we obtain a measure on the set of sections as before, but now with values in this Grothendieck ring. The corresponding zeta function is a very fine bookkeeping device, for it does its counting in a ring that is huge. There seems no a priori reason to restrict to the case of equal characteristic, for Kontsevich’s idea makes sense for any complete discrete valuation ring. Indeed, with little extra effort the material in Sections 2,3 and 9 can be generalized to that context.

This report concerns mainly work of Denef and Loeser. Some of their results are presented here somewhat differently, and this is why more proofs are provided than one perhaps expects of the write up of a seminar talk. References to the sources are in general given after the section titles, rather than in the statements of theorems.

I thank Jan Denef for inviting me for a short visit to Leuven to discuss the material exposed here. I also indebted to him and Maxim Kontsevich for comments on an earlier version.

2. THE ARC SPACE AND ITS MEASURE [14], [22]

Throughout the talk we fix a complete discrete valuation ring \mathcal{O} whose residue field k is assumed to be algebraically closed and of characteristic zero. The spectrum of \mathcal{O} is denoted \mathbb{D} with generic point \mathbb{D}^* and closed point o . A uniformizing parameter is often denoted by t so that $\mathcal{O} = k[[t]]$. The assumption that k be algebraically closed is for convenience only: in most situations this restriction is unnecessary or can be avoided.

The symbol \mathbb{N} stands for the set of nonnegative integers.

The Grothendieck ring of varieties

Consider the *Grothendieck ring* $K_0(\mathcal{V}_k)$ of reduced k -varieties: this is the abelian group generated by the isomorphism classes of such varieties, subject to the relations

$[X - Y] = [X] - [Y]$, where Y is a closed in X . The product over k turns it into a ring. Note that if we restrict ourselves to smooth varieties we get the same ring: the reason is that every k -variety X admits a stratification (i.e., a filtration by closed subschemes $X = X^0 \supset X^1 \supset \dots \supset X^{d+1} = \emptyset$ such that $X^k - X^{k+1}$ is smooth) and that any two such admit a common refinement. The latter property implies that $[X] := \sum_k [X^k - X^{k+1}]$ is unambiguously defined. In fact, $K_0(\mathcal{V}_k)$ is generated by the classes of complete nonsingular varieties, for any smooth variety U admits a completion \bar{U} by adding a normal crossing divisor and then $[U] = \sum (-1)^i [\bar{U}^i]$, where \bar{U}^i stands for the normalization of the codimension i skeleton of the resulting stratification. Włodarczyk's weak factorization theorem (in the form of the main theorem of [1]) can be used to show that relations of the following simple type suffice: if X is smooth projective and $\tilde{X} \rightarrow X$ is obtained by blowing up a smooth closed subvariety $Y \subset X$ with exceptional divisor \tilde{Y} , then $[\tilde{X}] - [\tilde{Y}] = [X] - [Y]$.

We denote the class of the affine line \mathbb{A}^1 by \mathbb{L} and we write M_k for the localization $K_0(\mathcal{V}_k)[\mathbb{L}^{-1}]$. Recall that a subset of a variety X is called *constructible* if it is a finite union of (locally closed) subvarieties. Any constructible subset C of X defines an element $[C] \in M_k$. The constructible subsets of X form a Boolean algebra and so we obtain in a tautological manner a M_k -valued measure μ_X defined on this Boolean algebra. More generally, a morphism $f : Y \rightarrow X$ defines on that same algebra an M_k -valued measure $f_*\mu_Y$: assign to a constructible subset of X its preimage in Y .

The ring M_k is interesting, big, and hard to grasp. Fortunately, there are several characteristics of M_k (i.e., ring homomorphisms from M_k to a ring) that are well understood. We describe some of these in decreasing order of complexity under the assumption that k is a subfield of \mathbb{C} . The first example is the Grothendieck ring $K_0(\text{HS})$ of the category of Hodge structures. A *Hodge structure* consists of a finite dimensional \mathbb{Q} -vector space H , a finite bigrading $H \otimes \mathbb{C} = \bigoplus_{p,q \in \mathbb{Z}} H^{p,q}$ such that $H^{p,q}$ is the complex conjugate of $H^{q,p}$ and each *weight summand*, $\bigoplus_{p+q=m} H^{p,q}$, is defined over \mathbb{Q} . There are evident notions of tensor product and morphism of Hodge structures so that we get an abelian category HS with tensor product. The Grothendieck construction produces a group $K_0(\text{HS})$, elements of which are representable as a formal difference of Hodge structures $[H] - [H']$ and $[H] = [H']$ if and only if H and H' are isomorphic. The tensor product makes it a ring.

For every complex variety X , the cohomology with compact supports, $H_c^r(X; \mathbb{Q})$, comes with a natural finite increasing filtration $W_\bullet H_c^r(X; \mathbb{Q})$, the *weight filtration*, such that the associated graded $\text{Gr}_\bullet^W H_c^r(X; \mathbb{Q})$ underlies a Hodge structure having $\text{Gr}_m^W H_c^r(X; \mathbb{Q})$ as weight m summand. We assign to X the *Hodge characteristic*¹

$$\chi_h(X) := \sum_r (-1)^r [H_c^r(X; \mathbb{Q})] \in K_0(\text{HS})$$

¹As all our characteristics are compactly supported we omit the otherwise desirable subscript c from the notation.

If $Y \subset X$ is closed subvariety, then the exact sequence

$$\cdots \rightarrow H_c^r(X - Y) \rightarrow H_c^r(X) \rightarrow H_c^r(Y) \rightarrow H_c^{r+1}(X - Y) \rightarrow \cdots$$

is compatible in a strong sense with the Hodge data. This implies the additivity property $\chi_h(X) = \chi_h(X - Y) + \chi_h(Y)$. For the affine line \mathbb{A}^1 , $H_c^r(\mathbb{A}^1; \mathbb{Q})$ is nonzero only for $r = 2$; the cohomology group $H_c^2(\mathbb{A}^1; \mathbb{Q})$ is one-dimensional and of type $(1, 1)$. So $\chi_h(\mathbb{A}^1)$ (usually denoted as $\mathbb{Q}(-1)$) is invertible. It follows that χ_h factorizes over M_k . If we only care for dimensions, then we compose with the ring homomorphism $K_0(\text{HS}) \rightarrow \mathbb{Z}[u, u^{-1}, v, v^{-1}]$, $[H] \mapsto \sum_{p,q} \dim(H^{p,q})u^p v^q$, to get the *Hodge number characteristic* $\chi_{\text{hn}} : M_k \rightarrow \mathbb{Z}[u, u^{-1}, v, v^{-1}]$. It takes \mathbb{L} to uv . The *weight characteristic* $\chi_{\text{wt}} : M_k \rightarrow \mathbb{Z}[w, w^{-1}]$ is obtained if we go further down along the map $\mathbb{Z}[u, u^{-1}, v, v^{-1}] \rightarrow \mathbb{Z}[w, w^{-1}]$ that sends both u and v to w . Evaluating the latter at $w = 1$ gives the ordinary² Euler characteristic $\chi_{\text{top}} : M_k \rightarrow \mathbb{Z}$.

In the spirit of this discussion is the following question raised by Kapranov [21]:

QUESTION 2.1. — Let X be a variety over k . If $\sigma_n(X) \in M_k$ denotes the class of its n th symmetric power, is then

$$Z_X(T) := 1 + \sum_{n=1}^{\infty} \sigma_n(X) T^n \in M_k[[T]]$$

a rational function in the sense that it determines an element in a suitable localization of $M_k[[T]]$? (Since the logarithmic derivative Z'/Z defines an additive map $M_k \rightarrow M_k[[T]]$, we may restrict ourselves here to the case of a smooth variety.) Does it satisfy a functional equation when X is smooth and complete? Kapranov shows that the answer to both questions is yes in case $\dim(X) \leq 1$.

A measure on the space of sections

Let us call a \mathbb{D} -variety a separated reduced scheme that is flat and of finite type over \mathbb{D} and whose closed fiber is reduced. Given a \mathbb{D} -variety \mathcal{X}/\mathbb{D} with closed fiber X , then the set of its sections up to order n , \mathcal{X}_n , is the set of closed points of a k -variety (also denoted \mathcal{X}_n) naturally associated to \mathcal{X} . It is obtained from \mathcal{X} modulo \mathfrak{m}^{n+1} essentially by Weil restriction of scalars [19]. So $\mathcal{X}_0 = X$. The set \mathcal{X}_∞ of sections of $\mathcal{X} \rightarrow \mathbb{D}$ is the projective limit of these and is therefore the set of closed points of a provariety. If \mathcal{X}/\mathbb{D} is of the form $X \times \mathbb{D} \rightarrow \mathbb{D}$, with X a k -variety, then we are dealing with the space of n -jets (of curves) on X and the *arc space* of X , here denoted by $\mathcal{L}_n(X)$ resp. $\mathcal{L}(X)$.

For $m \geq n$ we have a forgetful morphism $\pi_n^m : \mathcal{X}_m \rightarrow \mathcal{X}_n$. (When $n = 0$, we shall often write π_X^m, π_X instead of π_0^m, π_0 .) A fiber of π_n^{n+1} lies in an affine space over the Zariski tangent space of the base point. In case X is smooth, it is in fact an affine space over the

²A complex algebraic variety can be compactified within its homotopy type by giving it a topological boundary that is stratifiable into strata of odd dimension. This boundary has zero Euler characteristic, hence the compactly supported Euler characteristic of the variety is its ordinary Euler characteristic.

tangent space of the base point: π_n^{n+1} has then the structure of a torsor over the tangent bundle. A theorem of Greenberg [20] asserts that there exists a constant c such that the image of π_n equals the image of π_n^{cn} . So $\pi_n(\mathcal{X}_\infty)$ is constructible.

The goal is to define a measure on an interesting algebra of subsets of \mathcal{X}_∞ in such a way that its direct image under π_X is the tautological measure μ_X when X is smooth. (This will lead us to deviate from the definition of Denef-Loeser and Batyrev by a factor \mathbb{L}^d and to adopt the one used in [25] instead.) For this we assume that \mathcal{X} is of pure relative dimension d and we say that a subset A of \mathcal{X}_∞ is *stable* if for some $n \in \mathbb{N}$ we have

- $\pi_n(A)$ is constructible in \mathcal{X}_n and $A = \pi_n^{-1}\pi_n(A)$,
- for all $m \geq n$ the projection $\pi_{m+1}(A) \rightarrow \pi_m(A)$ is a piecewise trivial fibration (that is, trivial relative to a decomposition into subvarieties) with fiber an affine space of dimension d .

The second condition is of course superfluous in case \mathcal{X}/\mathbb{D} is smooth. It is clear that $\dim \pi_m(A) - md$ is independent of the choice of $m \geq n$; we call this the (*virtual*) *dimension* $\dim A$ of A . The same is true for the class $[\pi_m(A)]\mathbb{L}^{-md} \in M_k$; we denote that class by $\tilde{\mu}_{\mathcal{X}}(A)$. The stable subsets of \mathcal{X} form a Boolean algebra on which $\tilde{\mu}_{\mathcal{X}}$ defines a finite additive measure. A theorem of Denef-Loeser (see Theorem 9.1) ensures that there are plenty of stable sets.

In order to extend the measure to a bigger collection of interesting subsets of \mathcal{X}_∞ we need to complete M_k . Given $m \in \mathbb{Z}$, let $F_m M_k$ be the subgroup of M_k spanned by the $[Z]\mathbb{L}^{-r}$ with $\dim Z \leq m+r$. This is a filtration of M_k as a ring: $F_m M_k \cdot F_n M_k \subset F_{m+n} M_k$. So the separated completion of M_k with respect to this filtration,

$$\hat{M}_k := \varprojlim_{\leftarrow} M_k / F_m M_k \quad (m \rightarrow -\infty \text{ in this limit}),$$

to which we will refer as the *dimensional completion*, is also a ring. The kernel of the natural map $M_k \rightarrow \hat{M}_k$ is $\cap_m F_m M_k$, of course. It is not known whether this is zero³. In case $k \subset \mathbb{C}$, the Hodge characteristic extends to this completion:

$$\chi_h : \hat{M}_k \rightarrow \hat{K}_0(\text{HS}).$$

Here $\hat{K}_0(\text{HS})$ is defined in a similar way as \hat{M}_k with ‘dimension’ replaced by ‘weight’. The assertion follows from the fact that the weights in the compactly supported cohomology of a variety of dimension d are $\leq 2d$. Likewise we can extend the characteristics counting Hodge numbers or weight numbers (with values Laurent power series in the reciprocals of their variables). This does not apply to the Euler characteristic, but in many cases of interest the weight characteristic gives a rational function in w that has no pole at $w = 1$. Its value there is then a good substitute.

³This issue is avoided if we work with the adic completion $\mathbb{Z}((L^{-1})) \otimes_{\mathbb{Z}[L]} K_0(\mathcal{V}_k)$ instead, but in practice this is too small. Nevertheless, it seems that in all applications we are dealing with elements lying in the localization $\mathbb{Q}(L) \otimes_{\mathbb{Z}[L]} K_0(\mathcal{V}_k)$.

We will be mostly concerned with the composite of $\tilde{\mu}_{\mathcal{X}}$ and the completion map, for it is this measure that we shall extend. We call this the *motivic measure* on \mathcal{X} and denote it by $\mu_{\mathcal{X}}$. Let us say that a subset $A \subset \mathcal{X}_{\infty}$ is *measurable* if for every (negative) integer m there exist a stable subset $A_m \subset \mathcal{X}_{\infty}$ and a sequence $(C_i \subset \mathcal{X}_{\infty})_{i=0}^{\infty}$ of stable subsets such that the symmetric difference $A \Delta A_m$ is contained in $\cup_{i \in \mathbb{N}} C_i$ with $\dim C_i < m$ for all i and $\dim C_i \rightarrow -\infty$, for $i \rightarrow \infty$.

PROPOSITION 2.2. — The measurable subsets of \mathcal{X}_{∞} make up a Boolean subalgebra and $\mu_{\mathcal{X}}$ extends as a measure to this algebra by

$$\mu_{\mathcal{X}}(A) := \lim_{m \rightarrow -\infty} \mu_{\mathcal{X}}(A_m).$$

In particular, the above limit exists in \hat{M}_k and its value only depends on A .

The proof is based on

LEMMA 2.3. — Let \mathcal{X}/\mathbb{D} be of pure dimension and $A \subset \mathcal{X}_{\infty}$ a stable subset. If $\mathcal{C} = \{C_i\}_{i=1}^{\infty}$ is a countable covering of A by stable subsets with $\dim C_i \rightarrow -\infty$ as $i \rightarrow \infty$, then A is covered by a finite subcollection of \mathcal{C} .

Proof. — Let $n \in \mathbb{N}$ be such that $A = \pi_n^{-1}\pi_n(A)$. Suppose that A is not covered by a finite subcollection of \mathcal{C} . Choose $k \in \mathbb{N}$ such that $\dim C_i < -(n+2)d$ for $i > k$ and let $u_{n+1} \in \pi_{n+1}(A \setminus \cup_{i \leq k} C_i)$. We have $\pi_{n+1}^{-1}u_{n+1} \subset A$. This set is not covered by a finite subcollection of \mathcal{C} , for clearly $\pi_{n+1}^{-1}(u_{n+1})$ is not covered by $\{C_i\}_{i \leq k}$ and for $i > k$, $C_i \cap \pi_{n+1}^{-1}(u)$ is of positive codimension in $\pi_{n+1}^{-1}(u)$.

With induction we find a sequence $\{u_m \in \mathcal{L}_m(X)\}_{m > n}$ so that for all $m > n$ u_{m+1} lies over u_m and $\pi^{-1}(u_m)$ is not covered by a finite subcollection of \mathcal{C} . The sequence defines an element $u \in \mathcal{X}$. Since $\pi_n(u) \in \pi_n(A)$, we have $u \in A$ and so $u \in C_i$ for some i . But if C_i is stable at level $m > n$, then $\pi_m^{-1}(u_m) \subset C_i$, which contradicts a defining property of u_m . \square

For $k = \mathbb{C}$, the condition $\lim_{i \rightarrow \infty} \dim C_i = -\infty$ is unnecessary, for we may then use the Baire property of \mathbb{C} instead [5].

Proof of 2.2. — Suppose we have another solution $A \Delta A'_m \subset \cup_{i \in \mathbb{N}} C'_i$ with A'_m and C'_i stable, $\dim(C'_i) < m$ for all i and $\dim C'_i \rightarrow -\infty$ as $i \rightarrow \infty$. It is enough to prove that the dimension of the stable set $A_m \Delta A'_m$ is $< m$. Since $A_m \Delta A'_m \subset \cup_{i \in \mathbb{N}} (C_i \cup C'_i)$, Lemma 2.3 applies and we find that $A_m \Delta A'_m \subset \cup_{i \leq N} (C_i \cup C'_i)$ for some N . Since every term has dimension $< m$, this is also true for $A_m \Delta A'_m$. \square

So a countable union of stable sets $A = \cup_{n \in \mathbb{N}} A_n$ with $\lim_{n \rightarrow \infty} \dim A_n = -\infty$ is measurable and $\mu_{\mathcal{X}}(A) = \lim_{n \rightarrow \infty} \mu_{\mathcal{X}}(\cup_{k \leq n} A_k)$.

REMARK 2.4. — Given a \mathbb{D} -variety \mathcal{X} , then for any $d \in \mathbb{N}$ there is a d -measure $\mu_{\mathcal{X}}^d$ that induces $\mu_{\mathcal{Y}}$ on \mathcal{Y}_{∞} for any \mathbb{D} -subvariety \mathcal{Y} of pure dimension d . We expect this

measure to extend to a much bigger collection of subsets of \mathcal{X} so that if $f : \mathcal{X} \rightarrow \mathcal{S}$ is a dominant \mathbb{D} -morphism of pure relative dimension d , then every fiber of $f_* : \mathcal{X}_\infty \rightarrow \mathcal{S}_\infty$ is $\mu_{\mathcal{X}}^d$ -measurable.

Here is a sample of the results of Denef and Loeser on the rationality of Poincaré series [14].

THEOREM 2.5. — Let X be a k -variety. Then $\sum_{n=0}^\infty \mu_{\mathcal{X}}(\pi_n \mathcal{L}(X)) T^n \in M_k[[T]]$ is a rational expression in T with each factor in the denominator of the form $1 - \mathbb{L}^a T^b$ where $a \in \mathbb{Z}$ and b is a positive integer.

We will not discuss its proof, since this theorem is not used in what follows. Denef and Loeser derive this by means of Kontsevich’s transformation rule discussed below, which is applied to a suitable projective resolution \mathcal{X} , and a theorem about semialgebraic sets, due to Pas [24]. It is likely that this theorem still holds for the space of sections of any \mathbb{D} -variety.

3. THE TRANSFORMATION RULE [22], [14], [16]

We describe two results that are at the basis of the theory. The proofs are relegated to Section 9.

PROPOSITION 3.1. — For a \mathbb{D} -variety \mathcal{X}/\mathbb{D} of pure dimension, the preimage of any constructible subset under $\pi_n : \mathcal{X}_\infty \rightarrow \mathcal{X}_n$ is measurable. In particular, \mathcal{X}_∞ is measurable. If $\mathcal{Y} \subset \mathcal{X}$ is nowhere dense, then \mathcal{Y}_∞ is of measure zero.

For \mathcal{X}/\mathbb{D} of pure relative dimension we have the notion of an integrable function $\Phi : \mathcal{X}_\infty \rightarrow \hat{M}_k$: this requires the fibers of Φ to be measurable and the sum $\sum_a \mu_{\mathcal{X}}(\Phi^{-1}(a))a$ to converge, i.e., there are at most countably many nonzero terms $(\mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i)_{i \in \mathbb{N}}$ and we have $\mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i \in F_{m_i} \hat{M}_k$ with $\lim_{i \rightarrow \infty} m_i = -\infty$. The motivic integral of Φ is then by definition the value of this series:

$$\int \Phi d\mu_{\mathcal{X}} = \sum_i \mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i.$$

We have a similar notion for maps with values in topological \hat{M}_k -modules. An important example arises from an ideal $\mathcal{I} \subset \mathcal{O}_{\mathcal{X}}$: such an ideal defines a function $\text{ord}_{\mathcal{I}} : \mathcal{X}_\infty \rightarrow \mathbb{N} \cup \{\infty\}$ by assigning to $\gamma \in \mathcal{X}_\infty$ the multiplicity of $\gamma^* \mathcal{I}$. The condition $\text{ord}_{\mathcal{I}} \gamma = n$ only depends on the n -jet of γ and this defines a constructible subset $C_n \subset \mathcal{X}_n$. Hence the fibers of $\text{ord}_{\mathcal{I}}$ are measurable. We shall see that the function

$$\mathbb{L}^{-\text{ord}_{\mathcal{I}}} : \mathcal{X}_\infty \rightarrow \hat{M}_k$$

is integrable.

There is a beautiful transformation rule for motivic integrals under modifications. Let $H : \mathcal{Y} \rightarrow \mathcal{X}$ be a morphism of \mathbb{D} -varieties of pure dimension d . We define the *Jacobian ideal* $\mathcal{J}_H \subset \mathcal{O}_{\mathcal{Y}}$ of H as 0th Fitting ideal of $\Omega_{\mathcal{Y}/\mathcal{X}}$. This has the nice property that its formation commutes with base change. The following theorem generalizes an unpublished theorem of Kontsevich.

THEOREM 3.2. — Let $H : \mathcal{Y} \rightarrow \mathcal{X}$ be a \mathbb{D} -morphism of pure dimensional \mathbb{D} -varieties with \mathcal{Y}/\mathbb{D} smooth. If A is a measurable subset of \mathcal{Y}_{∞} with $H|_A$ injective, then HA is measurable and $\mu_{\mathcal{X}}(HA) = \int_A \mathbb{L}^{-\text{ord } \mathcal{J}_H} d\mu_{\mathcal{Y}}$.

4. THE BASIC FORMULA [14]

A relative Grothendieck ring

It is convenient to be able to work in a relative setting. Given a variety S , denote by $K_0(\mathcal{V}_S)$ the Grothendieck ring of S -varieties and by M_S its localization with respect \mathbb{L} . The ring M_S can be dimensionally completed as usual. Notice that an element of M_S defines a M_k -valued measure on the Boolean algebra of constructible subsets of S . Often measures are naturally represented this way. For instance, the preceding shows that for a measurable subset A of \mathcal{X}_{∞} , the direct image of $\mu_{\mathcal{X}}|_A$ on X is given by an element of \hat{M}_X .

A morphism $f : S' \rightarrow S$ induces a ring homomorphism $f^* : M_S \rightarrow M_{S'}$. This makes $M_{S'}$ a M_S -module. We also have a direct image $f_* : M_{S'} \rightarrow M_S$ that is a homomorphism of M_S -modules. Notice that f itself defines an element $[f] \in M_S$; this is also the image of $1 \in M_{S'}$ under f_* .

There are corresponding characteristics. For instance, the ordinary Euler characteristic χ_{top} becomes a ring homomorphism from M_S to the Grothendieck ring of constructible \mathbb{Q} -vector spaces on S . This ring is generated by direct images of irreducible local systems of \mathbb{Q} -vector spaces over smooth irreducible subvarieties Z of S . (A better choice is to take the intersection cohomology sheaf in S of this local system along Z ; this has the advantage that it only depends on the generic point of Z .)

Similarly, the Hodge characteristic χ_{h} takes values in a ring $K_0(\text{HS}_S)$ that is generated by variations of Hodge structures over a smooth subvariety of S . The homomorphisms f^* and f_* persist on this level: $f : S' \rightarrow S$ induces homomorphisms $f^* : K_0(\text{HS}_S) \rightarrow K_0(\text{HS}_{S'})$ and $f_* : K_0(\text{HS}_{S'}) \rightarrow K_0(\text{HS}_S)$.

The basic computation

A case of interest is when the base variety is $(\mathbb{N} \times \mathbb{G}_m)^r$. This fails to be finite type, but that is of no consequence and we identify $\hat{M}_{(\mathbb{N} \times \mathbb{G}_m)^r}$ with $\hat{M}_{\mathbb{G}_m^r}[[T_1, \dots, T_r]]$ in the obvious way.

We use a uniformizing parameter of \mathcal{O} to define

$$\text{ac} : \mathcal{L}(\mathbb{A}^1) - \{0\} \rightarrow \mathbb{N} \times \mathbb{G}_m,$$

by assigning to γ its order $\text{ord}(\gamma)$ resp. the first nonzero coefficient of γ (ac stands for *angular component*). Integration along ac sends a \hat{M}_k -valued measure on $\mathcal{L}(\mathbb{A}^1)$ to an element of $\hat{M}_{\mathbb{G}_m}[[T]]$. The prime example is when this measure is given by a regular function $f : \mathcal{X} \rightarrow \mathbb{A}^1$ on a \mathbb{D} -variety \mathcal{X} of pure relative dimension: this induces a map $f_* : \mathcal{X}_\infty \rightarrow \mathcal{L}(\mathbb{A}^1)$ and we then define

$$\text{ac}_f : \mathcal{X}_\infty \xrightarrow{f_*} \mathcal{L}(\mathbb{A}^1) \xrightarrow{\text{ac}} \mathbb{N} \times \mathbb{G}_m,$$

so that $[\text{ac}_f] \in \hat{M}_{\mathbb{G}_m}[[T]]$. More generally, given a morphism $f = (f_1, \dots, f_r) : \mathcal{X} \rightarrow \mathbb{A}^r$, we abbreviate

$$\text{ac}_{X,f} := (\pi_X, \text{ac}_{f_1}, \dots, \text{ac}_{f_r}) : \mathcal{X}_\infty \rightarrow X \times (\mathbb{N} \times \mathbb{G}_m)^r.$$

So $[\text{ac}_{X,f}] \in \hat{M}_{X \times \mathbb{G}_m^r}[[T_1, \dots, T_r]]$.

CONVENTIONS 4.1. — If E is a simple normal crossing hypersurface on a smooth k -variety Y , then we adhere to the following notation throughout the talk: $(E_i)_{i \in \text{irr}(E)}$ denotes the collection of irreducible components of E (so these are all smooth by assumption) and for any subset $I \subset \text{irr}(E)$, E_I° stands for the locus of $p \in \tilde{X}$ with $p \in E_i$ if and only if $i \in I$. (With this convention, $E_\emptyset^\circ = Y - E$.) We denote the complement of the zero section of the normal bundle of E_i by U_{E_i} (so this is a \mathbb{G}_m -bundle over E_i) and U_I designates the fiber product of the bundles $U_{E_i}|_{E_I^\circ}$, $i \in I$ (a \mathbb{G}_m^I -bundle whose total space has the same dimension as Y).

If \mathcal{E} is a simple normal crossing hypersurface on a \mathbb{D} -variety \mathcal{Y}/\mathbb{D} with \mathcal{Y} smooth, then we shall always assume that its union with the closed fiber Y has also normal crossings. The notational conventions are as above to the extent that restriction or intersection with Y is indicated by switching from calligraphic to roman font (e.g., $E_i = \mathcal{E}_i \cap Y$). If Y is smooth, then we may identify $\text{irr}(E)$ with a subset of $\text{irr}(\mathcal{E})$. (An equality if \mathcal{E} has no component in Y .)

The following proposition accounts for many of the rationality assertions in [14].

PROPOSITION 4.2. — Let \mathcal{X}/\mathbb{D} be a \mathbb{D} -variety of pure relative dimension and $H : \mathcal{Y} \rightarrow \mathcal{X}$ a resolution of singularities. Let \mathcal{E} be a simple normal crossing hypersurface on \mathcal{Y} that has no irreducible component in the closed fiber Y . Assume that the Jacobian ideal \mathcal{J}_H of H is principal and has divisor $\sum_i (\nu_i - 1)\mathcal{E}_i$ (so $\nu_i \geq 1$). Let for $\rho = 1, \dots, r$, $f_\rho : \mathcal{X} \rightarrow \mathbb{A}^1$ be a regular function such that $f_\rho H$ has zero divisor $\sum_i N_{i,\rho}\mathcal{E}_i$ and put $N_i := (N_{i,1}, \dots, N_{i,r}) \in \mathbb{N}^r$, $i \in \text{irr}(E)$. Then

$$[\text{ac}_{X,f}] = \sum_{I \subset \text{irr}(E)} [U_I/X \times \mathbb{G}_m^r] \prod_{i \in I} (\mathbb{L}^{\nu_i} T^{-N_i} - 1)^{-1} \text{ in } \hat{M}_{X \times \mathbb{G}_m^r}[[T_1, \dots, T_r]],$$

where $U_I \rightarrow X \times \mathbb{G}_m^r$ has first component projection onto $E_I^\circ \subset X$ followed by the restriction of H and second component induced by fH .

Proof. — Given $m \in \mathbb{N}^{\text{irr}(E)}$, consider the set $\mathcal{Y}(m)$ of $\gamma \in \mathcal{Y}_\infty$ with order m_i along \mathcal{E}_i . So for $\gamma \in \mathcal{Y}(m)$ we have $\text{ord}_{\mathcal{J}_H}(\gamma) = \sum_i m_i(\nu_i - 1)$ and $\text{ord}_{f_\rho H}(\gamma) = \sum_i m_i N_{\rho,i}$. If $\text{supp}(m) \subset \text{irr}(E)$ is the support of m , then we have a natural projection $e_m : \mathcal{Y}(m) \rightarrow U_{\text{supp}(m)}$. Its composite with the morphism $U_{\text{supp}(m)} \rightarrow X \times \mathbb{G}_m^r$ is a restriction of $\text{ac}_{X,fH} := (\pi_X H, \text{ac}_{f_1 H}, \dots, \text{ac}_{f_r H}) : \mathcal{Y}_\infty \rightarrow X \times (\mathbb{N} \times \mathbb{G}_m)^r$ with \mathbb{N}^r -component $\sum_i m_i N_i$. In other words,

$$[\text{ac}_{X,fH} |_{\mathcal{Y}(m)}] = [U_{\text{supp}(m)}/X \times \mathbb{G}_m^r] \mathbb{L}^{-\sum_i m_i} T^{\sum_i m_i N_i}.$$

So the transformation formula 3.2 yields

$$\begin{aligned} [\text{ac}_{X,f}] &= \sum_{m \in \mathbb{N}^{\text{irr}(E)}} [U_{\text{supp}(m)}/X \times \mathbb{G}_m^r] \prod_{i \in \text{supp}(m)} \left(\mathbb{L}^{-m_i - m_i(\nu_i - 1)} T^{m_i N_i} \right) \\ &= \sum_{I \subset \text{irr}(E)} [U_I/X \times \mathbb{G}_m^r] \prod_{i \in I} \frac{\mathbb{L}^{-\nu_i} T^{N_i}}{1 - \mathbb{L}^{-\nu_i} T^{N_i}}. \end{aligned}$$

□

If we drop the assumption that \mathcal{E} has no irreducible component in Y , then the above formula must be somewhat modified: now each irreducible component of Y contributes with an expression of the above form times a monomial in \mathbb{L}^{-1} and T_1, \dots, T_r .

COROLLARY 4.3. — In the situation of 4.2, the class of $(\pi_X, \text{ord}_f) : \mathcal{X}_\infty \rightarrow X \times \mathbb{N}^r$ in $\hat{M}_X[[T_1, \dots, T_r]]$ equals

$$\sum_{I \subset \text{irr}(E)} [E_I^\circ/X] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i} T^{-N_i} - 1}.$$

In particular, the direct image of $\mu_{\mathcal{X}}$ on X is represented by

$$\sum_{I \subset \text{irr}(E)} [E_I^\circ/X] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i} - 1}.$$

Proof. — Since U_I is a \mathbb{G}_m^I -bundle over E_I° , the class of the projection $U_I \rightarrow X$ is $(\mathbb{L} - 1)^{|I|}$ times the class of $E_I^\circ \rightarrow X$. □

This corollary also shows that the Hodge character of the full measure of \mathcal{X}_∞ lies in $\mathbb{Q}[u, v][[uv^{-1}]]$ and that the ‘Euler characteristic’ of \mathcal{X}_∞ is $\sum_{I \subset \text{irr}(E)} \chi_{\text{top}}(E_I^\circ) \prod_{i \in I} \nu_i^{-1}$.

5. THE MOTIVIC NEARBY FIBER [13], [18]

An equivariant Grothendieck ring

Let G be a finite group. We consider varieties X with *good* G -action, where ‘good’ means that every orbit is contained in an affine open subset. This property ensures that

the orbit space exists as a variety. For instance, a representation of G on a k -vector space V is good. For a fixed variety S with trivial G -action, we define the Grothendieck group $K_0^G(\mathcal{V}_S)$ as generated by isomorphism types of S -varieties with good G -action modulo the usual equivalence relation (defined by pairs) and the relation that declares that every finite dimensional representation ρ of G has the same class as the trivial representation of the same degree (i.e., $\mathbb{L}^{\deg(\rho)}$). The product makes $K_0^G(\mathcal{V}_S)$ a $K_0(\mathcal{V}_S)$ -algebra. We may change our point of view by looking up instead of down: $K_0^G(\mathcal{V}_S)$ is then thought of as a Grothendieck ring of S -varieties equipped with a G -covering (the G -action on the cover is automatically good).

If G is abelian, then assigning to a variety X with good G -action its G -orbit space $\overline{X} := G \backslash X$ augments this as a $K_0(\mathcal{V}_S)$ -module:

$$K_0^G(\mathcal{V}_S) \rightarrow K_0(\mathcal{V}_S), \quad a \mapsto \bar{a}.$$

(Not as an algebra, for the orbit space of a product is in general not the product of orbit spaces.) That this is well-defined follows from the lemma below. (We do not know whether this holds for arbitrary finite G .)

LEMMA 5.1. — Let be given a representation of a finite abelian group G on a k -vector space V of finite dimension n . Then the class of \overline{V} in $K_0(\mathcal{V}_k)$ is \mathbb{L}^n .

Proof. — Let $V = \bigoplus_{\chi \in \hat{G}} V_\chi$ be the eigen space decomposition of the G -action. Given a subset $I \subset \hat{G}$, denote by V_I the set of vectors in V whose V_χ -component is nonzero if and only if $\chi \in I$. We have a natural projection $\overline{V_I} \rightarrow \prod_{\chi \in I} \mathbb{P}(V_\chi)$. This has the structure of a torus bundle, the torus in question being a quotient of \mathbb{G}_m^I by a finite subgroup. So the class of $\overline{V_I}$ in M_k is $(\mathbb{L} - 1)^{|I|}$ times the class of $\prod_{\chi \in I} \mathbb{P}(V_\chi)$. Since V_I has also that structure, the classes of $\overline{V_I}$ and V_I in M_k coincide. Hence the same is true for \overline{V} and V . \square

Similarly we can form the M_S -algebra $M_S^G := K_0^G(\mathcal{V}_S)[\mathbb{L}^{-1}]$ and its dimensional completion. If G is abelian we have corresponding augmentations taking values in M_S and its completion. If G acts on an S -variety Z , then we denote by $[Z/S; G]$ the corresponding element of M_S^G or its image in \hat{M}_S^G .

There are corresponding characteristics in case $k \subset \mathbb{C}$. For instance, the ordinary Euler characteristic defines a ring homomorphism from M_k^G to the Grothendieck ring $K_0^G(\mathbb{Q})$ of finite dimensional representations of G over \mathbb{Q} and more generally, we have a ring homomorphism χ_{top}^G from M_S^G to the Grothendieck ring of constructible sheaves with G -action on S , $K_0^G(\mathbb{Q}_S)$. Similarly, there is a Hodge character $\chi_h^G : M_S^G \rightarrow K_0^G(\text{HS}_S)$.

The case $G = \hat{\mu}$

We will mostly (but not exclusively) be concerned with the case when G is a group of roots of unity. We have the Grothendieck ring $M_S^{\hat{\mu}}$ of varieties with a topological action

of the procyclic group $\hat{\mu} = \lim_{\leftarrow} \mu_n$ (such an action factorizes through a finite quotient μ_n). The inverse automorphism of $\hat{\mu}$, $\zeta \mapsto \zeta^{-1}$, defines an involution $*$ in $M_S^\hat{\mu}$.

The group of continuous characters of $\hat{\mu}$ is naturally isomorphic with \mathbb{Q}/\mathbb{Z} , with the involution $*$ acting as multiplication by -1 ; the projection $\hat{\mu} \rightarrow \mu_n$ followed by the inclusion $\mu_n \subset \mathbb{G}_m$ corresponds to $\frac{1}{n} \pmod{\mathbb{Z}}$. In other words, $K_0^{\hat{\mu}}(\mathbb{C}) \cong \mathbb{Z}[e^\alpha \mid \alpha \in \mathbb{Q}/\mathbb{Z}]$. For every positive integer n there is a rational irreducible representation χ_n of μ_n , namely the field $\mathbb{Q}(\mu_n)$, regarded as \mathbb{Q} -vector space. These make up an additive basis of $K_0^{\hat{\mu}}(\mathbb{Q})$. The image of χ_n in $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ is $\sum_{(k,n)=1} e^{k/n}$, which allows us to regard $K_0^{\hat{\mu}}(\mathbb{Q})$ as a subring of $\mathbb{Z}[e^\alpha \mid \alpha \in \mathbb{Q}/\mathbb{Z}]$.

Equivariance

If X is a k -variety, then $\text{Aut}(\mathbb{D})$ acts on $\mathcal{L}(X)$ by composition: $h(\gamma) := \gamma h^{-1}$. If X is of pure dimension, then this action is free outside negligible subset. Clearly, a morphism of k -varieties induces an $\text{Aut}(\mathbb{D})$ -equivariant map between their arc spaces. This is of particular interest when the target is the affine line: Suppose we are given a regular function $f : X \rightarrow \mathbb{A}^1$. Then the image of the resulting map $\mathcal{L}(f) : \mathcal{L}(X) \rightarrow \mathcal{L}(\mathbb{A}^1)$ intersected with $\mathcal{L}(\mathbb{A}^1, 0)$ will be a union of $\text{Aut}(\mathbb{D})$ -orbits. Here $\mathcal{L}(\mathbb{A}^1, 0)$ stands for the set of arcs based at the origin $0 \in \mathbb{A}^1$. If we assign to $\gamma \in \mathcal{L}(\mathbb{A}^1, 0)$ the pull-back under γ of the defining parameter of \mathbb{A}^1 , then we obtain a natural bijection $\mathcal{L}(\mathbb{A}^1, 0) \cong \mathfrak{m}$. Under this bijection an $\text{Aut}(\mathbb{D})$ -orbit in $\mathcal{L}(\mathbb{A}^1, 0) - \{0\}$ corresponds to $\mathfrak{m}^n - \mathfrak{m}^{n+1}$ for some positive integer n . Now $\mathcal{L}(f)$ is an $\text{Aut}(\mathbb{D})$ -fibration over this orbit (if nonempty) and hence the direct image of $\mu_{\mathcal{L}(X)}$ (and more generally, of any $\text{Aut}(\mathbb{D})$ -invariant measure on $\mathcal{L}(\mathbb{A}^1, 0)$) restricted to that orbit is completely understood if we know its value on $t^n + \mathfrak{m}^{n+1}$ equipped with the monodromy action of $\mu_n \subset \text{Aut}(\mathbb{D})$. Here t is a uniformizing parameter for \mathcal{O} . So if X_0 is the zero set of the morphism $X \rightarrow \mathbb{A}^1$, then we get an element of $\hat{M}_{X_0}^{\mu_n}$ (or in $M_{X_0}^{\mu_n}$ in case X is smooth) for every $n \geq 1$.

A motivic zeta function

We now take X smooth of dimension d and let $f : X \rightarrow \mathbb{A}^1$ be a nonconstant regular function. Given a positive integer n , then for $\gamma \in \mathcal{X}$, the property $\text{ord}_f(\gamma) = n$ only depends on its n -jet and hence $\text{ord}_f^{-1}(n)$ is the preimage of a subset $\mathcal{L}(X)_n^f$ of $\mathcal{L}(X)_n$. The action of $\mathbb{G}_m \subset \text{Aut}(\mathbb{D})$ on $\mathcal{L}(X)_n$ induces one on $\mathcal{L}(X)_n^f$ and the projection $\text{ac}_{f,n} : \text{ord}_f^{-1}(n) \rightarrow \mathbb{G}_m$ (a restriction of ac_f) factorizes over a morphism $\mathcal{L}(X)_n^f \rightarrow \mathbb{G}_m$ that is homogeneous of degree n with respect to the given \mathbb{G}_m -action. So the fiber over 1, $\mathcal{L}(X)_n^f(1)$ is preserved by $\mu_n \subset \mathbb{G}_m$ and this action describes the monodromy. It is clear that

$$[\text{ac}_{f,n}^{-1}(1)/X_0; \mu_n] = [\mathcal{L}(X)_n^f(1)/X_0; \mu_n] \mathbb{L}^{-dn}$$

holds in $M_{X_0}^{\mu_n}$. We form the generating series

$$S(f) := \sum_{n=1}^{\infty} [\text{ac}_{f,n}^{-1}(1)/X_0; \mu_n] T^n = \sum_{n=1}^{\infty} [\mathcal{L}(X)_{f,n}(1)/X_0; \mu_n] (\mathbb{L}^{-d}T)^n \in M_{X_0}^{\hat{\mu}}[[T]].$$

This series can be computed from an embedded resolution of the zero set of f , $H : Y \rightarrow X$ of X , as in 4.1. We assume here that the preimage E of $X_0 := f^{-1}(0)$ is a simple normal crossing hypersurface that contains the exceptional set. Let m be a positive integer that is divided by all the coefficients N_i of the divisor (f) on the irreducible components of E . If we make a base change of $\tilde{f} := fH$ over the m th power map $\mathbb{A}^1 \rightarrow \mathbb{A}^1$ and normalize, then we get a μ_m -covering $\tilde{Y} \rightarrow Y$. Let \tilde{E}_I° be a connected component of the preimage of E_I° in \tilde{Y} . The restriction $\tilde{E}_I^\circ \rightarrow E_I^\circ$ is unramified, and has μ_m -stabilizer of \tilde{E}_I° as its Galois group. The latter is easily seen to be the subgroup μ_{N_I} , where $N(I) := \text{gcd}\{N_i \mid i \in I\}$. This defines

$$[\tilde{E}_I^\circ/Y; \mu_{N(I)}] \in M_Y^{\mu_{N(I)}}.$$

This element lies over X_0 if I is nonempty, an assumption we make from now on. We wish to compare it with $U_I(1) \subset U_I$, the fiber over 1 of the projection $U_I \rightarrow \mathbb{G}_m$ induced by \tilde{f} . This projection has weights $(N_i)_{i \in I}$ relative to the \mathbb{G}_m^I -action and so $\prod_{i \in I} \mu_{N_i} \subset \mathbb{G}_m^I$ preserves $U_I(1)$. This finite group contains a monodromy action by $\mu_{N(I)}$: write $N(I) = \sum_{i \in I} \alpha_i N_i$ and embed \mathbb{G}_m in \mathbb{G}_m^I by $\lambda \rightarrow (\lambda^{\alpha_i})_{i \in I}$ (since the $(\alpha_i)_{i \in I}$ are relatively prime, this is an embedding indeed). Notice that the projection $U_I \rightarrow \mathbb{G}_m$ is homogeneous of degree N_I relative to the action of this one parameter subgroup. This implies that $\mu_{N(I)} \subset \mathbb{G}_m$ may serve as monodromy group. (There are a priori several choices for this action, but they are all E_I° -isomorphic.)

LEMMA 5.2. — In $M_{Y_0}^{\hat{\mu}}$ we have $[U_I(1)/Y_0; \mu_{N(I)}] = (\mathbb{L} - 1)^{|I|-1} [\tilde{E}_I^\circ/Y_0; \mu_{N(I)}]$.

Proof. — One verifies that the Stein factorization of the projection $U_I(1) \rightarrow E_I^\circ$ has $\tilde{E}_I^\circ \rightarrow E_I^\circ$ as finite factor with $U_I(1) \rightarrow \tilde{E}_I^\circ$ being an algebraic torus bundle of rank $|I| - 1$. In view of Lemma 5.1 the equivariant class of the latter is $(\mathbb{L} - 1)^{|I|-1}$ times the equivariant class of the base. The lemma follows. \square

Much of the work of Denef-Loeser on motivic integration centers around the following

THEOREM 5.3. — The following identity holds in $M_{X_0}^{\hat{\mu}}[[T]]$:

$$S(f) = \sum_{\emptyset \neq I \subset \text{irr}(E)} (\mathbb{L} - 1)^{|I|-1} [\tilde{E}_I^\circ/X_0; \mu_{N(I)}] \prod_{i \in I} (\mathbb{L}^{\nu_i} T^{-N_i} - 1)^{-1}.$$

Proof. — Start with the identity of Proposition 4.2 (with $r = 1$). Omit at both sides the constant terms (on the right this amounts to summing over nonempty I only), and restrict the resulting identity to the fiber over $1 \in \mathbb{G}_m$. If we take into account the monodromies and use Lemma 5.2, we get the asserted identity, at least if we take our coefficients in $\hat{M}_{X_0}^{\hat{\mu}}$. Inspection of the proof shows that this actually holds in $M_{X_0}^{\hat{\mu}}[[T]]$. (The reason

is that X is smooth, so that in this derivation no coefficient involves an infinite sum of stable sets.) □

So the expression at the righthand side is independent of the resolution, something that is not at all evident a priori. Notice that it lies in a rather small subring of $M_{X_0}^{\hat{\mu}}[[T]]$: if R denotes the $\mathbb{Z}[L]$ -subalgebra of $\mathbb{Q}(L, T)$ generated by the fractions $(L^\nu T^{-N} - 1)^{-1}$ with $\nu, N > 0$, then $S(f)$ lies in $R \otimes_{\mathbb{Z}[L]} M_{X_0}^{\hat{\mu}}$. This allows us for instance to take $T = \infty$, that is, we apply the $\mathbb{Z}[L]$ -algebra homomorphism $R \rightarrow \mathbb{Z}[L]$ that sends $(L^\nu T^{-N} - 1)^{-1}$ to -1 . Thus

$$S(f)|_{T=\infty} = - \sum_{\emptyset \neq I \subset \text{Cirr}(E)} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0; \mu_{N(I)}]$$

Comparison with ordinary monodromy

The element $-S(f)|_{T=\infty}$ has an interpretation in terms of the nearby cycle sheaf of f as we shall now explain.

Suppose first that $k = \mathbb{C}$. Let $\widetilde{X - X_0} \rightarrow X - X_0 \subset X$ be the pull-back along f of the universal covering $\text{exp} : \mathbb{C} \rightarrow \mathbb{C}^\times \subset \mathbb{C}$. Take the full direct image of the constant sheaf $\mathbb{Q}_{\widetilde{X - X_0}}$ on X and restrict to X_0 : this defines ψ_f as an element of the derived category of constructible sheaves on X_0 . Let $\sigma : \widetilde{X - X_0} \rightarrow \widetilde{X - X_0}$ be a generator of the covering transformation that induces in \mathbb{C} translation over $-2\pi\sqrt{-1}$. This generator has the property that its action in ψ_f is the monodromy.

Let $H : Y \rightarrow X$ be a resolution as in 4.1. In the same way, $\psi_{\tilde{f}}$ is defined as an element of the derived category of constructible sheaves on the zero set Y_0 of \tilde{f} . The full direct image of $\psi_{\tilde{f}}$ on X_0 is equal to ψ_f .

An elementary calculation shows that the stalk of $\psi_{\tilde{f}}$ at a point of E_I° is the cohomology of N_I copies of a real torus of dimension $N_I - 1$. More precisely, the restriction of $\psi_{\tilde{f}}$ to E_I° is naturally representable as the full direct image of the constant sheaf on $U_I(1)$ (an algebraic torus bundle of dimension $N_I - 1$ over \tilde{E}_I°) under the projection $U_I(1) \rightarrow E_I^\circ$. We have a canonical isomorphism $H^k(\mathbb{G}_m^r; \mathbb{Q}) \cong H_c^{k+r}(\mathbb{G}_m^r; \mathbb{Q})$ and hence the Euler characteristic $\sum_k (-1)^k [H^k(\mathbb{G}_m^r; \mathbb{Q})]$ in $K_0(HS)$ is $(-1)^r$ times the Euler characteristic $\sum_k (-1)^k [H_c^k(\mathbb{G}_m^r; \mathbb{Q})]$. In other words, it is the value of χ_h on $(1 - \mathbb{L})^r$. Hence, if Z is a subvariety of E_I° with preimage \tilde{Z} in \tilde{E}_I° , then $\sum_k (-1)^k [H_c^k(Z; \psi_{\tilde{f}})]$ is the value of χ_h on $(1 - \mathbb{L})^{|I|-1} [\tilde{Z}; \mu_{N(I)}]$. This shows that ψ_f and $-S(f)|_{T=\infty}$ have the same Hodge characteristic. We therefore put

$$[\psi_f] := -S(f)|_{T=\infty} = \sum_{\emptyset \neq I \subset \text{Cirr}(E)} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0; \mu_{N(I)}].$$

We refer to $[\psi_f]$ as the *nearby cycle class* of f along X_0 . Its component in the augmentation submodule,

$$[\phi_f] := [\psi_f] - \overline{[\psi_f]} \in M_{X_0}^{\hat{\mu}},$$

is by definition the *vanishing cycle class* of f .

Let n be a positive integer. Given a S -variety Z with a good topological $\hat{\mu}$ -action, then the fixed point locus of $\ker(\hat{\mu} \rightarrow \mu_n)$ in Z is a S -variety which inherits a good μ_n -action. This defines a homomorphism of M_S -algebras

$$\mathrm{Tr}_n : M_S^{\hat{\mu}} \rightarrow M_S^{\mu_n}.$$

If $\sigma \in \hat{\mu}$ generates a dense subgroup of $\hat{\mu}$, then the fixed point locus of $\ker(\hat{\mu} \rightarrow \mu_n)$ is also the fixed point locus of σ^n . In case $k \subset \mathbb{C}$, a Lefschetz fixed point formula (applied to a partition of Z by orbit type) implies that $\chi_h[Z^{\sigma^n}]$ equals the trace of σ^n in $\chi_h[Z]$. So we may then think of $\mathrm{Tr}_n[Z]$ as the motivic trace of σ^n . This is why the following proposition is a motivic version of a result of A'Campo [2].

PROPOSITION 5.4 (see Denef-Loeser [18]). — For every positive integer n , the classes $[\mathcal{L}(X)_n^f(1)/X_0]$ and $\mathrm{Tr}_n[\psi_f]$ have the same image in $M_{X_0}^{\mu_n}/(\mathbb{L} - 1)$.

Proof. — The monodromy σ acts on \tilde{E}_I° as a covering transformation of order N_I . So σ^n has no fixed point if N_I does not divide n and is equal to all of \tilde{E}_I° otherwise. It follows from formula for the nearby cycle class that

$$\mathrm{Tr}_n[\psi_f] = \sum_{I \subset \mathrm{Cirr}(E), N_I | n} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0].$$

So

$$\begin{aligned} \sum_{n=1}^{\infty} \mathrm{Tr}_n[\psi_f] T^n &= \sum_{n=1}^{\infty} \sum_{I \subset \mathrm{Cirr}(E), N_I | n} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0] T^n \\ &= \sum_{\emptyset \neq I \subset \mathrm{Cirr}(E)} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0] \sum_{k \geq 1} T^{kN_I} = \sum_{\emptyset \neq I \subset \mathrm{Cirr}(E)} (1 - \mathbb{L})^{|I|-1} [\tilde{E}_I^\circ / X_0] \frac{T^{N_I}}{1 - T^{N_I}}. \end{aligned}$$

If we reduce modulo $(\mathbb{L} - 1)$ only the terms with I a singleton remain. Theorem 5.3 shows that this has the same reduction modulo $(\mathbb{L} - 1)$ as $S(f)$. □

6. THE MOTIVIC ZETA FUNCTION OF DENEFF-LOESER [13]

This function is a motivic analogue of Igusa's local zeta function. It captures slightly less than the function $S(f)$, but has the virtue that it is defined in greater generality. First we introduce three homomorphisms of Grothendieck rings.

An arrow $M_S^{\mu_{rn}} \rightarrow M_S^{\mu_n}$ is defined by assigning to a variety with good μ_{rn} -action its orbit space with respect to the subgroup $\mu_r \subset \mu_{rn}$ (with a residual action of μ_n). The totality of these arrows forms a projective system whose limit we denote by $M_S(\hat{\mu})$. This is not the same as $M_S^{\hat{\mu}}$, but there is certainly a natural ring homomorphism

$$\rho : M_S^{\hat{\mu}} \rightarrow M_S(\hat{\mu}).$$

It is given by assigning to a variety X with good $\hat{\mu}$ -action, the system $(X_n)_n$, where X_n is the orbit space of X by the kernel of $\hat{\mu} \rightarrow \mu_n$ endowed with the residual action of μ_n .

We next define the *Kummer map*

$$M_{S \times \mathbb{G}_m} \rightarrow M_S(\hat{\mu}), \quad [f] \mapsto [f]^{1/\infty}.$$

Given a S -variety Y and a morphism $f : Y \rightarrow \mathbb{G}_m$, then for every positive integer n , let $f^{1/n} : Y(f^{1/n}) \rightarrow \mathbb{G}_m$ be the pull-back of f over the n th power map $[n] : \mathbb{G}_m \rightarrow \mathbb{G}_m$. So $Y(f^{1/n})$ is the hypersurface in $\mathbb{G}_m \times Y$ defined by $f(z) = u^n$. The projection of $Y(f^{1/n}) \rightarrow Y$ is a μ_n -covering and thus defines an element $[f]^{1/n}$ of $M_S^{\mu_n}$. Notice that $Y(f^{1/n})$ is the orbit space of $Y(f^{1/nr})$ relative to the subgroup $\mu_r \subset \mu_{rn}$. Hence the $[f]^{1/n}$'s define an element $[f]^{1/\infty} \in M_S(\hat{\mu})$.

We also need the *mapping torus* of a variety with μ_n -action X : this is the étale locally trivial fibration $\mathbb{G}_m \times^{\mu_n} X \rightarrow \mathbb{G}_m$ whose total space is the orbit space of the μ_n -action on $\mathbb{G}_m \times X$ defined by $\zeta(\lambda, x) = (\lambda\zeta^{-1}, \zeta x)$ and for which the projection is induced by $(\lambda, x) \mapsto \lambda^n$. Notice that the fiber over $1 \in \mathbb{G}_m$ can be identified with X and that the monodromy is given by the action of μ_n on X . If $m = kn$ is a positive multiple of n and we let μ_m act on X via $\mu_m \rightarrow \mu_n$, then $(\lambda, x) \mapsto (\lambda^k, x)$ identifies the two fibrations. So we have in fact an M_S -linear map

$$M_S^{\hat{\mu}} \rightarrow M_{S \times \mathbb{G}_m}.$$

It is probably injective. The following lemma is a straightforward exercise.

LEMMA 6.1. — The composition of the mapping torus construction and the Kummer map is equal to $(\mathbb{L} - 1)\rho$.

For \mathcal{X} a smooth \mathbb{D} -variety of pure relative dimension d , define the *Denef-Loeser zeta function* by

$$I(f) := \mathbb{L}^{-d} \sum_{n=0}^{\infty} [\text{ac}_{f,n}]^{1/\infty} \mathbb{L}^{-sn} \in M_X(\hat{\mu})[[\mathbb{L}^{-s}]],$$

where \mathbb{L}^{-s} is just a variable with a suggestive notation. Then Corollary 4.3 and Lemma 6.1 yield

THEOREM 6.2. — The following identity holds in $M_X(\hat{\mu})[[\mathbb{L}^{-s}]]$:

$$I(f) = \mathbb{L}^{-d} \sum_{I \subset \text{irr}(E)} \rho[\tilde{E}_I^\circ/X; \mu_{N(I)}] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i + sN_i} - 1}.$$

Putting $\mathbb{L} = 1$

Consider the $\mathbb{Z}[L, L^{-1}]$ -subalgebra S of $\mathbb{Q}(L, L^{-s})$ generated by the rational functions $(L - 1)(L^{n+sN} - 1)^{-1}$, $n, N \geq 1$. The spectrum of S contains the generic point of the exceptional divisor of the blow up of $(1, 1)$ in $\mathbb{G}_m \times \mathbb{A}^1$. The corresponding specialization is

the evaluation homomorphism $S \rightarrow \mathbb{Q}(s)$ which sends $(L - 1)(L^{n+sN} - 1)^{-1}$ to $(n + sN)^{-1}$. According to Theorem 6.2, $I(f)$ lies in $S \otimes_{\mathbb{Z}[L, L^{-1}]} M_X(\hat{\mu})$. Evaluation at $\mathbb{L} = 1$ yields

$$I(f)|_{\mathbb{L}=1} = \sum_{I \subset \text{Irr}(D)} \rho[\tilde{E}_I^\circ/X; \mu_{N(I)}] \prod_{i \in I} \frac{1}{\nu_i + sN_i} \in M_X(\hat{\mu})/(\mathbb{L} - 1) \otimes_{\mathbb{Z}} \mathbb{Q}(s).$$

This is the motivic incarnation of the topological zeta function considered earlier by Denef and Loeser in [12]. At the time the resolution independence of this function was established using Theorem 6.3 below.

Comparison with Igusa’s p -adic zeta function

Suppose we are given a complete discrete valuation ring (R, m) of characteristic zero whose residue field $F = R/m$ has finite cardinality q . Then R contains all the $(q - 1)$ st roots of unity μ_{q-1} and this group projects isomorphically onto F^\times . Let K be the quotient field of R . If we choose a uniformizing parameter $\pi \in m - m^2$, then then the collection $(\zeta\pi^k)_{\zeta \in \mu_{q-1}, k \in \mathbb{Z}}$ is a system of representatives of $\bigoplus_{k \in \mathbb{Z}} m^k/m^{k+1}$. Define

$$\text{ac}^s : K \rightarrow \mathbb{Z}[\mu_{q-1}][q^{-s}]$$

by assigning to $u \in \zeta\pi^k + m^{k+1}$ the value ζq^{-ks} and 0 to 0. (Here q^{-s} is just the name of a variable; the righthand side can be more canonically understood as the group algebra of $K^\times/(1 + m)$.) There is a natural (additive) Haar measure μ on the the Boolean algebra of subsets of K generated by the cosets of powers of m that takes the value 1 on R . It takes values in $\mathbb{Z}[q^{-1}]$. Given $f \in R[x_1, \dots, x_d]$ whose reduction mod m is nonzero, then its *Igusa local zeta function* is defined by

$$Z(f) := \int_{R^m} \text{ac}^s f(x) d\mu(x),$$

where R^m is endowed with the product measure. We regard this as an element of $\mathbb{Q}[\mu_{q-1}][[q^{-s}]]$: the coefficient of ζq^{-ns} is the volume of $f^{-1}(\zeta\pi^n + m^{n+1})$. (It is customary to let s be a complex number—the series then converges in a right half plane—and to compose with a complex character $\alpha : \mu_{q-1} \rightarrow \mathbb{C}^\times$.)

Let us write \mathcal{X} for $\text{Spec}(R[x_1, \dots, x_d])$ and regard f as a morphism $\mathcal{X} \rightarrow \mathbb{A}_R^1$ over $\text{Spec}(R)$. Suppose we have an embedded resolution $H : \mathcal{Y} \rightarrow \mathcal{X}$ of the zero locus of f over $\text{Spec}(R)$ with a simple normal crossing hypersurface \mathcal{E} relative to $\text{Spec}(R)$ (so no irreducible component in the closed fiber). Then we get an embedded resolution of the closed fiber $Y \rightarrow X$ with simple normal crossing divisor E . Make a base change of $fH : \mathcal{Y} \rightarrow \mathbb{A}_R^1$ over the $(q - 1)$ st power map $[q - 1] : \mathbb{A}_R^1 \rightarrow \mathbb{A}_R^1$ and normalize; this gives a morphism $\tilde{\mathcal{Y}} \rightarrow \mathcal{Y}$ over a finite extension of R and defines a μ_{q-1} -covering $\tilde{\mathcal{E}}_I^\circ \rightarrow \mathcal{E}_I^\circ$ in much the same way as before. This covering is defined over R . Let E_I° be its reduction modulo the maximal ideal of R . This is a variety defined over F and so we choose an algebraic closure $F \subset \bar{F}$, then we can consider the trace of the Frobenius acting on the ℓ -adic Euler characteristic $\sum_i (-1)^i [H^i(\tilde{E}_I^\circ(\bar{F}); \mathbb{Q}_\ell)]$, where $\ell \in \mathbb{N}$ is relatively prime to

q . This trace is in fact a rational integer, for it equals the cardinality of $\hat{E}_I^\circ(F)$, the set of F -rational points of \hat{E}_I° . The latter interpretation makes it evident that the presence of the μ_{q-1} -action lifts the trace to an element in the group algebra $\mathbb{Z}[\mu_{q-1}]$, namely $\#(E_I^\circ(F)) \sum_{\zeta \in \mu_{q-1}} \zeta$. Let us denote that element $[\hat{E}_I^\circ(F); \mu_{q-1}]$.

Denef proved earlier [10] the following R -analogue of 6.2:

THEOREM 6.3 (Denef). — In this situation we have

$$Z(f) = q^{-d} \sum_{I \subset \text{Irr}(E)} [\hat{E}_I^\circ(F); \mu_{q-1}] \prod_{i \in I} \frac{q-1}{q^{\nu_i + sN_i} - 1},$$

where ν_i and N_i have the usual meaning.

As appears from 6.2, $Z(f)$ is what we get from the value of $I(f_{\bar{K}})$ on $X_0(\bar{K})$ (with \bar{K} an algebraic closure of K) if we replace classes in $M_{\bar{K}}$ by the number of F -rational points in their F -counterparts (so that we substitute q for \mathbb{L}) and pass from $\hat{\mu}$ to μ_{q-1} . This should be understood on a more conceptual level that involves a Grothendieck ring $M_{\text{Spec}(R)}^{\mu_{q-1}}$ which specializes to both $M_{\text{Spec}(\bar{K})}^{\mu_{q-1}}$ and $\mathbb{Q}[\mu_{q-1}][[q^{-s}]]$, and avoids resolution.

7. MOTIVIC CONVOLUTION [15]

An ulterior motivic ring

Consider the Fermat curve J_n in \mathbb{G}_m^2 defined by $u^n + v^n = 1$. Notice that it is invariant under the subgroup $\mu_n^2 \subset \mathbb{G}_m^2$. If d is a positive divisor of n , then the μ_d^2 -orbit space of J_n is $J_{n/d}$. In particular, the μ_n^2 -orbit space of J_n is J_1 , an affine line less two points. Given varieties X and Y with good μ_n -action, then we have the variety with $\mu_n \times \mu_n$ -action

$$J_n(X, Y) := J_n \times^{(\mu_n \times \mu_n)} (X \times Y).$$

(If a group G acts well on varieties A and B , then $A \times^G B$ stands for quotient of $A \times B$ by the equivalence relation $(ga, b) \sim (a, gb)$ with G acting well on it by $g[a, b] := [ga, b] = [a, gb]$.) Let μ_n act on $J_n(X, Y)$ diagonally: $\zeta[(u, v), (x, y)] := [(\zeta u, \zeta v), (x, y)]$. The natural map $J_n(X, Y) \rightarrow J_1$ is étale locally trivial. If Y has trivial μ_n -action, then $J_n(X, Y) = J_n(X, pt) \times Y$ and the variety $J_n(X, pt)$ can be identified with $(\mathbb{G}_m - \{\mu_n\}) \times^{\mu_n} X$. The latter has the structure of a piecewise \mathbb{G}_m -bundle over \bar{X} from which a copy of X has been removed. Similarly, the natural projection of $\overline{J_n(X, Y)} \rightarrow \bar{X} \times \bar{Y}$ is a piecewise \mathbb{G}_m -bundle from which a copy of $\bar{X} \times \bar{Y}$ has been removed.

Clearly, $J_n(X, Y) \cong J_n(Y, X)$. The operation J_n is also associative: this is seen as follows: the orbit space of $J_n \times J_n$ with respect to the μ_n action defined by $\zeta((u, v), (w, t)) = ((u, \zeta^{-1}v), \zeta w, \zeta t)$ gives the Fermat surface J_n^2 in \mathbb{G}_m^3 defined by $u^n + x^n + y^n = 1$ (put $x := \alpha v w, y := \alpha v t$, where $\alpha^n = -1$). So $J_n(X, J_n(Y, Z)) = J_n^2 \times^{(\mu_n)^3} X \times Y \times Z$, and likewise for $J_n(J_n(X, Y), Z)$.

If m is a divisor of n and the action of μ_n on X and Y is through μ_m , then $J_m(X, Y) = J_n(X, Y)$. So this induces a binary operation

$$J : M_k^{\hat{\mu}} \times M_k^{\hat{\mu}} \rightarrow M_k^{\hat{\mu}}.$$

Summing up:

LEMMA 7.1. — The operation J is associative, commutative and bilinear over M_k . Moreover

- (i) $J(a, 1) = (\mathbb{L} - 1)\bar{a} - a$ and
- (ii) $\overline{J(a, b)} = (\mathbb{L} - 1)\bar{a}\bar{b} - \overline{ab}$,

where we recall that $a \in M_k^{\hat{\mu}} \mapsto \bar{a} \in M_k$ is the augmentation defined by ‘passing to the orbit space’.

The construction is perhaps better understood in terms of the fibrations over \mathbb{G}_m defined by the mapping torus construction. Recall that for a variety X with μ_n -action, its mapping torus $\mathbb{G}_m \times^{\mu_n} X$ fibers over \mathbb{G}_m by $[\lambda, x] \mapsto \lambda^n$ with $\{1\} \times X$ mapping to the fiber over 1. The monodromy is the given μ_n -action on X . If Y is another variety with μ_n -action, then the composite

$$(\mathbb{G}_m \times^{\mu_n} X) \times (\mathbb{G}_m \times^{\mu_n} Y) \longrightarrow \mathbb{G}_m \times \mathbb{G}_m \subset \mathbb{G}_a \times \mathbb{G}_a \xrightarrow{+} \mathbb{G}_a$$

is a fibration over \mathbb{G}_m . The fiber over $1 \in \mathbb{G}_a$ is identified as $J_n(X, Y)$ and the monodromy is the given μ_n -action on $J_n(X, Y)$ defined above.

For reasons that we hope become clear later we are interested in maps G from $M_k^{\hat{\mu}}$ to commutative rings U with unit satisfying the following two properties:

- G is linear over a unitary homomorphism $M_k \rightarrow U$ and
- $G(a)G(b) + GJ(a, b) + G((1 - \mathbb{L})\bar{a}\bar{b}) = 0$ for all $a, b \in M_k^{\hat{\mu}}$.

Anderson’s notion of ulterior motive [3] suggests to call such a map *ulterior*.

EXAMPLE 7.2. — It follows from Lemma 7.1-ii that the map $M_k^{\hat{\mu}} \rightarrow M_k$ defined by $a \mapsto (1 - \mathbb{L})\bar{a}$ is ulterior.

EXAMPLE 7.3. — A *Hodge structure with fractional weights* is a finite dimensional rational vector space H with a decomposition of its complexification $\mathbb{C} \otimes H = \bigoplus_{p,q \in \mathbb{Q}; p+q \in \mathbb{Z}} H^{p,q}$ with the usual conditions: $H^{q,p}$ is the complex conjugate of $H^{p,q}$ and $\bigoplus_{p+q=n} H^{p,q}$ is defined over \mathbb{Q} . We denote the corresponding category $\text{HS}(\mathbb{Q})$ and the associated Grothendieck ring by $K_0(\text{HS}(\mathbb{Q}))$. (The latter can also be described as the virtual representation ring of the universal profinite cover of the Deligne group $\text{Res}_{\mathbb{R}|\mathbb{C}} \mathbb{G}_m$.) Every Hodge structure H with $\hat{\mu}$ -action can be turned into a Hodge structure with fractional weights: decompose $\mathbb{C} \otimes H^{p,q}$ according to the characters of $\hat{\mu}$:

$$\mathbb{C} \otimes H^{p,q} = \bigoplus_{\alpha \in \mathbb{Q}/\mathbb{Z}} H_{\alpha}^{p,q}.$$

For $\alpha = 0$, the fractional bidegree of $H_\alpha^{p,q}$ will remain (p, q) ; otherwise we choose it to be $(p + \tilde{\alpha}, q + 1 - \tilde{\alpha})$, where $\tilde{\alpha}$ is the representative of α in the interval $[0, 1[$. We define $G_h : K_0^{\hat{\mu}}(\text{HS}) \rightarrow K_0(\text{HS}(\mathbb{Q}))$ by applying the above recipe. The eigen space of $\mu_n \times \mu_n$ in $H_c^1(J_n)$ with character $(\alpha, \beta) \in (n^{-1}\mathbb{Z}/\mathbb{Z})^2$ is of dimension two for $(\alpha, \beta) = (0, 0)$ and one-dimensional in all other cases; the Hodge bidegree of this eigen space is $(0, 1)$ if $0 < \tilde{\alpha} + \tilde{\beta} < 1$ and $\alpha \neq 0 \neq \beta$, $(1, 0)$ if $1 < \tilde{\alpha} + \tilde{\beta} < 2$, and $(0, 0)$ in the remaining cases ($\alpha = 0, \beta = 0$ or $\alpha + \beta = 0$). It is then straightforward to verify that the composite of the Hodge character with G_h is ulterior.

The universal ulterior map is obtained as follows. Consider the ring $M_k^{\hat{\mu}}\langle G \rangle$ of noncommutative polynomials over $M_k^{\hat{\mu}}$ in a single variable G and let M_k^{ult} be the quotient of the two sided ideal generated by G in $M_k^{\hat{\mu}}\langle G \rangle$ by the left ideal generated by $Ga - aG$ with $a \in M_k$ and $GaGb + GJ(a, b) + G(1 - \mathbb{L})\overline{ab}$ with $a, b \in M_k^{\hat{\mu}}$. As a right $M_k^{\hat{\mu}}$ -module, M_k^{ult} is generated by the image of G . So if we think of G as a map

$$G : M_k^{\hat{\mu}} \rightarrow M_k^{\text{ult}}, \quad a \mapsto Ga,$$

then G is ulterior and surjective. Example 7.2 shows that $(1 - \mathbb{L})$ times the augmentation factorizes over G . So the kernel of $G|_{M_k}$ is $(\mathbb{L} - 1)$ -torsion.

Convolution

In what follows we need the (additive) group structure on the affine line, so we write \mathbb{G}_a instead of \mathbb{A}^1 . Remember that we have a bijection $\mathcal{L}(\mathbb{G}_a, 0) \cong \mathfrak{m}$, defined by assigning to $\gamma \in \mathcal{L}(\mathbb{G}_a, 0)$ the pull-back of the standard coordinate on \mathbb{G}_a . We fix a uniformizing parameter $t \in \mathfrak{m}$. An $\text{Aut}(\mathbb{D})$ -invariant M_k -valued measure λ on $\mathcal{L}(\mathbb{G}_a, 0)$ is determined by giving for every positive integer n its value $a_n(\lambda)$ on $t^n + \mathfrak{m}^{n+1}$, viewed as an element of $M_k^{\mu_n}$, and its value $A_n(\lambda)$ on \mathfrak{m}^n , an element of M_k . These values are not independent, for we have the relation $A_n(\lambda) - A_{n+1}(\lambda) = (\mathbb{L} - 1)\overline{a}_n(\lambda)$. The value of λ on all of $\mathcal{L}(\mathbb{G}_a, 0)$, $A_1(\lambda)$, is called the *mass* of λ . So the measure λ is completely determined by the series $\sum_n a_n(\lambda)T^n$ and its mass. As we shall take coefficients in $\hat{M}_k^{\hat{\mu}}$, we can dispense with the mass since the latter is $\sum_n (\mathbb{L} - 1)\overline{a}_n(\lambda)$ in $\hat{M}_k^{\hat{\mu}}$. We therefore need not make a distinction between an $\text{Aut}(\mathbb{D})$ -invariant \hat{M}_k -valued measure on $\mathcal{L}(\mathbb{G}_a, 0)$ and a series $a(T) = \sum_{n>0} a_n T^n$ with $a_n \in \hat{M}_k^{\mu_n}$, for which $\sum_n (\mathbb{L} - 1)\overline{a}_n$ converges in \hat{M}_k . These series define a \hat{M}_k -submodule of $\hat{M}_k^{\hat{\mu}}[[T]]$ that we shall denote $\hat{M}_k^{\hat{\mu}}\{T\}$. Addition $\mathbb{G}_a \times \mathbb{G}_a \rightarrow \mathbb{G}_a$ defines a map

$$\text{add} : \mathcal{L}(\mathbb{G}_a, 0) \times \mathcal{L}(\mathbb{G}_a, 0) \rightarrow \mathcal{L}(\mathbb{G}_a, 0).$$

The convolution of the measures λ and λ' is the direct image of the product measure under this map: $\lambda * \lambda' := \text{add}_*(\lambda \times \lambda')$. This will be an $\text{Aut}(\mathbb{D})$ -invariant measure and thus $\hat{M}_k^{\hat{\mu}}\{T\}$ becomes a commutative \hat{M}_k -algebra (whose product we continue to denote by $*$).

LEMMA 7.4. — We have

$$(a * b)_n = \sum_{i < n} (\mathbb{L} - 1) \overline{a_i b_i} \mathbb{L}^{i-n} + J(a_n, b_n) + a_n B_{n+1} + A_{n+1} b_n.$$

Proof. — Let $\alpha = \sum_{k \geq i} \alpha_k t^k \in \mathfrak{m}^i - \mathfrak{m}^{i+1}$ and $\beta = \sum_{k \geq j} \beta_k t^k \in \mathfrak{m}^j - \mathfrak{m}^{j+1}$, (so α_i and β_j nonzero), such that their sum is in $t^n + \mathfrak{m}^{n+1}$. If one of i, j is $> n$, say $i > n$, then we must have $j = n$ and $\beta_n = 1$. It is clear that then $a_i T^i * b_n T^n = (\mathbb{L} - 1) \overline{a_i b_n} T^n$.

If i and j are both $\leq n$, then $i = j$ and we must have $\alpha_k + \beta_k = 0$ for $k = i, \dots, n-1$ and $\alpha_n + \beta_n = 1$. Let us first consider the case $i < n$. Then these conditions define a subvariety $\cong \mathbb{G}_m \times \mathbb{A}^{n-i}$ in $(\mathfrak{m}^i - \mathfrak{m}^{i+1} / \mathfrak{m}^{n+1})^2 \cong (\mathbb{G}_m \times \mathbb{A}^{n-i})^2$ whose image in M_k is $(\mathbb{L} - 1) \mathbb{L}^{n-i}$. Since there is no monodromy here, we find that $(a_i T^i * b_i T^i)_n = (\mathbb{L} - 1) \overline{a_i b_i}$.

When $i = j = n$, the subvariety in $(\mathfrak{m}^n - \mathfrak{m}^{n+1} / \mathfrak{m}^{n+1})^2 \cong \mathbb{G}_m^2$ is defined by $\alpha_n + \beta_n = 1$, in other words is equal to J_1 . From the above discussion one sees that $(a_n T^n * b_n T^n)_n = J(a_n, b_n)$. \square

Let us write \tilde{R} for the localization of $\mathbb{Z}[L, L^{-1}, T]$ obtained by inverting the monic polynomials $T^N - L^\nu$, where N is a positive integer and $\nu \in \mathbb{N}$. We are only interested in measures in

$$\hat{M}_k^{\hat{\mu}}(T) := \hat{M}_k^{\hat{\mu}}\{T\} \cap (\tilde{R} \otimes_{\mathbb{Z}[L, L^{-1}]} \hat{M}_k^{\hat{\mu}}).$$

Denote by $\hat{M}_k^{\text{ult}}(T)$ its image under G .

THEOREM 7.5 (Abstract Thom-Sebastiani property). — Convolution preserves the algebra $\hat{M}_k^{\text{ult}}(T)$ and if $a, b \in \hat{M}_k^{\text{ult}}(T)$ come from massless measures on $\mathcal{L}(\mathbb{G}_a, 0)$, then $(a * b)(\infty) = a(\infty)b(\infty)$.

COROLLARY 7.6. — Let X and Y be smooth connected varieties and $f : X \rightarrow \mathbb{G}_a$, $g : Y \rightarrow \mathbb{G}_a$ nonconstant morphisms with zero fibers X_0 and Y_0 . Let $f * g : X \times Y \rightarrow \mathbb{G}_a$ be defined by $(f * g)(x, y) := f(x) + g(y)$. Then the restriction of $[\phi_{f * g}]$ to $X_0 \times Y_0$ and the exterior product $[\phi_f] \times [\phi_g] \in M_{X_0 \times Y_0}$ have the same image under G . In particular, their images in the Grothendieck ring of variations of fractional Hodge structures over subvarieties of $X_0 \times Y_0$ are the same.

Thus we recover the Thom-Sebastiani property for the spectrum, proved earlier by Varchenko in case f and g have isolated singularities and by M. Saito [26] in general.

For the proof of Theorem 7.5 we need the following

LEMMA 7.7. — Let $a(T)$ and $b(T)$ be rational functions of a single complex variable T which are both zero in 0 and regular at ∞ . If $\sum_{k > 0} a_k T^k$ resp. $\sum_{k > 0} b_k T^k$ are their expansions at 0, then $\sum_{k > 0} a_k b_k T^k$ is the expansion at zero of a rational function $c(T)$ in z whose value at ∞ equals $-a(\infty)b(\infty)$. Moreover, if a and b also depend on a second complex variable L and lie in \tilde{R} (viewed as a subalgebra of $\mathbb{C}(L, T)$), then so does c .

Proof. — The idea of the proof is inspired by a paper of Deligne [9]. Let $r_0 > 0$ be a radius of convergence for the two expansions. Let $T \in \mathbb{C}$ be such that $|T| < r_0^2$ and choose $|T|/r_0 < r < r_0$. Consider the integral

$$c(T) := \frac{1}{2\pi\sqrt{-1}} \int_{|\tau|=r} a(T/\tau)b(\tau) \frac{d\tau}{\tau}.$$

On the circle of integration the expansions converge uniformly and absolutely and so

$$c(T) = \frac{1}{2\pi\sqrt{-1}} \int_{|\tau|=r} \sum_{k,l \in \mathbb{N}} a_k T^k b^l \tau^{k-l} \frac{d\tau}{\tau}.$$

Since summation and integration may be interchanged, only the terms with $k = l$ remain and hence $c(T) = \sum_{k \in \mathbb{N}} a_k b_k T^k$. If P_a resp. P_b denotes the set of poles of a resp. b , then the integrand has polar set $TP_a^{-1} \cup P_b$ (there is no pole in 0 or ∞) and the poles enclosed by the circle of integration are those in TP_a^{-1} . By the theory of residues, $-c(T)$ must then be equal to the sum of the residues of the integrand at P_b . This description no longer requires $|T| < r_0^2$ and defines an analytic extension of c to the complement of $P_a P_b$. This extension is easily seen to be meromorphic at $P_a P_b$. To compute its behavior at ∞ , we note that $a(T/\tau)$ converges for $T \rightarrow \infty$ on a neighborhood of P_b absolutely (with all its derivatives) to the constant function $a(\infty)$. So as $T \rightarrow \infty$, $-c(T)$ tends to the sum of the residues of $a(\infty)b(\tau)\tau^{-1} d\tau$ at P_b . This sum is opposite to the residue at the remaining pole ∞ , hence equal to $a(\infty)b(\infty)$. In particular, c is a rational function with polar set contained in $P_a P_b$.

Assume now that $a, b \in \tilde{R}$. A pole of an element of R in $\mathbb{C}^\times \times \mathbb{C}$ satisfies an equation $T^N = L^\nu$ for certain integers $N > 0$, $\nu \geq 0$. A product of such poles satisfies a similar equation, and this implies that a product of c and a finite set of polynomials of the form $T^N - L^\nu$ is in $\mathbb{C}[L, L^{-1}, T]$. Since the expansion of c at $T = 0$ has integral coefficients, this product lies in $\mathbb{Z}[L, L^{-1}, T]$. \square

Proof of Theorem 7.5. — We start with the convolution formula 7.4. It says that

$$(a * b)(T) = \sum_{0 < i < n} (\mathbb{L} - 1) \overline{a_i b_i} \mathbb{L}^{i-n} T^n + \sum_{n > 0} J(a_n, b_n) T^n + \sum_{n > 0} (a_n B_{n+1} + A_{n+1} b_n) T^n.$$

We now assume that a and b are massless so that $A_{n+1} = -(\mathbb{L} - 1) \sum_{i=1}^n \overline{a_i}$ and similarly $B_{n+1} = -(\mathbb{L} - 1) \sum_{i=1}^n \overline{b_i}$. Since we work in $\hat{M}_k^{\text{ult}}(T)$, we have

$$(a * b)(T) = (\mathbb{L} - 1) \sum_{0 < i < n} \overline{a_i b_i} \mathbb{L}^{i-n} T^n + \\ - \sum_{n > 0} a_n b_n T^n + (\mathbb{L} - 1) \sum_{n > 0} \overline{a_n b_n} T^n - (\mathbb{L} - 1) \sum_{0 < i \leq n} (a_n \overline{b_i} + \overline{a_i} b_n) T^n.$$

We consider each series on the right separately. We have

$$\begin{aligned}
 (\mathbb{L} - 1) \sum_{0 < i < n} \overline{a_i b_i} \mathbb{L}^{i-n} T^n &= (\mathbb{L} - 1) \sum_{i > 0} \sum_{k > 0} \overline{a_i b_i} \mathbb{L}^{-k} T^{k+i} \\
 &= -(\mathbb{L} - 1) T (T - \mathbb{L})^{-1} \sum_{i > 0} \overline{a_i b_i} T^i.
 \end{aligned}$$

By Lemma 7.7 the last expression lies in \tilde{R} and takes the value $+(\mathbb{L} - 1)\overline{a(\infty)b(\infty)}$ at ∞ . This is the opposite of the value of $(\mathbb{L} - 1) \sum_{n > 0} \overline{a_n b_n} T^n$ at ∞ , so these terms cancel. Similarly $-\sum_{n > 0} a_n b_n T^n$ takes the value $a(\infty)b(\infty)$ at ∞ . Since

$$\sum_{0 < i \leq n} \bar{a}_i T^n = -(T - 1)^{-1} \sum_{i > 0} \bar{a}_i T^i$$

is a rational function in T with value zero at ∞ it follows from Lemma 7.7 that the same is true for $(\mathbb{L} - 1) \sum_{0 < i \leq n} (\bar{a}_i b_n) T^n$. Likewise for $(\mathbb{L} - 1) \sum_{0 < i \leq n} (a_n \bar{b}_i) T^n$. So $(a * b)(T)$ is in \tilde{R} and has value $a(\infty)b(\infty)$ at ∞ . \square

8. THE MCKAY CORRESPONDENCE [6], [16], [25]

Suppose a group G of finite order m acts well and effectively on a smooth connected variety U of dimension d . This defines an orbifold $p : U \rightarrow U_G$ with underlying variety $G \backslash U$. Let us write X for the orbifold U_G . We also fix a primitive m th root of unity ζ_m .

Let $g \in G$ and let U^g be its fixed point set in U . The action of g in the normal bundle of U^g decomposes that bundle into a direct sum of eigen subbundles

$$\nu_{U/U^g} = \bigoplus_{k=1}^{m-1} \nu_g^k,$$

where ν_g^k has eigen value ζ_m^k . We like to think of ν_g^k as the pull-back of a fractional bundle on a subvariety of X whose virtual rank is k/m times that of ν_g^k . A more formal discussion involves the extension $M_X[\mathbb{L}^{1/m}]$ of M_X obtained by adjoining an m th root of \mathbb{L} . To be precise, let $w(g) := \sum_k \frac{k}{m} \text{rk}(\nu_g^k)$, considered as locally constant function $U^g \rightarrow m^{-1}\mathbb{Z}$, and let $\mathbb{L}_{U^g}^{w(g)}$ be the element of $M_{U^g}[\mathbb{L}^{1/m}] \subset M_U[\mathbb{L}^{1/m}]$ that this defines. Then $\sum_{g \in G} \mathbb{L}_{U^g}^{w(g)}$ is the image under p^* of

$$W(X) = \sum_{[g] \in \text{conj}(G)} \sum_{i \in \pi_0(U^g)} [(G_i \backslash U_i^g) / X] \mathbb{L}^{w_i} \in M_X[\mathbb{L}^{1/m}].$$

Here U_i^g is the connected component of U^g labeled by i , G_i is the G -stabilizer of this component, and w_i the value of $w(g)$ on U_i^g . The sum is over a system of representatives of the conjugacy classes of G and can be rewritten as one over the orbifold strata of X (see Reid [25]): the decomposition of U into connected strata by orbit type (a stratum is a connected component of the locus of points with given G -stabilizer) induces a partition of X into orbifolds and $W(X)$ has the form $\sum_S [S] W_S$, where the sum is over the orbifold

strata, and W_S is a polynomial in $\mathbb{L}^{1/m}$. We will see that $W(X)$ can be understood as the class of an obstruction bundle for lifting arcs in X to arcs in U .

The McKay correspondence identifies $W(X)$ in terms of a resolution of X :

THEOREM 8.1 (Batyrev [6], Denef-Loeser [16]). — Let $H : Y \rightarrow X$ be a resolution of the orbifold X whose exceptional divisor E has simple normal crossings. With the usual meaning of E_I° and with ν_i^* as defined below we have the following identity in $\hat{M}_X[\mathbb{L}^{1/m}]$:

$$W(X) = \sum_{I \subset \text{irr}(D)} [E_I^\circ/X] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i^*} - 1}.$$

The statement does not involve arc spaces, but the proof does. It could well be that the identity is already valid in $M_X[\mathbb{L}^{1/m}]$. The relative simplicity of the lefthand side has implications for the righthand side, one of which is that all the ‘non-Tate’ material in a fiber of H must cancel out in the sum. For that same reason the lefthand side is hardly affected if we apply the weight character relative to X to it, that is, if we take the image of $W(X)$ in the Grothendieck ring of constructible $\mathbb{Z}((w^{-1/m}))$ -modules on X : just substitute w^2 for \mathbb{L} .

We first seek an orbifold measure on $\mu_{\mathcal{L}(X)}^{\text{orb}}$ on $\mathcal{L}(X)$ with the property that for every G -invariant measurable $A \subset \mathcal{L}(U)$ we have

$$\mu_{\mathcal{L}(X)}^{\text{orb}}(p_*A) := \overline{\mu_{\mathcal{L}(U)}(A)},$$

where the righthand side should be interpreted as follows: think of $\mu_{\mathcal{L}(U)}(A)$ as an element of \hat{M}_k^G , and then let $\overline{\mu_{\mathcal{L}(U)}(A)}$ be the image of $\mu_{\mathcal{L}(U)}(A)$ under the augmentation $\hat{M}_k^G \rightarrow \hat{M}_k$. Since $p_* : \mathcal{L}(U) \rightarrow \mathcal{L}(X)$ need not be surjective, this will not characterize the orbifold measure a priori. But it suggests how to define it: suppose that the Jacobian ideal \mathcal{J}_p has constant order e along A . Then the usual measure of $\mathcal{L}(X)$ pulled back to A is $\mathbb{L}^{-e} \mu_{\mathcal{L}(U)}|_A$. We therefore want the orbifold measure restricted to $p_*(A)$ to be the restriction of $\mathbb{L}^e \mu_{\mathcal{L}(X)}$. This can be done as follows. Let r be a positive integer such that $(\Omega_U^d)^{\otimes r}$ descends to an invertible sheaf $\omega_X^{(r)}$ on X . (So for every $u \in U$, G_u acts on the tangent space $T_u U$ with determinant an r th root of unity.) There is a natural homomorphism $(\Omega_X^d)^{\otimes r} \rightarrow \omega_X^{(r)}$ whose kernel is the torsion of $(\Omega_X^d)^{\otimes r}$. The image of this homomorphism has the form $\mathcal{I}^{(r)} \omega_X^{(r)}$ for an ideal $\mathcal{I}^{(r)}$. We set

$$\mu_{\mathcal{L}(X)}^{\text{orb}} := \mathbb{L}^{\text{ord}_{\mathcal{I}^{(r)}}/r} \mu_{\mathcal{L}(X)}.$$

It is a measure that takes values in $\hat{M}_k[\mathbb{L}^{1/r}]$.

LEMMA 8.2. — The pull-back of $\mu_{\mathcal{L}(X)}^{\text{orb}}$ under p^* is a measure that assigns to any G -invariant measurable subset A of $\mathcal{L}(U)$ the image of A under the augmentation map $\hat{M}_k^G \rightarrow \hat{M}_k$.

Proof. — If we apply p^* to the identity $(\Omega_X^d)^{\otimes r}/tors = \mathcal{I}^{(r)}\omega_X^{(r)}$ we get $\mathcal{J}_p^r(\Omega_U^d)^{\otimes r} = p^*(\mathcal{I}^{(r)}\omega_U^{\otimes r})$. Since $\Omega_U^d = \omega_U$, it follows that $p^*(\mathcal{I}^{(r)}) = \mathcal{J}_p^r$. So $\mu_{\mathcal{L}(X)}^{orb}$ pulls back under p^* to $\mathbb{L}^{-ord_{\mathcal{J}_p} + p^*(\mathcal{I}^{(r)})/r} \mu_{\mathcal{L}(U)} = \mu_{\mathcal{L}(U)}$. The rest is left to the reader. \square

The following lemma describes the direct image of $\mu_{\mathcal{L}(X)}^{orb}$ on X in terms of a resolution of X : let $Y \rightarrow X$ be a resolution of singularities with simple normal crossing divisor E . We have $H^*\omega_X^{(r)} = \tilde{\mathcal{I}}^{(r)}\omega_Y^{\otimes r}$ for some fractional ideal $\tilde{\mathcal{I}}^{(r)}$ on Y . It is known that the multiplicity m_i of E_i in this ideal is $> -r$. So $\nu_i^* := 1 + m_i/r$ is positive. Entirely analogous to the proof of Theorem 4.2 one derives:

LEMMA 8.3. — The direct image of $\mu_{\mathcal{L}(X)}^{orb}$ on X is represented by the class

$$\sum_{I \subset \text{Cirr}(E)} [E_I^\circ/X] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i^*} - 1} \in \hat{M}_X[\mathbb{L}^{1/r}].$$

Let $\mathcal{L}'(X)$ be the set of arcs in X not contained in the discriminant of $p : U \rightarrow X$. This is a subset of full measure. We decompose $\mathcal{L}'(X)$ according to the ramification behavior of $p : U \rightarrow X$. Let $[m] : \mathbb{D} \rightarrow \mathbb{D}$ be the m th power map and denote the parameter of the domain by $t^{1/m}$. We regard ζ_m (through its action on the domain) as generator of the Galois group of $[m]$. For $\gamma \in \mathcal{L}'(X)$, $\gamma[m]$ lifts to a morphism $\tilde{\gamma} : \mathbb{D} \rightarrow \tilde{X}$ and this lift is unique up to conjugation with G . Given the lift, there is a $g \in G$ such that $g\tilde{\gamma} = \tilde{\gamma}\zeta_m$. Its conjugacy class $[g]$ in G only depends on γ . This conjugacy class determines the isomorphism type of the G -covering over γ : if m' is the order of g , then $\gamma^*(p)$ is isomorphic to $G \times^{(g)} \mathbb{D} \rightarrow \mu_{m'} \backslash \mathbb{D}$, with g acting on \mathbb{D} as multiplication by $\zeta^{m/m'}$. Notice that $\tilde{\gamma}(o)$ is in the fixed point set U^g . The ‘fractional lifts’ $\tilde{\gamma}$ that so arise are like arcs in the total space of the normal bundle $\oplus_k \nu_g^k$ of U^g (based at the zero section) which in the ν_g^k -direction develop as $t^{k/m}$ times a power series in t .

Denote the set of arcs in $\mathcal{L}'(X)$ belonging to the conjugacy class of $[g]$ of g by $\mathcal{L}(X, [g])$. The McKay correspondence now results from:

LEMMA 8.4. — The subset $\mathcal{L}(X, [g])$ is measurable for $\mu_{\mathcal{L}(X)}^{orb}$ and the restriction of $\mu_{\mathcal{L}(X)}^{orb}$ to this subset is represented by the class $[(G_g \backslash U^g)/X] \mathbb{L}^{w(g^{-1})} \in M_X[\mathbb{L}^{1/m}]$, where G_g is the G -stabilizer of U^g .

The proof is a calculation which we only discuss in a heuristic fashion. The elements of $\mathcal{L}(X, [g])$ correspond to G_g -orbits of fractional lifts as described above. In view of our definition of orbifold measure, we need to argue that these fractional lifts are represented by the element $\mathbb{L}_{U^g}^{w(g^{-1})}$. If r_1, \dots, r_{m-1} are positive integers, then the arcs in $\oplus_k \nu_g^k$ of U^g based at the zero section and which in the ν_g^k -direction have order r_k make up a constructible subset of $\mathcal{L}(\oplus_k \nu_g^k)$ whose class is easily seen to be equal to $\mathbb{L}_{U^g}^w$, with $w = \sum_k (1 - r_k) \text{rk}(\nu_g^k)$. The fact is that this also holds for the fractional values $r_k = k/m$. So in that case we have $w = \sum_k (1 - k/m) \text{rk}(\nu_g^k) = w(g^{-1})$.

9. PROOF OF THE TRANSFORMATION RULE [14]

Let \mathcal{X}/\mathbb{D} be a \mathbb{D} -variety of pure relative dimension d . The d th Fitting ideal of $\Omega_{\mathcal{X}/\mathbb{D}}$ defines the locus where \mathcal{X} fails to be smooth over \mathbb{D} ; we denote that ideal by $\mathcal{J}(\mathcal{X}/\mathbb{D})$. Locally this ideal is obtained as follows: if \mathcal{X} is given as a closed subset of $(\mathbb{A}^{d+l})_{\mathbb{D}}$, then $\mathcal{J}_{\mathcal{X}/\mathbb{D}}$ is the restriction to \mathcal{X} of the ideal generated by the determinants $\det((\partial f_j/\partial x_{i_k})_{j,k=1}^l)$, where f_1, \dots, f_l are taken from the ideal $I_{\mathcal{X}} \subset \mathcal{O}[x_1, \dots, x_{d+l}]$ defining \mathcal{X} and $1 \leq i_1 < \dots < i_l \leq d+l$.

Let $\gamma \in \mathcal{X}_{\infty}$ be such that $\gamma^*\mathcal{J}(\mathcal{X}/\mathbb{D})$ has finite order e . This implies that γ maps \mathbb{D}^* to the part $(\mathcal{X}/\mathbb{D})_{\text{reg}}$ where \mathcal{X} is smooth over \mathbb{D} . In particular, $\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}$ is a \mathcal{O} -module of rank d . Since the formation of a Fitting ideal commutes with base change, the d th Fitting ideal of $\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}$ will be \mathfrak{m}^e . This means that the torsion of $\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}$ has length e .

It is clear that $\text{Der}_{\mathcal{O}}(\mathcal{O}_{\mathcal{X},\gamma(o)}, \mathcal{O}) \cong \text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O})$ is a free \mathcal{O} -module of rank d (where \mathcal{O} is a $\mathcal{O}_{\mathcal{X},\gamma(o)}$ -module via γ^*). The fiber over o , $\text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O}) \otimes_{\mathcal{O}} k$, is d -dimensional subspace of the Zariski tangent space $T_{\mathcal{X},\gamma(o)}$, which we shall denote by $\hat{T}_{\mathcal{X},\gamma}$. Any \mathcal{O} -homomorphism $\gamma^*\Omega_{\mathcal{X}/\mathbb{D}} \rightarrow \mathcal{O}/\mathfrak{m}^{n+1}$ that kills the torsion lifts to a \mathcal{O} -homomorphism $\gamma^*\Omega_{\mathcal{X}/\mathbb{D}} \rightarrow \mathcal{O}$. This is automatic when $n \geq e$ and so $\hat{T}_{\mathcal{X},\gamma}$ only depends on the e -jet of γ . The space $\hat{T}_{\mathcal{X},\gamma}$ has a simple geometric interpretation: it is the ‘limiting position’ of the tangent space along the fibers of \mathcal{X}/\mathbb{D} at the generic point of $\gamma(\mathbb{D})$ in the closed point $\gamma(o)$.

If $\gamma' : \mathbb{D} \rightarrow \mathcal{X}$ has the same n -jet as γ , then γ^* and γ'^* differ by a homomorphism $\mathcal{O}_{\mathcal{X},\gamma(o)} \rightarrow \mathfrak{m}^{n+1}$. The reduction modulo $\mathfrak{m}^{2(n+1)}$ of this homomorphism is a \mathcal{O} -derivation, i.e., defines an element of $\text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathfrak{m}^{n+1}/\mathfrak{m}^{2(n+1)})$. Its reduction modulo \mathfrak{m}^{n+2} will lie in $\hat{T}_{\mathcal{X},\gamma} \otimes \mathfrak{m}^{n+1}/\mathfrak{m}^{n+2}$, provided that $n \geq e$. The next lemma shows that every element of this k -vector space so arises.

LEMMA 9.1. — Assume that $n \geq e$. The fiber of $\pi_{n+1}\mathcal{X}_{\infty} \rightarrow \pi_n\mathcal{X}_{\infty}$ over $\pi_n(\gamma)$ is an affine space with translation space $\hat{T}_{\mathcal{X},\gamma} \otimes_k \mathfrak{m}^{n+1}/\mathfrak{m}^{n+2}$. This defines an affine space bundle of rank d over the locus of $\pi_n\mathcal{X}_{\infty}$ defined by $\text{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})} \leq n$.

Proof. — Assume that \mathcal{X} is given as a closed subset of $(\mathbb{A}^{d+l})_{\mathbb{D}}$ as above. There exist $f_1, \dots, f_l \in I_{\mathcal{X}}$ and $1 \leq i_1 < \dots < i_l \leq d+l$ such that the Jacobian matrix $\det((\partial f_j/\partial x_{i_k})_{j,k=1}^l)$ has order e along γ , whereas for any other matrix thus formed the order is $\geq e$. By means of a coordinate change we may arrange that

$$\gamma^*df_j \equiv t^{e_j}dx_j \pmod{t^{e_j+1}(dx_{j+1}, \dots, dx_{d+l})}, \quad j = 1, \dots, l,$$

so that $e = \sum_j e_j$. The subspace of \mathbb{A}_k^{d+l} spanned by the last d basis vectors is then just $\hat{T}_{\mathcal{X},\gamma}$.

We investigate which $u_0 \in k^{d+l}$ appear as the constant coefficient of an $u \in \mathcal{O}^{d+l}$ with the property that $\gamma + t^{n+1}u \in \mathcal{X}_{\infty}$. We first do this for the complete intersection defined by f_1, \dots, f_l . This complete intersection contains \mathcal{X} and the irreducible component that

contains the image of γ lies in \mathcal{X} . So we want $f_j(\gamma + t^{n+1}u) = 0$ for $j = 1, \dots, l$. By expanding at γ this amounts to identities of the form

$$t^{n+1}D_\gamma f_j(u) + t^{2(n+1)}F_j(u) = 0, \quad j = 1, \dots, l,$$

with $D_\gamma f_j$ the derivative of f_j at γ and $F_j \in \mathcal{O}[x_1, \dots, x_{d+l}]$. Equivalently:

$$t^{-e_j}D_\gamma f_j(u) + t^{n+1-e_j}F_j(u) = 0, \quad j = 1, \dots, l.$$

All the terms are regular and the reduction modulo t yields the j th unit vector in k^{d+l} . Hensel's lemma says that a solution u exists if and only if u_0 solves this set of equations modulo t . This just means that $u_0 \in \hat{T}_{X,\gamma}$. In particular, we see that for all $k \in \mathbb{N}$, $\pi_{n+k}\pi_n^{-1}\pi_n(\gamma)$ is isomorphic to an affine space and hence is irreducible. This implies that all elements of $\pi_n^{-1}\pi_n(\gamma)$ map to the same irreducible component of the common zero locus of f_1, \dots, f_l . It follows that $\pi_n^{-1}\pi_n(\gamma) \subset \mathcal{X}_\infty$. The last assertion is easy. \square

Proof of Proposition 3.1. — Suppose that \mathcal{X} is of pure relative dimension d . Let C_e denote the subset of \mathcal{X}_∞ defined by $\text{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})} = e$. It is clear that $C_e = \pi_e^{-1}\pi_e(C_e)$. It follows from Greenberg's theorem [20] that $\pi_e(C_e)$ is constructible. Hence C_e is stable by Lemma 9.1. We have $\cup_e C_e = \mathcal{X}_\infty - (\mathcal{X}_{\text{sing}})_\infty$. In view of Lemma 2.3 it now suffices to see that $\dim \pi_e(C_e) - de \rightarrow -\infty$ as $e \rightarrow \infty$. This is not difficult. \square

Let $H : \mathcal{Y} \rightarrow \mathcal{X}$ be a \mathbb{D} -morphism of \mathbb{D} -varieties of pure relative dimension d . Recall that the Jacobian ideal \mathcal{J}_H of H is the 0th Fitting ideal of $\Omega_{\mathcal{Y}/\mathcal{X}}$. Suppose $\gamma \in \mathcal{Y}_\infty$ is such that \mathcal{J}_H has finite order e along γ . Then γ resp. $H\gamma$ maps the generic point \mathbb{D}^* to $(\mathcal{Y}/\mathbb{D})_{\text{reg}}$ resp. $(\mathcal{X}/\mathbb{D})_{\text{reg}}$. We have an exact sequence of \mathcal{O} -modules

$$(H\gamma)^*\Omega_{\mathcal{X}/\mathbb{D}} \rightarrow \gamma^*\Omega_{\mathcal{Y}/\mathbb{D}} \rightarrow \gamma^*\Omega_{\mathcal{Y}/\mathcal{X}} \rightarrow 0.$$

The base change property of Fitting ideals implies that the length of $\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}$ must be e . So if $\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}$ is torsion free and $n \geq e$, then the kernel of the map

$$D_\gamma^{(n)} : \text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathcal{O}/\mathfrak{m}^{n+1}) \rightarrow \text{Hom}_{\mathcal{O}}((\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O}/\mathfrak{m}^{n+1})$$

induced by the derivative of H is contained in $\text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathfrak{m}^{n+1-e}/\mathfrak{m}^{n+1})$, can be identified with $\text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}, \mathcal{O}/\mathfrak{m}^{n+1})$, and is of length e . The proof of Theorem 3.2 now rests on the

KEY LEMMA 9.2. — Suppose \mathcal{Y}/\mathbb{D} smooth and let $A \subset \mathcal{Y}_\infty$ be a stable subset of level l : $A = \pi_l^{-1}\pi_l(A)$. Assume that $H|_A$ is injective and that $\text{ord}_{\mathcal{J}_H}|_A$ is constant equal to $e < \infty$. If $n \geq \sup\{2e, l + e, \text{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})}|_{HA}\}$, then $H_n : \pi_n A \rightarrow H_n \pi_n A$ has the structure of affine-linear bundle of dimension e .

Proof. — Let $\gamma \in A$ and put $x := \gamma(o)$, $y := H(x)$. Suppose $\gamma' \in A$ is such that $H\gamma'$ and $H\gamma$ have the same n -jet. We first show that γ and γ' have the same $(n - e)$ -jet. We do this by constructing a $\gamma_1 \in \mathcal{Y}_\infty$ (by successive approximation) with the same $(n - e)$ -jet

as γ and with $H\gamma_1 = H\gamma'$. Since $n - e \geq l$, we will have $\gamma_1 \in A$ and our injectivity assumption then implies $\gamma_1 = \gamma'$.

The difference $(H\gamma)^* - (H\gamma')^*$ defines a \mathcal{O} -derivation $\mathcal{O}_{\mathcal{X},x} \rightarrow \mathfrak{m}^{n+1}/\mathfrak{m}^{2(n+1)}$ over γ^* and hence a $\tilde{v} \in \text{Hom}_{\mathcal{O}}((\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathfrak{m}^{n+1}/\mathfrak{m}^{2n+2})$. Since $n \geq \text{ord}_{H\gamma} \mathcal{J}(\mathcal{X}/\mathbb{D})$, this element annihilates the torsion of $(\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}$. This is then also true for its reduction modulo \mathfrak{m}^{n+2} and it follows from the fact that $n \geq e$ that this reduction is of the form $D_\gamma^{(n+1)}(u)$ for some $u \in \text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+2})$. Regard u as a \mathcal{O} -derivation $\mathcal{O}_{\mathcal{Y},y} \rightarrow \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+2}$ and let $\gamma_1 \in \mathcal{Y}_\infty$ be such that $\gamma_1^* - \gamma^*$ represents u . Then $\pi_{n-e}(\gamma_1) = \pi_{n-e}(\gamma)$ and $\pi_{n+1}(H\gamma_1) = \pi_{n+1}(H\gamma)$. Replace γ by γ_1 and continue with induction on n .

So $(\gamma')^* - \gamma^*$ defines a \mathcal{O} -derivation $\mathcal{O}_{\mathcal{X},x} \rightarrow \mathfrak{m}^{n-e+1}/\mathfrak{m}^{2(n-e+1)}$ and hence a \mathcal{O} -derivation $\mathcal{O}_{\mathcal{X},x} \rightarrow \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+1}$ (because $n \geq 2e$). The latter is zero if and only if $\pi_n(\gamma') = \pi_n(\gamma)$. This proves that the fiber of $H_n|_{\pi_n A}$ through $\pi_n(\gamma)$ is an affine space over the kernel of $D_\gamma^{(n)}$, $\text{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}, \mathcal{O}/\mathfrak{m}^{n+1})$, which has length e .

The last assertion is easy. □

Proof of 3.2. — It is enough to prove this for A stable. In that case the theorem follows in a straightforward manner from Lemma 9.2. □

REFERENCES

- [1] D. Abramovich, K. Karu, K. Matsuki, J. Włodarczyk, *Torification and Factorization of Birational Maps*, available at math.AG/9904135.
- [2] N. A'Campo, *La fonction zêta d'une monodromie*, Comment. Math. Helv. **50** (1975), 233–248.
- [3] G. Anderson, *Cyclotomy and an extension of the Taniyama group*, Compositio Mathematica **57** (1986), 153–217.
- [4] V. Batyrev, *Birational Calabi–Yau n -folds have equal Betti numbers*, in New trends in algebraic geometry, Klaus Hulek *et al.*, eds., CUP, 1999, 1–11.
- [5] V. Batyrev, *Stringy Hodge numbers of varieties with Gorenstein canonical singularities*, in Integrable systems and algebraic geometry (Kobe/Kyoto, 1997), 1–32, World Sci. Publishing, River Edge, NJ, 1998.
- [6] V. Batyrev, *Non-Archimedean integrals and stringy Euler numbers of log-terminal pairs*, J. Eur. Math. Soc. **1** (1999), 5–33.
- [7] V. Batyrev and D. Dais, *Strong McKay correspondence, string-theoretic Hodge numbers and mirror symmetry*, Topology **35** (1996), 901–929.
- [8] A. Craw, *An introduction to motivic integration*, preliminary draft available at www.maths.warwick.ac.uk~craw, 23 pp. (1999).

- [9] P. Deligne, *Intégration sur un cycle évanescant*. Invent. Math. **76** (1984), no. 1, 129–143.
- [10] J. Denef, *Local zeta functions and Euler characteristics*. Duke Math. J. **63** (1991), no. 3, 713–721.
- [11] J. Denef, *Report on Igusa’s local zeta function*, in *Séminaire Bourbaki, volume 1990/91*, exposé 741. Astérisque **201-202-203** (1991), 359–386.
- [12] J. Denef, F. Loeser, *Caractéristiques d’Euler-Poincaré, fonctions zêta locales et modifications analytiques*, J. Amer. Math. Soc. **5** (1992), 705–720.
- [13] J. Denef, F. Loeser, *Motivic Igusa zeta functions*, J. Algebraic Geom., **7** (1998), 505–537.
- [14] J. Denef, F. Loeser, *Germes of arcs on singular algebraic varieties and motivic integration*, Invent. Math., **135** (1999), 201–232.
- [15] J. Denef, F. Loeser, *Motivic exponential integrals and a motivic Thom-Sebastiani Theorem*, Duke Math. J., **99** (1999), 285–309.
- [16] J. Denef, F. Loeser, *Motivic integration, quotient singularities and the McKay correspondence*, 20 p., available at math.AG/9903187.
- [17] J. Denef, F. Loeser, *Definable sets, motives and P-adic integrals*, 45 p., available at math.AG/9910107.
- [18] J. Denef, F. Loeser, *Lefschetz numbers of the monodromy and truncated arcs*, 10 p., available at math.AG/0001105.
- [19] M. Greenberg, *Schemata over local rings*, Ann. Math. **73** (1961), 624–648.
- [20] M. Greenberg, *Rational points in discrete valuation rings*, Publ. Math. I.H.E.S. **31** (1966), 59–64.
- [21] M. Kapranov, *The elliptic curve in the S-duality theory and Eisenstein series for Kac-Moody groups*, available at math.AG/0001005.
- [22] M. Kontsevich, Lecture at Orsay (December 7, 1995).
- [23] J. Nash Jr., *Arc structure of singularities*, Duke Math. J., **81** (1995), 31–38.
- [24] J. Pas, *Uniform p-adic cell decomposition and local zeta functions*, J. f. d. reine u. angew. Math. **399** (1989), 137–172.
- [25] M. Reid, *La correspondance de McKay*, in *Séminaire Bourbaki*, exposé 867, Novembre 1999, 20 p., available at math.AG/9911165 (en anglais).
- [26] J. Steenbrink, *The spectrum of hypersurface singularities*, in *Théorie de Hodge, Luminy 1987*, Astérisque **179-180** (1989), 163–184.
- [27] A. Varchenko, *Asymptotic Hodge structure in the vanishing cohomology*, Math. USSR Izvestija **18** (1982), 469–512.
- [28] W. Veys, *The topological zeta function associated to a function on a normal surface germ*, Topology **38** (1999), no. 2, 439–456.

- [29] W. Veys, *Zeta functions and 'Kontsevich invariants' on singular varieties*, available at math.AG/0003025.

Eduard LOOIJENGA

Adresse courante : Mathematisch Instituut

Universiteit Utrecht

Postal Box 80.010, NL-3508 TA Utrecht

Nederland

E-mail : looijeng@math.uu.nl

VERTEX ALGEBRAS AND ALGEBRAIC CURVES

by **Edward FRENKEL**

1. INTRODUCTION

Vertex operators appeared in the early days of string theory as local operators describing propagation of string states. Mathematical analogues of these operators were discovered in representation theory of affine Kac-Moody algebras in the works of Lepowsky–Wilson [LW] and I. Frenkel–Kac [FK]. In order to formalize the emerging structure and motivated in particular by the I. Frenkel–Lepowsky–Meurman construction of the Moonshine Module of the Monster group, Borchers gave the definition of vertex algebra in [B1]. The foundations of the theory were subsequently laid down in [FLM, FHL]; in particular, it was shown in [FLM] that the Moonshine Module indeed possessed a vertex algebra structure. In the meantime, Belavin, Polyakov and Zamolodchikov [BPZ] initiated the study of two-dimensional conformal field theory (CFT). Vertex algebras can be seen in retrospect as the mathematical equivalent of the central objects of CFT called the chiral symmetry algebras. Moreover, the key property of associativity of vertex algebras is equivalent to the property of operator product expansion in CFT, which goes back to the pioneering works of Polyakov and Wilson. Thus, vertex algebras may be thought of as the mathematical language of two-dimensional conformal field theory.

Vertex algebras have turned out to be extremely useful in many areas of mathematics. They are by now ubiquitous in representation theory of infinite-dimensional Lie algebras. They have also found applications in such fields as algebraic geometry, theory of finite groups, modular functions and topology. Recently Beilinson and Drinfeld have introduced a remarkable geometric version of vertex algebras which they called chiral algebras [BD3]. Chiral algebras give rise to some novel concepts and techniques which are likely to have a profound impact on algebraic geometry.

In this talk we review the theory of vertex algebras with a particular emphasis on their algebro-geometric interpretation and applications. We start in §2 with the axiomatic definition of vertex algebra, which is somewhat different from, but equivalent to Borchers’

original definition (see [FKRW, K2]). We then discuss some of their most important properties and give the first examples, which come from infinite-dimensional Lie algebras, such as Heisenberg, affine Kac-Moody and the Virasoro algebra. More unconventional examples of vertex algebras which are not generated by a Lie algebra, such as the \mathcal{W} -algebras, are reviewed in §3.

In §4 we explain how to make vertex algebras coordinate independent, thus effectively getting rid of the formal variable omnipresent in the standard algebraic approach. This is achieved by attaching to each conformal vertex algebra a vector bundle with a flat connection on a formal disc, equipped with an intrinsic operation. The formal variable is restored when we choose a coordinate on the disc; the independence of the operation from the choice of coordinate follows from the fact that the group of changes of coordinates is an “internal symmetry” of the vertex algebra. Once we obtain a coordinate independent object, we can give a rigorous definition of the space of conformal blocks associated to a conformal vertex algebra and an algebraic curve X (see §5). From the physics point of view, conformal blocks give rise to “chiral correlation functions” on powers of X with singularities along the diagonals.

As we vary the curve X and other data on X reflecting the internal symmetry of a vertex algebra (such as G -bundles), the corresponding spaces of coinvariants, which are the duals of the spaces of conformal blocks, combine into a sheaf on the relevant moduli space. This sheaf carries the structure of a (twisted) \mathcal{D} -module, as explained in §6. One can gain new insights into the structure of moduli spaces from the study of these \mathcal{D} -modules. For instance, one obtains a description of the formal deformation spaces of the complex structure or a G -bundle on a curve in terms of certain sheaves on the symmetric powers of the curve [BG, BD2, Gi]. Thus, vertex algebras appear as the local objects controlling deformations of curves with various extra structures. This raises the possibility that more exotic vertex algebras, such as \mathcal{W} -algebras, also correspond to some still unknown moduli spaces.

Finally, in §7 we discuss the relation between vertex algebras and the Beilinson–Drinfeld chiral algebras. We review briefly the description of chiral algebras as factorization algebras, i.e., sheaves on the Ran space of finite subsets of a curve, satisfying certain compatibilities. Using this description, Beilinson and Drinfeld have introduced the concept of chiral homology, which can be thought of as a derived functor of the functor of coinvariants.

The formalism of vertex and chiral algebras appears to be particularly suitable for the construction of the conjectural geometric Langlands correspondence between automorphic \mathcal{D} -modules on the moduli space of G -bundles on a smooth projective curve X over \mathbb{C} and flat ${}^L G$ -bundles on X , where G is a simple algebraic group and ${}^L G$ is the Langlands dual group (see [BD2]). The idea is to construct the automorphic \mathcal{D} -modules corresponding to flat ${}^L G$ -bundles as the sheaves of coinvariants (or, more generally, chiral homology) of a

suitable vertex algebra. We mention below two such constructions (both due to Beilinson and Drinfeld): in one of them the relevant vertex algebra is associated to an affine Kac-Moody algebra of critical level (see §6.5), and in the other it is the chiral Hecke algebra (see §7).

Another application of vertex algebras to algebraic geometry is the recent construction by Malikov, Schechtman and Vaintrob [MSV] of a sheaf of vertex superalgebras on an arbitrary smooth algebraic variety, called the chiral deRham complex, which is reviewed in §6.6.

Vertex algebras form a vast and rapidly growing subject, and it is impossible to cover all major results (or even give a comprehensive bibliography) in one survey. For example, because of lack of space, I have not discussed such important topics as the theory of conformal algebras [K2, K3] and their chiral counterpart, Lie* algebras [BD2]; quantum deformations of vertex algebras [B3, EK, FR]; and the connection between vertex algebras and integrable systems.

Most of the material presented below (note that §§4–6 contain previously unpublished results) is developed in the forthcoming book [BF].

I thank A. Beilinson for answering my questions about chiral algebras and D. Ben-Zvi for helpful comments on the draft of this paper. The support from the Packard Foundation and the NSF is gratefully acknowledged.

2. DEFINITION AND FIRST PROPERTIES OF VERTEX ALGEBRAS

2.1. Let R be a \mathbb{C} -algebra. An R -valued *formal power series* (or formal distribution) in variables z_1, z_2, \dots, z_n is an arbitrary infinite series

$$A(z_1, \dots, z_n) = \sum_{i_1 \in \mathbb{Z}} \cdots \sum_{i_n \in \mathbb{Z}} a_{i_1, \dots, i_n} z_1^{i_1} \cdots z_n^{i_n},$$

where each $a_{i_1, \dots, i_n} \in R$. These series form a vector space denoted by $R[[z_1^{\pm 1}, \dots, z_n^{\pm 1}]]$. If $P(z_1, \dots, z_n) \in R[[z_1^{\pm 1}, \dots, z_n^{\pm 1}]]$ and $Q(w_1, \dots, w_m) \in R[[w_1^{\pm 1}, \dots, w_m^{\pm 1}]]$, then their product is a well-defined element of $R[[z_1^{\pm 1}, \dots, z_n^{\pm 1}, w_1^{\pm 1}, \dots, w_m^{\pm 1}]]$. In general, a product of two formal power series in the same variables does not make sense, but the product of a formal power series by a polynomial (i.e., a series, such that $a_{i_1, \dots, i_n} = 0$ for all but finitely many indices) is always well-defined.

Let V be a \mathbb{Z}_+ -graded vector space $V = \bigoplus_{n=0}^{\infty} V_n$ with finite-dimensional homogeneous components. An endomorphism T of V is called homogeneous of degree n if $T(V_m) \subset V_{n+m}$. Denote by $\text{End } V$ the vector space of linear endomorphisms of V , which are finite linear combinations of homogeneous endomorphisms. This is a \mathbb{Z} -graded algebra.

A *field* of conformal dimension $\Delta \in \mathbb{Z}_+$ is an $\text{End } V$ -valued formal power series in z ,

$$\phi(z) = \sum_{j \in \mathbb{Z}} \phi_j z^{-j-\Delta}$$

where each ϕ_j is a homogeneous linear endomorphism of V of degree $-j$. Two fields $\phi(z)$ and $\psi(z)$ are called mutually *local* if there exists $N \in \mathbb{Z}_+$, such that

$$(1) \quad (z-w)^N[\phi(z), \psi(w)] = 0$$

(as an element of $\text{End } V[[z^{\pm 1}, w^{\pm 1}]]$).

Now we can formulate the axioms of vertex algebra.

DEFINITION 2.1. — *A vertex algebra is a collection of data:*

- space of states, a \mathbb{Z}_+ -graded vector space $V = \bigoplus_{n=0}^{\infty} V_n$, with $\dim(V_n) < \infty$;
- vacuum vector $|0\rangle \in V_0$;
- shift operator $T : V \rightarrow V$ of degree one;
- vertex operation, a linear map $Y(\cdot, z) : V \rightarrow \text{End } V[[z, z^{-1}]]$ taking each $A \in V_n$ to a field of conformal dimension n .

These data are subject to the following axioms:

- (vacuum axiom) $Y(|0\rangle, z) = \text{Id}_V$. Furthermore, for any $A \in V$ we have $Y(A, z)|0\rangle \in A + zV[[z]]$ (i.e., $Y(A, z)|0\rangle$ has a well-defined value at $z = 0$, which is equal to A).
- (translation axiom) For any $A \in V$, $[T, Y(A, z)] = \partial_z Y(A, z)$ and $T|0\rangle = 0$.
- (locality axiom) All fields $Y(A, z)$ are mutually local with each other.

REMARK 2.2. — It is easy to adopt the above definition to the supercase (see, e.g., [K2]). Then V is a superspace, and the above structures and axioms should be modified appropriately; in particular, we need to replace the commutator by the supercommutator in the definition of locality. Then we obtain the definition of vertex superalgebra. We mostly consider below purely even vertex superalgebras, but general vertex superalgebras are very important; for instance, $N = 2$ superconformal vertex superalgebras appear in physical models relevant to mirror symmetry.

The above conditions on V can be relaxed: it suffices to require that for any $A \in V$, $v \in V$, we have: $A_n \cdot v = 0$ for n large enough. It is not necessary to require that $\dim V_n < \infty$ and even that V is graded (in that case however one needs to be careful when dealing with dual spaces). We impose the above stronger conditions in order to simplify the exposition.

It is straightforward to define homomorphisms between vertex algebras, vertex subalgebras, ideals and quotients. If V_1 and V_2 are two vertex algebras, then $V_1 \otimes V_2$ carries a natural vertex algebra structure.

2.2. Example: commutative vertex algebras

Let V be a \mathbb{Z} -graded commutative algebra (with finite-dimensional homogeneous components) with a unit and a derivation T of degree 1. Then V carries a canonical structure

of vertex algebra. Namely, we take the unit of V as the vacuum vector $|0\rangle$, and define the operation Y by the formula

$$Y(A, z) = \sum_{n \geq 0} \frac{1}{n!} m(T^n A) z^n,$$

where for $B \in V$, $m(B)$ denotes the operator of multiplication by B on V . It is straightforward to check that all axioms of vertex algebra are satisfied; in fact, the locality axiom is satisfied in a strong sense: for any $A, B \in V$, we have: $[Y(A, z), Y(B, w)] = 0$ (so we have $N = 0$ in formula (1)).

Conversely, let V be a vertex algebra, in which locality holds in the strong sense (we call such a vertex algebra *commutative*). Then locality and vacuum axioms imply that $Y(A, z) \in \text{End } V[[z]]$ for all $A \in V$ (i.e., there are no terms with negative powers of z). Denote by Y_A the endomorphism of V , which is the constant term of $Y(A, z)$, and define a bilinear operation \circ on V by setting $A \circ B = Y_A \cdot B$. By construction, $Y_A Y_B = Y_B Y_A$ for all $A, B \in V$. This implies commutativity and associativity of \circ (see, e.g., [Li]). We obtain a commutative and associative product on V , which respects the \mathbb{Z} -gradation. Furthermore, the vacuum vector $|0\rangle$ is a unit, and the operator T is a derivation with respect to this product. Thus, we see that the notion of commutative vertex algebra is equivalent to that of \mathbb{Z} -graded commutative associative algebra with a unit and a derivation of degree 1.

REMARK 2.3. — The operator T may be viewed as the generator of infinitesimal translations on the formal additive group with coordinate z . Therefore a commutative vertex algebra is the same as a commutative algebra equipped with an action of the formal additive group. Thus one may think of general vertex algebras as meromorphic generalizations of commutative algebras with an action of the formal additive group. This point of view has been developed by Borchers [B3], who showed that vertex algebras are “singular commutative rings” in a certain category. He has also considered generalizations of vertex algebras, replacing the formal additive group by other (formal) groups or Hopf algebras.

2.3. Non-commutative example: Heisenberg vertex algebra

Consider the space $\mathbb{C}((t))$ of Laurent series in one variable as a commutative Lie algebra. We define the Heisenberg Lie algebra \mathcal{H} as follows. As a vector space, it is the direct sum of the space of formal Laurent power series $\mathbb{C}((t))$ and a one-dimensional space $\mathbb{C}\mathbf{1}$, with the commutation relations

$$(2) \quad [f(t), g(t)] = -\text{Res } fdg\mathbf{1}, \quad [\mathbf{1}, f(t)] = 0.$$

Here Res denotes the (-1) st Fourier coefficient of a Laurent series. Thus, \mathcal{H} is a one-dimensional central extension of the commutative Lie algebra $\mathbb{C}((t))$. Note that the relations (2) are independent of the choice of local coordinate t . Thus we may define a

Heisenberg Lie algebra canonically as a central extension of the space of functions on a punctured disc without a specific choice of formal coordinate t .

The Heisenberg Lie algebra \mathcal{H} is topologically generated by the generators $b_n = t^n$, $n \in \mathbb{Z}$, and the central element $\mathbf{1}$, and the relations between them read

$$(3) \quad [b_n, b_m] = n\delta_{n,-m}\mathbf{1}, \quad [\mathbf{1}, b_n] = 0.$$

The subspace $\mathbb{C}[[t]] \oplus \mathbf{1}$ is a commutative Lie subalgebra of \mathcal{H} . Let π be the \mathcal{H} -module induced from the one-dimensional representation of $\mathbb{C}[[t]] \oplus \mathbf{1}$, on which $\mathbb{C}[[t]]$ acts by 0 and $\mathbf{1}$ acts by 1. Equivalently, we may describe π as the polynomial algebra $\mathbb{C}[b_{-1}, b_{-2}, \dots]$ with $b_n, n < 0$, acting by multiplication, b_0 acting by 0, and $b_n, n > 0$, acting as $n \frac{\partial}{\partial b_{-n}}$. The operators b_n with $n < 0$ are known in this context as creation operators, since they “create the state b_n from the vacuum 1 ”, while the operators b_n with $n \geq 0$ are the annihilation operators, repeated application of which will kill any vector in π . The module π is called the Fock representation of \mathcal{H} .

We wish to endow π with a structure of vertex algebra. This involves the following data (Definition 2.1):

- \mathbb{Z}_+ grading: $\deg b_{j_1} \dots b_{j_k} = -\sum_i j_i$.
- Vacuum vector: $|0\rangle = 1$.
- The shift operator T defined by the rules: $T \cdot 1 = 0$ and $[T, b_i] = -ib_{i-1}$.

We now need to define the fields $Y(A, z)$. To the vacuum vector $|0\rangle = 1$, we are required to assign $Y(|0\rangle, z) = \text{Id}$. The key definition is that of the field $b(z) = Y(b_{-1}, z)$, which generates π in an appropriate sense. Since $\deg(b_{-1}) = 1$, $b(z)$ needs to have conformal dimension one. We set

$$b(z) = \sum_{n \in \mathbb{Z}} b_n z^{-n-1}.$$

The Reconstruction Theorem stated below implies that once we have defined $Y(b_{-1}, z)$, there is (at most) a unique way to extend this definition to other vectors of π . This is not very surprising, since as a commutative algebra with derivation T , π is freely generated by b_{-1} . Explicitly, the fields corresponding to other elements of π are constructed by the formula

$$(4) \quad Y(b_{-n_1} b_{-n_2} \dots b_{-n_k}, z) = \frac{1}{(n_1 - 1)! \dots (n_k - 1)!} : \partial_z^{n_1-1} b(z) \dots \partial_z^{n_k-1} b(z) :.$$

The columns in the right hand side of the formula stand for the *normally ordered product*, which is defined as follows. First, let $: b_{n_1} \dots b_{n_k} :$ be the monomial obtained from $b_{n_1} \dots b_{n_k}$ by moving all b_{n_i} with $n_i < 0$ to the left of all b_{n_j} with $n_j \geq 0$ (in other words, moving all “creation operators” $b_n, n < 0$, to the left and all “annihilation operators” $b_n, n \geq 0$, to the right). The important fact that makes this definition correct is that the operators b_n with $n < 0$ (resp., $n \geq 0$) commute with each other, hence it does not matter how we order the creation (resp., annihilation) operators among themselves (this

property does not hold for more general vertex algebras, and in those cases one needs to use a different definition of normally ordered product, which is given below). Now define $:\partial_z^{m_1}b(z)\dots\partial_z^{m_k}b(z):$ as the power series in z obtained from the ordinary product $\partial_z^{m_1}b(z)\dots\partial_z^{m_k}b(z)$ by replacing each term $b_{n_1}\dots b_{n_k}$ with $:b_{n_1}\dots b_{n_k}:$.

PROPOSITION 2.4. — *The Fock representation π with the structure given above satisfies the axioms of vertex algebra.*

The most difficult axiom to check is locality. The Reconstruction Theorem stated below allows us to reduce it to verifying locality property for the “generating” field $b(z)$. This is straightforward from the commutation relations (3):

$$\begin{aligned} [b(z), b(w)] &= \sum_{n,m \in \mathbb{Z}} [b_n, b_m] z^{-n-1} w^{-m-1} = \sum_{n \in \mathbb{Z}} [b_n, b_{-n}] z^{-n-1} w^{n-1} \\ &= \sum_{n \in \mathbb{Z}} n z^{-n-1} w^{n-1} = \partial_w \delta(z - w), \end{aligned}$$

where we use the notation $\delta(z - w) = \sum_{m \in \mathbb{Z}} w^m z^{-m-1}$.

Now $\delta(z - w)$ has the property that $(z - w)\delta(z - w) = 0$. In fact, the annihilator of the operator of multiplication by $(z - w)$ in $R[[z^{\pm 1}, w^{\pm 1}]]$ equals $R[[w^{\pm 1}]] \cdot \delta(z - w)$ (note that the product of $\delta(z - w)$ with any formal power series in z or in w is well-defined). More generally, the annihilator of the operator of multiplication by $(z - w)^N$ in $R[[z^{\pm 1}, w^{\pm 1}]]$ equals $\oplus_{n=0}^{N-1} R[[w^{\pm 1}]] \cdot \partial_w^n \delta(z - w)$ (see [K1]). This implies in particular that $(z - w)^2 [b(z), b(w)] = 0$, and hence the field $b(z)$ is local with itself. Proposition 2.4 now follows from the Reconstruction Theorem below.

2.4. Reconstruction Theorem

We state a general result, which provides a “generators–and–relations” approach to the construction of vertex algebras. Let V be a \mathbb{Z}_+ -graded vector space, $|0\rangle \in V_0$ a non-zero vector, and T a degree 1 endomorphism of V . Let S be a countable ordered set and $\{a^\alpha\}_{\alpha \in S}$ be a collection of homogeneous vectors in V , with a^α of degree Δ_α . Suppose we are also given fields

$$a^\alpha(z) = \sum_{n \in \mathbb{Z}} a_n^\alpha z^{-n-\Delta_\alpha}$$

on V , such that the following hold:

- For all α , $a^\alpha(z)|0\rangle \in a^\alpha + zV[[z]]$;
- $T|0\rangle = 0$ and $[T, a^\alpha(z)] = \partial_z a^\alpha(z)$ for all α ;
- All fields $a^\alpha(z)$ are mutually local;
- V is spanned by lexicographically ordered monomials

$$a_{-\Delta_{\alpha_1}-j_1}^{\alpha_1} \cdots a_{-\Delta_{\alpha_m}-j_m}^{\alpha_m} |0\rangle$$

(including $|0\rangle$), where $j_1 \geq j_2 \geq \dots \geq j_m \geq 0$, and if $j_i = j_{i+1}$, then $\alpha_i \geq \alpha_{i+1}$ with respect to the order on the set S .

THEOREM 2.5 ([FKRW]). — *Under the above assumptions, the above data together with the assignment*

$$(5) \quad Y(a_{-\Delta_{\alpha_1}-j_1}^{\alpha_1} \cdots a_{-\Delta_{\alpha_m}-j_m}^{\alpha_m} |0\rangle, z) = \frac{1}{j_1! \cdots j_m!} : \partial_z^{j_1} a^{\alpha_1}(z) \cdots \partial_z^{j_m} a^{\alpha_m}(z) :,$$

define a vertex algebra structure on V .

Here we use the following general definition of the normally ordered product of fields. Let $\phi(z), \psi(w)$ be two fields of respective conformal dimensions Δ_ϕ, Δ_ψ and Fourier coefficients ϕ_n, ψ_n . The normally ordered product of $\phi(z)$ and $\psi(z)$ is by definition the formal power series

$$:\phi(z)\psi(z): = \sum_{n \in \mathbb{Z}} \left(\sum_{m \leq -\Delta_\phi} \phi_m \psi_{n-m} + \sum_{m > -\Delta_\phi} \psi_{n-m} \phi_m \right) z^{-n-\Delta_\phi-\Delta_\psi}.$$

This is a field of conformal dimension $\Delta_\phi + \Delta_\psi$. The normal ordering of more than two fields is defined recursively from right to left, so that by definition $:A(z)B(z)C(z): = :A(z)(:B(z)C(z):):$. It is easy to see that in the case of the Heisenberg vertex algebra π this definition of normal ordering coincides with the one given in §2.3.

2.5. The meaning of locality

The product $\phi(z)\psi(w)$ of two fields is a well-defined $\text{End } V$ -valued formal power series in $z^{\pm 1}$ and $w^{\pm 1}$. Given $v \in V$ and $v^\vee \in V^\vee = \bigoplus_{n \geq 0} V_n^*$, consider the matrix coefficient

$$\langle v^\vee | \phi(z)\psi(w) | v \rangle \in \mathbb{C}[[z^{\pm 1}, w^{\pm 1}]].$$

Since $V_n = 0$ for $n < 0$, we find by degree consideration that it belongs to $\mathbb{C}((z))((w))$, the space of formal Laurent series in w , whose coefficients are formal Laurent series in z . Likewise, $\langle v^\vee | \psi(w)\phi(z) | v \rangle$ belongs to $\mathbb{C}((w))((z))$.

As we have seen in §2.2, the condition that the fields $\phi(z)$ and $\psi(w)$ literally commute is too strong, and it essentially keeps us in the realm of commutative algebra. However, there is a natural way to relax this condition, which leads to the more general notion of locality. Let $\mathbb{C}((z, w))$ be the field of fractions of the ring $\mathbb{C}[[z, w]]$; its elements may be viewed as meromorphic functions in two formal variables. We have natural embeddings

$$(6) \quad \mathbb{C}((z))((w)) \longleftarrow \mathbb{C}((z, w)) \longrightarrow \mathbb{C}((w))((z)),$$

which are simply the inclusions of $\mathbb{C}((z, w))$ into its completions in two different topologies, corresponding to the z and w axes. For example, the images of $1/(z-w) \in \mathbb{C}((z, w))$ in $\mathbb{C}((z))((w))$ and in $\mathbb{C}((w))((z))$ are equal to

$$\delta(z-w)_- = \sum_{n \geq 0} w^n z^{-n-1}, \quad -\delta(z-w)_+ = -\frac{1}{z} \sum_{n < 0} w^n z^{-n-1},$$

respectively.

Now we can relax the condition of commutativity of two fields by requiring that for any $v \in V$ and $v^\vee \in V^\vee$, the matrix elements $\langle v^\vee | \phi(z)\psi(w) | v \rangle$ and $\langle v^\vee | \psi(w)\phi(z) | v \rangle$ are the images of the same element $f_{v^\vee, v}$ of $\mathbb{C}((z, w))$ in $\mathbb{C}((z))((w))$ and $\mathbb{C}((w))((z))$, respectively. In the case of vertex algebras, we additionally require that

$$f_{v^\vee, v} \in \mathbb{C}[[z, w]][z^{-1}, w^{-1}, (z - w)^{-1}]$$

for all $v \in V, v^\vee \in V^\vee$, and that the order of pole of $f_{v^\vee, v}$ is universally bounded, i.e., there exists $N \in \mathbb{Z}_+$, such that $(z - w)^N f_{v^\vee, v} \in \mathbb{C}[z^{\pm 1}, w^{\pm 1}]$ for all v, v^\vee . The last condition is equivalent to the condition of locality given by formula (1).

From the analytic point of view, locality of the fields $\phi(z)$ and $\phi(w)$ means that the matrix element $\langle v^\vee | \phi(z)\psi(w) | v \rangle$ is well-defined in the domain $|z| > |w|$ whereas $\langle v^\vee | \psi(w)\phi(z) | v \rangle$ is well-defined in the domain $|w| > |z|$, but both can be analytically continued to the same function $f_{v^\vee, v}(z, w)$. Then their commutator (considered as a distribution) is the difference between boundary values of holomorphic functions, and hence is a delta-like distribution supported on the diagonal. In the simplest case, this amounts to the formula $\delta(z - w) = \delta(z - w)_- + \delta(z - w)_+$, which is a version of the Sokhotsky–Plemelj formula well-known in complex analysis.

2.6. Associativity

Now we state the “associativity” property of vertex algebras. Consider the $\text{End } V$ -valued formal power series

$$Y(Y(A, z - w)B, w) = \sum_{n \in \mathbb{Z}} Y(A_n \cdot B, w)(z - w)^{-n - \Delta_A}$$

in w and $z - w$. By degree reasons, this is an element of $\text{End } V((w))((z - w))$ (i.e., the order of pole at $z = w$ is bounded). By mapping $(z - w)^{-j}$ to $(\delta(z - w)_-)^j$ (its expansion in w/z), we obtain an embedding of $\text{End } V((w))((z - w))$ into $\text{End } V[[z^{\pm 1}, w^{\pm 1}]]$.

PROPOSITION 2.6 ([FHL, K2]). — *Any vertex algebra V satisfies the following associativity property: for any $A, B \in V$ we have the equality of formal power series*

$$(7) \quad Y(A, z)Y(B, w) = Y(Y(A, z - w)B, w) = \sum_{n \in \mathbb{Z}} \frac{Y(A_n \cdot B, w)}{(z - w)^{n + \Delta_A}},$$

where by the right hand side we understand its image in $\text{End } V[[z^{\pm 1}, w^{\pm 1}]]$.

Formula (7) is called the operator product expansion (OPE). From the physics point of view, it manifests the important property in quantum field theory that the product of two fields at nearby points can be expanded in terms of other fields and the small parameter $z - w$. From the analytic point of view, this formula expresses the fact that the matrix elements of the left and right hand sides of the formula, well-defined in the appropriate

domains, can be analytically continued to the same rational functions in z, w , with poles only at $z = 0, w = 0$, and $z = w$.

For example, in the case of Heisenberg algebra, we obtain:

$$(8) \quad b(z)b(w) = \frac{1}{(z-w)^2} + \sum_{n \geq 0} \frac{1}{n!} : \partial_w^n b(w)b(w) : (z-w)^n.$$

Using the Cauchy formula, one can easily extract the commutation relations between the Fourier coefficients of the fields $Y(A, z)$ and $Y(B, w)$ from the singular (at $z = w$) terms of their OPE. The result is

$$(9) \quad [A_m, B_k] = \sum_{n > -\Delta_A} \binom{m + \Delta_A - 1}{n + \Delta_A - 1} (A_n \cdot B)_{m+k}.$$

2.7. Correlation functions

Formula (7) and the locality property have the following “multi-point” generalization. Let V^* be the space of all linear functionals on V .

PROPOSITION 2.7 ([FHL]). — *Let $A_1, \dots, A_n \in V$. For any $v \in V, \varphi \in V^*$, and any permutation σ on n elements, the formal power series in z_1, \dots, z_n ,*

$$(10) \quad \varphi \left(Y(A_{\sigma(1)}, z_{\sigma(1)}) \dots Y(A_{\sigma(n)}, z_{\sigma(n)}) | 0 \right)$$

is the expansion in $\mathbb{C}((z_{\sigma(1)})) \dots ((z_{\sigma(n)}))$ of an element $f_{A_1, \dots, A_n}^\sigma(z_1, \dots, z_n)$ of $\mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}$, which satisfies the following properties: it does not depend on σ (so we suppress it in the notation);

$$f_{A_1, \dots, A_n}(z_1, \dots, z_n) = f_{Y(A_i, z_i - z_j) A_j, A_1, \dots, \widehat{A}_i, \dots, \widehat{A}_j, \dots, A_n}(z_j, z_1, \dots, \widehat{z}_i, \dots, \widehat{z}_j, \dots, z_n)$$

for all $i \neq j$; and $\partial_{z_i} f_{A_1, \dots, A_n}(z_1, \dots, z_n) = f_{A_1, \dots, T A_i, \dots, A_n}(z_1, \dots, A_n)$.

Thus, to each $\varphi \in V^*$ we can attach a collection of matrix elements (10), the “ n -point functions” on $\text{Spec } \mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}$. They satisfy a symmetry condition, a “bootstrap” condition (which describes the behavior of the n -point function near the diagonals in terms of $(n-1)$ -point functions) and a “horizontality” condition. We obtain a linear map from V^* to the vector space \mathcal{F}_n of all collections $\{f_{A_1, \dots, A_m}(z_1, \dots, z_m), A_i \in V\}_{m=1}^n$ satisfying the above conditions. This map is actually an isomorphism for each $n \geq 1$. The inverse map $\mathcal{F}_n \rightarrow V^*$ takes $\{f_{A_1, \dots, A_m}(z_1, \dots, z_m), A_i \in V\}_{m=1}^n \in \mathcal{F}_n$ to the functional φ on V , defined by the formula $\varphi(A) = f_A(0)$. Thus, we obtain a “functional realization” of V^* . In the case when V is generated by fields such that the singular terms in their OPEs are linear combinations of the same fields and their derivatives, we can simplify this functional realization by considering only the n -point functions of the generating fields.

For example, in the case of the Heisenberg vertex algebra we consider for each $\varphi \in \pi^*$ the n -point functions

$$(11) \quad \omega_n(z_1, \dots, z_n) = \varphi(b(z_1) \dots b(z_n)|0).$$

By Proposition 2.7 and the OPE (8), these functions are symmetric and satisfy the bootstrap condition

$$(12) \quad \omega_n(z_1, \dots, z_n) = \left(\frac{\omega_{r-2}(z_1, \dots, \widehat{z}_i, \dots, \widehat{z}_j, \dots, z_n)}{(z_i - z_j)^2} + \text{regular} \right).$$

Let Ω_∞ be the vector space of infinite collections $(\omega_n)_{n \geq 0}$, where

$$\omega_n \in \mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}$$

satisfy the above conditions. Using the functions (11), we obtain a map $\pi^* \rightarrow \Omega_\infty$. One can show that this map is an isomorphism.

3. MORE EXAMPLES

3.1. Affine Kac-Moody algebras

Let \mathfrak{g} be a simple Lie algebra over \mathbb{C} . Consider the formal loop algebra $L\mathfrak{g} = \mathfrak{g}((t))$ with the obvious commutator. The affine algebra $\widehat{\mathfrak{g}}$ is defined as a central extension of $L\mathfrak{g}$. As a vector space, $\widehat{\mathfrak{g}} = L\mathfrak{g} \oplus \mathbb{C}K$, and the commutation relations read: $[K, \cdot] = 0$ and

$$(13) \quad [A \otimes f(t), B \otimes g(t)] = [A, B] \otimes f(t)g(t) + (\text{Res}_{t=0} fdg(A, B))K.$$

Here (\cdot, \cdot) is an invariant bilinear form on \mathfrak{g} (such a form is unique up to a scalar multiple; we normalize it as in [K1] by the requirement that $(\alpha_{\max}, \alpha_{\max}) = 2$).

Consider the Lie subalgebra $\mathfrak{g}[[t]]$ of $L\mathfrak{g}$. If $f, g \in \mathbb{C}[[t]]$, then $\text{Res}_{t=0} fdg = 0$. Hence $\mathfrak{g}[[t]]$ is a Lie subalgebra of $\widehat{\mathfrak{g}}$. Let \mathbb{C}_k be the one-dimensional representation of $\mathfrak{g}[[t]] \oplus \mathbb{C}K$, on which $\mathfrak{g}[[t]]$ acts by 0 and K acts by the scalar $k \in \mathbb{C}$. We define the vacuum representation of $\widehat{\mathfrak{g}}$ of level k as the representation induced from \mathbb{C}_k : $V_k(\mathfrak{g}) = \text{Ind}_{\mathfrak{g}[[t]] \oplus \mathbb{C}K}^{\widehat{\mathfrak{g}}} \mathbb{C}_k$.

Let $\{J^a\}_{a=1, \dots, \dim \mathfrak{g}}$ be a basis of \mathfrak{g} . Denote $J_n^a = J^a \otimes t^n \in L\mathfrak{g}$. Then $J_n^a, n \in \mathbb{Z}$, and K form a (topological) basis for $\widehat{\mathfrak{g}}$. By the Poincaré–Birkhoff–Witt theorem, $V_k(\mathfrak{g})$ has a basis of lexicographically ordered monomials of the form $J_{n_1}^{a_1} \dots J_{n_m}^{a_m} v_k$, where v_k is the image of $1 \in \mathbb{C}_k$ in $V_k(\mathfrak{g})$, and all $n_i < 0$. We are now in the situation of the Reconstruction Theorem, and hence we obtain a vertex algebra structure on $V_k(\mathfrak{g})$, such that

$$Y(J_{-1}^a v_k, z) = J^a(z) := \sum_{n \in \mathbb{Z}} J_n^a z^{-n-1}.$$

The fields corresponding to other monomials are obtained by formula (5). Explicit computation shows that the fields $J^a(z)$ (and hence all other fields) are mutually local.

3.2. Virasoro algebra

The Virasoro algebra Vir is a central extension of the Lie algebra $\text{Der } \mathbb{C}((t)) = \mathbb{C}((t)) \frac{\partial}{\partial t}$. It has (topological) basis elements $L_n = -t^{n+1}, n \in \mathbb{Z}$, and C satisfying the relations that C is central and

$$[L_n, L_m] = (n - m)L_{n+m} + \frac{n^3 - n}{12} \delta_{n,-m} C$$

(a coordinate independent version is given in §5.2). These relations imply that the Lie algebra $\text{Der } \mathbb{C}[[t]] = \mathbb{C}[[t]] \frac{\partial}{\partial t}$ generated by $L_n, n \geq 0$, is a Lie subalgebra of Vir . Consider the induced representation $\text{Vir}_c = \text{Ind}_{\text{Der } \mathbb{C}[[t]] \oplus \mathbb{C}C}^{Vir} \mathbb{C}_c$, where \mathbb{C}_c is a one-dimensional representation, on which C acts as c and $\text{Der } \mathbb{C}[[t]]$ acts by zero. By the Poincaré–Birkhoff–Witt theorem, Vir_c has a basis consisting of monomials of the form $L_{j_1} \dots L_{j_m} v_c, j_1 \leq j_2 \leq j_m \leq -2$, where v_c is the image of $1 \in \mathbb{C}_c$ in the induced representation. By Reconstruction Theorem, we obtain a vertex algebra structure on Vir_c , such that

$$Y(L_{-2}v_c, z) = T(z) := \sum_{n \in \mathbb{Z}} L_n z^{-n-2}.$$

Note that the translation operator is equal to L_{-1} .

The Lie algebra $\text{Der } \mathcal{O}$ generates infinitesimal changes of coordinates. As we will see in §4, it is important to have this Lie algebra (and even better, the whole Virasoro algebra) act on a given vertex algebra by “internal symmetries”. This property is formalized by the following definition.

DEFINITION 3.1. — *A vertex algebra V is called conformal, of central charge $c \in \mathbb{C}$, if V contains a non-zero conformal vector $\omega \in V_2$, such that the corresponding vertex operator $Y(\omega, z) = \sum_{n \in \mathbb{Z}} L_n z^{-n-2}$ satisfies: $L_{-1} = T$, $L_0|_{V_n} = n \text{Id}$, and $L_2\omega = \frac{1}{2}c|0\rangle$.*

These conditions imply that the operators $L_n, n \in \mathbb{Z}$, satisfy the commutation relations of the Virasoro algebra with central charge c . We also obtain a non-trivial homomorphism $\text{Vir}_c \rightarrow V$, sending $L_{-2}v_c$ to ω .

For example, the vector $\frac{1}{2}b_{-1}^2 + \lambda b_{-2}$ is a conformal vector in π for any $\lambda \in \mathbb{C}$. The corresponding central charge equals $1 - 12\lambda^2$. The Kac-Moody vertex algebra $V_k(\mathfrak{g})$ is conformal if $k \neq -h^\vee$, where h^\vee is the dual Coxeter number of \mathfrak{g} . The conformal vector is given by the Sugawara formula $\frac{1}{2(k+h^\vee)} \sum_a (J_{-1}^a)^2 v_k$, where $\{J^a\}$ is an orthonormal basis of \mathfrak{g} . Thus, each $\widehat{\mathfrak{g}}$ -module from the category \mathcal{O} of level $k \neq -h^\vee$ is automatically a module over the Virasoro algebra.

3.3. Boson–fermion correspondence

Let \mathcal{C} be the Clifford algebra associated to the vector space $\mathbb{C}((t)) \oplus \mathbb{C}((t))dt$ equipped with the inner product induced by the residue pairing. It has topological generators $\psi_n = t^n, \psi_n^* = t^{n-1}dt, n \in \mathbb{Z}$, satisfying the anti-commutation relations

$$(14) \quad [\psi_n, \psi_m]_+ = [\psi_n^*, \psi_m^*]_+ = 0, \quad [\psi_n, \psi_m^*]_+ = \delta_{n,-m}.$$

Denote by \bigwedge the fermionic Fock representation of \mathcal{C} , generated by vector $|0\rangle$, such that $\psi_n|0\rangle = 0, n \geq 0, \psi_n^*|0\rangle = 0, n > 0$. This is a vertex superalgebra with

$$Y(\psi_{-1}|0\rangle, z) = \psi(z) := \sum_{n \in \mathbb{Z}} \psi_n z^{-n-1}, \quad Y(\psi_0^*|0\rangle, z) = \psi^*(z) := \sum_{n \in \mathbb{Z}} \psi_n^* z^{-n}.$$

Boson–fermion correspondence establishes an isomorphism between \bigwedge and a vertex superalgebra built out of the Heisenberg algebra \mathcal{H} from §2.3. For $\lambda \in \mathbb{C}$, let π_λ be the \mathcal{H} -module generated by a vector $\mathbf{1}_\lambda$, such that $b_n \cdot \mathbf{1}_\lambda = \lambda \delta_{n,0} \mathbf{1}_\lambda, n \geq 0$. For $N \in \mathbb{Z}_{>0}$, set $V_{\sqrt{N}\mathbb{Z}} = \bigoplus_{m \in \mathbb{Z}} \pi_{m\sqrt{N}}$. This is a vertex algebra (resp., superalgebra) for any even N (resp., odd N), which contains $\pi_0 = \pi$ as a vertex subalgebra. The gradation is given by the formulas $\deg b_n = -n, \deg \mathbf{1}_\lambda = \lambda^2/2$ (note that it can be half-integral). The translation operator is $T = \frac{1}{2} \sum_{n \in \mathbb{Z}} b_n b_{-1-n}$, so $T \cdot \mathbf{1}_\lambda = \lambda b_{-1} \mathbf{1}_\lambda$. In order to define the vertex operation on $V_{\sqrt{N}\mathbb{Z}}$, it suffices to define the fields $Y(\mathbf{1}_{m\sqrt{N}}, z)$. They are determined by the identity

$$\partial_z Y(\mathbf{1}_\lambda, z) = Y(T \cdot \mathbf{1}_\lambda, z) = \lambda : b(z) Y(\mathbf{1}_\lambda, z) : \quad .$$

Explicitly,

$$Y(\mathbf{1}_\lambda, z) = S_\lambda z^{\lambda b_0} \exp \left(-\lambda \sum_{n < 0} \frac{b_n}{n} z^{-n} \right) \exp \left(-\lambda \sum_{n > 0} \frac{b_n}{n} z^{-n} \right),$$

where $S_\lambda : \pi_\mu \rightarrow \pi_{\mu+\lambda}$ is the shift operator, $S_\lambda \cdot \mathbf{1}_\mu = \mathbf{1}_{\mu+\lambda}, [S_\lambda, b_n] = 0, n \neq 0$.

The *boson–fermion correspondence* is an isomorphism of vertex superalgebras $\bigwedge \simeq V_{\mathbb{Z}}$, which maps $\psi(z)$ to $Y(\mathbf{1}_{-1}, z)$ and $\psi^*(z)$ to $Y(\mathbf{1}_1, z)$. For more details, see, e.g., [K2].

3.4. Rational vertex algebras

Rational vertex algebras constitute an important class of vertex algebras, which are particularly relevant to conformal field theory [dFMS]. In order to define them, we first need to give the definition of a module over a vertex algebra.

Let V be a vertex algebra. A vector space M is called a V -module if it is equipped with the following data:

- gradation: $M = \bigoplus_{n \in \mathbb{Z}_+} M_n$;
- operation $Y_M : V \rightarrow \text{End } M[[z, z^{-1}]]$, which assigns to each $A \in V_n$ a field $Y_M(A, z)$ of conformal dimension n on M ;

subject to the following conditions:

- $Y_M(|0\rangle, z) = \text{Id}_M$;
- $Y_M(A, z)Y_M(B, w) = Y_M(Y(A, z-w)B, w)$, in the sense of Proposition 2.6.

A conformal vertex algebra V (see Definition 3.1) is called *rational* if every V -module is completely reducible. It is shown in [DLM] that this condition implies that V has *finitely many* inequivalent simple modules, and the graded components of each simple V -module M are finite-dimensional. Furthermore, the gradation operator on M coincides with the

Virasoro operator L_0^M up to a shift. Hence we can attach to a simple V -module M its character $\text{ch } M = \text{Tr}_M q^{L_0^M - c/24}$, where c is the central charge of V . Set $q = e^{2\pi i\tau}$. Zhu [Z1] has shown that if V satisfies a certain finiteness condition, then the characters give rise to holomorphic functions in τ on the upper-half plane. Moreover, he proved the following remarkable

THEOREM 3.2. — *Let V be a rational vertex algebra satisfying Zhu's finiteness condition, and $\{M_1, \dots, M_n\}$ be the set of all inequivalent simple V -modules. Then the vector space spanned by $\text{ch } M_i, i = 1, \dots, n$, is invariant under the action of $SL_2(\mathbb{Z})$.*

This result has the following heuristic explanation. To each vertex algebra we can attach the sheaves of coinvariants on the moduli spaces of curves (see §6). It is expected that the sheaves associated to a rational vertex algebra satisfying Zhu's condition are vector bundles with projectively flat connection. The characters of simple V -modules should form the basis of the space of horizontal sections of the corresponding bundle on the moduli of elliptic curves near $\tau = \infty$. The $SL_2(\mathbb{Z})$ action is just the monodromy action on these sections.

Here are some examples of rational vertex algebras.

(1) Let L be an even positive definite lattice in a real vector space W . One can attach to it a vertex algebra V_L in the same way as in §3.3 for $L = \sqrt{N}\mathbb{Z}$. Namely, we define the Heisenberg Lie algebra associated to $W_{\mathbb{C}}$ as the Kac-Moody type central extension of the commutative Lie algebra $W_{\mathbb{C}}((t))$. Its Fock representation π_W is a vertex algebra. The vertex structure on π_W can be extended to $V_L = \pi_W \otimes \mathbb{C}[L]$, where $\mathbb{C}[L]$ is the group algebra of L [B1, FLM]. Then V_L is rational, and its inequivalent simple modules are parameterized by L'/L where L' is the dual lattice [D1]. The corresponding characters are theta-functions. The vertex algebra V_L is the chiral symmetry algebra of the free bosonic conformal field theory compactified on the torus W/L .

Note that if we take as L an arbitrary integral lattice, then V_L is a vertex superalgebra.

(2) Let \mathfrak{g} be a simple Lie algebra and $k \in \mathbb{Z}_+$. Let $L_k(\mathfrak{g})$ be the irreducible quotient of the $\widehat{\mathfrak{g}}$ -module $V_k(\mathfrak{g})$. This is a rational vertex algebra, whose modules are integrable representations of $\widehat{\mathfrak{g}}$ of level k [FZ]. The corresponding conformal field theory is the Wess-Zumino-Witten model (see [dFMS]).

(3) Let $L_{c(p,q)}$ be the irreducible quotient of $\text{Vir}_{c(p,q)}$ (as a module over the Virasoro algebra), where $c(p,q) = 1 - 6(p-q)^2/pq$, $p, q > 1$, $(p,q) = 1$. This is a rational vertex algebra [Wa], whose simple modules form the "minimal model" of conformal field theory defined by Belavin, Polyakov, and Zamolodchikov [BPZ] (see also [dFMS]).

(4) The Moonshine Module vertex algebra V^{\natural} (see below).

Zhu [Z1] has attached to each vertex algebra V an associative algebra $A(V)$, such that simple V -modules are in one-to-one correspondence with simple $A(V)$ -modules.

3.5. Orbifolds and the Monster

Let G be a group of automorphisms of a vertex algebra V . Then the space V^G of G -invariants in V is a vertex subalgebra of V . When G is a finite group, *orbifold theory* allows one to construct V^G -modules from twisted V -modules. For a finite order automorphism g of V , one defines a g -twisted V -module by modifying the above axioms of V -module in such a way that $Y_M(A, z)$ has monodromy λ^{-1} around the origin, if $g \cdot A = \lambda v$ (see [DLM]). Then conjecturally all V^G -modules can be obtained as V^G -submodules of g -twisted V -modules for $g \in G$.

Now suppose that V is a *holomorphic* vertex algebra, i.e., rational with a unique simple module, namely itself. Let G be a cyclic group of automorphisms of V of order n generated by g . Then the following pattern is expected to hold (see [D2]): (1) for each $h \in G$ there is a unique simple h -twisted V -module V_h ; (2) V_h has a natural G -action, so we can write $V_h = \bigoplus_{i=0}^{n-1} V_h(i)$, where $V_h(i) = \{v \in V_h | g \cdot v = e^{2\pi i k/n} v\}$; then each $V_h(i)$ is a simple V^G -module, and these are all simple V^G -modules; (3) The vector space $\tilde{V} = \bigoplus_{h \in G} V_h(0)$ carries a canonical vertex algebra structure, which is holomorphic.

A spectacular example is the construction of the *Moonshine Module* vertex algebra V^\natural by I. Frenkel, Lepowsky and Meurman [FLM]. In this case V is the vertex algebra V_Λ associated to the Leech lattice Λ (it is self-dual, hence V_Λ is holomorphic), and g is constructed from the involution -1 on Λ . Then $V^\natural = V_\Lambda(0) \oplus V_{\Lambda, g}(0)$ is a vertex algebra, whose group of automorphisms is the Monster group [FLM]. Moreover, V^\natural is holomorphic [D2]. Conjecturally, V^\natural is the unique holomorphic vertex algebra V with central charge 24, such that $V_1 = 0$. Its character is the modular function $j(\tau) - 744$. More generally, for each element x of the Monster group, consider the Thompson series $\text{Tr}_{V^\natural} x q^{L_0 - 1}$. Conway and Norton conjectured, and Borcherds proved [B2] that these are Hauptmoduls for genus zero subgroups of $SL_2(\mathbb{R})$.

3.6. Coset construction

Let V be a vertex algebra, and W its vertex subalgebra. Let $C(V, W)$ be the vector subspace of V spanned by vectors v , such that $Y(A, z) \cdot v \in V[[z]]$ for all $A \in W$. Then $C(V, W)$ is a vertex subalgebra of V , which is called the *coset vertex algebra* of the pair (V, W) . In the case $W = V$, the vertex algebra $C(V, V)$ is commutative, and is called the *center* of V . An example of a coset vertex algebra is provided by the famous Goddard–Kent–Olive construction [GKO], which identifies $C(L_1(\mathfrak{sl}_2) \otimes L_k(\mathfrak{sl}_2), L_{k+1}(\mathfrak{sl}_2))$ with $L_{c(k+2, k+3)}$ (see §3.4). For other examples, see [dFMS].

3.7. BRST construction and \mathcal{W} -algebras

For $A \in V$, let $y(A) = \text{Res } Y(A, z)$. Then associativity implies: $[y(A), Y(B, z)] = Y(y(A) \cdot B, z)$, and so $y(A)$ is an infinitesimal automorphism of V . Now suppose that V^\bullet is a vertex (super)algebra with an additional \mathbb{Z} -gradation and $A \in V^1$ is such that

$y(A)^2 = 0$. Then $(V^\bullet, y(A))$ is a complex, and its cohomology is a graded vertex (super)algebra. Important examples of such complexes are provided by *topological vertex superalgebras* introduced by Lian and Zuckerman [LZ]. In this case the cohomology is a graded commutative vertex algebra, but it has an additional structure of Batalin-Vilkovisky algebra.

Another example is the BRST complex of quantum hamiltonian reduction. We illustrate it in the case of quantum Drinfeld-Sokolov reduction [FF3], which leads to the definition of \mathcal{W} -algebras. Let \mathfrak{g} be a simple Lie algebra of rank ℓ , and \mathfrak{n} its upper nilpotent subalgebra.

Let \mathcal{C} be the Clifford algebra associated to the vector space $\mathfrak{n}((t)) \oplus \mathfrak{n}^*((t))dt$ equipped with the inner product induced by the residue pairing. It has topological generators $\psi_{\alpha,n} = e_\alpha \otimes t^n, \psi_{\alpha,n}^* = e_\alpha^* \otimes t^{n-1}dt, \alpha \in \Delta_+, n \in \mathbb{Z}$, satisfying the anti-commutation relations (14). Let $\bigwedge_{\mathfrak{n}}$ be its Fock representation, generated by vector $|0\rangle$, such that $\psi_{\alpha,n}|0\rangle = 0, n \geq 0, \psi_{\alpha,n}^*|0\rangle = 0, n > 0$. This is a vertex superalgebra which is the tensor product of several copies of the vertex superalgebra \bigwedge from §3.3. Introduce an additional \mathbb{Z} -gradation on \mathcal{C} and $\bigwedge_{\mathfrak{n}}$ by setting $\deg \psi_{\alpha,n}^* = -\deg \psi_{\alpha,n} = 1, \deg |0\rangle = 0$.

Now consider the vertex superalgebra $C_k^\bullet(\mathfrak{g}) = V_k(\mathfrak{g}) \otimes \bigwedge_{\mathfrak{n}}^\bullet$. It carries a canonical differential d_{st} of degree 1, the standard differential of semi-infinite cohomology of $\mathfrak{n}((t))$ with coefficients in V_k [Fe1, FGZ]; it is equal to the residue of a field from $C_k^\bullet(\mathfrak{g})$ (see [FF3]). Let $\chi = \sum_{i=1}^{\ell} \text{Res} \psi_{\alpha_i}^*(z)$ be the Drinfeld-Sokolov character [DS] of $\mathfrak{n}((t))$. Then $d = d_{\text{st}} + \chi$ is a differential on $C_k^\bullet(\mathfrak{g})$, and the cohomology $H_k^\bullet(\mathfrak{g})$ of this differential is a vertex superalgebra.

THEOREM 3.3. — $H_k^0(\mathfrak{g})$ is a vertex algebra generated by elements W_i of degrees $d_i+1, i = 1, \dots, \ell$, where d_i is the i th exponent of \mathfrak{g} (in the sense of the Reconstruction Theorem), and $H_k^i(\mathfrak{g}) = 0, i \neq 0$.

This theorem is proved in [FF3, FF4] for generic k and in [dBT] for all k . The vertex algebra $H_k^0(\mathfrak{g})$ is called the \mathcal{W} -algebra associated to \mathfrak{g} and denoted by $\mathcal{W}_k(\mathfrak{g})$. We have: $\mathcal{W}_k(\mathfrak{sl}_2) = \text{Vir}_{c(k)}$, where $c(k) = 1 - 6(k+1)^2/(k+2)$. For $\mathfrak{g} = \mathfrak{sl}_3$, the \mathcal{W} -algebra was first constructed by Zamolodchikov, and for $\mathfrak{g} = \mathfrak{sl}_N$ by Fateev and Lukyanov [FL]. Since $V_k(\mathfrak{g})$ is conformal for $k \neq -h^\vee$ (see §3.2), $\mathcal{W}_k(\mathfrak{g})$ is also conformal, with W_1 playing the role of conformal vector. On the other hand, $\mathcal{W}_{-h^\vee}(\mathfrak{g})$ is a commutative vertex algebra, which is isomorphic to the center of $V_{-h^\vee}(\mathfrak{g})$ [FF3] (see §6.5). The simple quotient of $\mathcal{W}_k(\mathfrak{g})$ for $k = -h^\vee + p/q$, where p, q are relatively prime integers greater than or equal to h^\vee , is believed to be a rational vertex algebra. Moreover, if \mathfrak{g} is simply-laced and $q = p + 1$, this vertex algebra is conjecturally isomorphic to the coset algebra of $(L_1(\mathfrak{g}) \otimes L_{p-2}(\mathfrak{g}), L_{p-1}(\mathfrak{g}))$, see [FKW].

For any $V_k(\mathfrak{g})$ -module M , the cohomology of the complex $(M \otimes \bigwedge_{\mathfrak{n}}^\bullet, d)$ is a $\mathcal{W}_k(\mathfrak{g})$ -module. This defines a functor, which was studied in [FKW].

The \mathcal{W} -algebras exhibit a remarkable duality: $\mathcal{W}_k(\mathfrak{g}) \simeq \mathcal{W}_{k'}({}^L\mathfrak{g})$, where ${}^L\mathfrak{g}$ is the Langlands dual Lie algebra to \mathfrak{g} and $(k + h^\vee)r^\vee = (k' + {}^Lh^\vee)^{-1}$ (here r^\vee denotes the maximal number of edges connecting two vertices of the Dynkin diagram of \mathfrak{g}), see [FF3]. In the limit $k \rightarrow -h^\vee$ it becomes the isomorphism of Theorem 6.3.

4. COORDINATE INDEPENDENT DESCRIPTION OF VERTEX ALGEBRA STRUCTURE

Up to this point, we have discussed vertex algebras in the language of formal power series. The vertex operation is a map $Y : V \rightarrow \text{End } V[[z, z^{-1}]]$, or, equivalently, an element of the completed tensor product $V^*((z)) \widehat{\otimes} \text{End } V$. Thus, we can view Y as an $\text{End } V$ -valued section of a vector bundle over the punctured disc $D^\times = \text{Spec } \mathbb{C}((z))$ with fiber V^* . The question is whether one can define this bundle in such a way that this section is canonical, i.e., independent of the choice of z . In order to do that, we need a precise description of the transformation properties of Y . Once we understand what type of geometric object Y is, we will be able to give a global geometric meaning to vertex operators on arbitrary curves.

4.1. The group $\text{Aut } \mathcal{O}$

Denote by \mathcal{O} the complete topological \mathbb{C} -algebra $\mathbb{C}[[z]]$, and let $\text{Aut } \mathcal{O}$ be the group of continuous automorphisms of \mathcal{O} . Such an automorphism is determined by its action on the generator $z \in \mathbb{C}[[z]]$. Thus as a set this group consists of elements $\rho(z) \in \mathbb{C}[[z]]$ of the form $\rho_1 z + \rho_2 z^2 + \dots$, with $\rho_1 \in \mathbb{C}^\times$, endowed with the composition law $(\rho * \mu)(z) = \mu(\rho(z))$. It is easy to see that $\text{Aut } \mathcal{O}$ is a proalgebraic group, $\varprojlim \text{Aut } \mathbb{C}[[z]]/(z^n)$. It is the semi-direct product $\mathbb{G}_m \ltimes \text{Aut}_+ \mathcal{O}$, where $\text{Aut}_+ \mathcal{O}$ is the subgroup that consists of the transformations of the form $z \rightarrow z + a_2 z^2 + \dots$, and \mathbb{G}_m is the group of rescalings $z \rightarrow az, a \neq 0$. Furthermore, $\text{Aut}_+ \mathcal{O}$ is a pronipotent group. The Lie algebra of $\text{Aut } \mathcal{O}$ is $\text{Der}_0 \mathcal{O} = z\mathbb{C}[[z]]\partial_z$, which is a semi-direct product of $\mathbb{C}\partial_z$ and $\text{Der}_+ \mathcal{O} = z^2\mathbb{C}[[z]]\partial_z$. Any representation of $\text{Der}_0 \mathcal{O}$, on which $z\partial_z$ is diagonalizable with integral eigenvalues and $\text{Der}_+ \mathcal{O}$ acts locally nilpotently, can be exponentiated to a representation of $\text{Aut } \mathcal{O}$.

Given a vertex algebra and a field $Y(A, z)$, it makes sense to consider a new field $Y(A, \rho(z))$. We now seek an action of $\text{Aut } \mathcal{O}$ on V , $A \mapsto R(\rho) \cdot A$, such that $Y(A, \rho(z))$ is related to $Y(R(\rho)A, z)$ in some reasonable way. In the theory of vertex algebras, actions of $\text{Aut } \mathcal{O}$ usually arise from the action of the Virasoro algebra. Namely, let V be a conformal vertex algebra (see Definition 3.1). Then the Fourier coefficients L_n of $Y(\omega, z)$ satisfy the commutation relations of the Virasoro algebra with central charge c . The operators $L_n, n \geq 0$, then define an action of the Lie algebra $\text{Der}_+ \mathcal{O}$ (L_n corresponds to $-t^{n+1}\partial_t$). It follows from the axioms of vertex algebra that this action can be exponentiated to an action of $\text{Aut } \mathcal{O}$. For $f(z) \in \text{Aut } \mathcal{O}$, denote by $R(f) : V \rightarrow V$ the corresponding operator.

Given a vector field $v = \sum_{n \geq -1} v_n z^{n+1} \partial_z$, we assign to it the operator $\mathbf{v} = - \sum_{n \geq -1} v_n L_n$.

From formula (9) we obtain:

$$(15) \quad [\mathbf{v}, Y(A, w)] = - \sum_{m \geq -1} \frac{1}{(m+1)!} (\partial_w^{m+1} v(w)) Y(L_m A, w).$$

By “exponentiating” this formula, we obtain the following identity, due to Y.-Z. Huang [Hu]:

$$(16) \quad Y(A, t) = R(\rho) Y(R(\rho_t)^{-1} A, \rho(t)) R(\rho)^{-1},$$

for any $\rho \in \text{Aut } \mathbb{C}[[z]]$ (here $\rho_t(z) = \rho(t+z) - \rho(t)$, considered as formal power series in z with coefficients in $\mathbb{C}[[t]]$).

4.2. Vertex algebra bundle

Now we can give a coordinate-free description of the operation Y . Let X be a smooth complex curve. Given a point $x \in X$, denote by \mathcal{O}_x the completion of the local ring of x , by \mathfrak{m}_x its maximal ideal, and by \mathcal{K}_x the field of fractions of \mathcal{O}_x . A formal coordinate t_x at x is by definition a topological generator of \mathfrak{m}_x . Consider the set of pairs (x, t_x) , where $x \in X$ and t_x is a formal coordinate at x . This is the set of points of a scheme \widehat{X} of infinite type, which is a principal $\text{Aut } \mathcal{O}$ -bundle over X . Its fiber at $x \in X$ is the $\text{Aut } \mathcal{O}$ -torsor \mathcal{L}_x of all formal coordinates at x . Given a finite-dimensional $\text{Aut } \mathcal{O}$ -module V , let $\mathcal{V} = \widehat{X} \times_{\text{Aut } \mathcal{O}} V$ be the vector bundle associated to V and \widehat{X} . The fiber of \mathcal{V} at $x \in X$ is the \mathcal{L}_x -twist of V , $\mathcal{V}_x = \mathcal{L}_x \times_{\text{Aut } \mathcal{O}} V$.

A conformal vertex algebra is an $\text{Aut } \mathcal{O}$ -module, which has a filtration $V_{\leq i} := \bigoplus_{k=0}^i V_k$ by finite-dimensional $\text{Aut } \mathcal{O}$ -submodules. We obtain the directed system $(\mathcal{V}_{\leq i})$ of the corresponding vector bundles of finite rank on X and embeddings $\mathcal{V}_{\leq i} \hookrightarrow \mathcal{V}_{\leq j}, i \leq j$. We will denote this system simply by \mathcal{V} , thinking of it as an inductive limit of bundles of finite rank. Likewise, by \mathcal{V}^* we will understand the inverse system of bundles $(\mathcal{V}_{\leq i})^*$ and surjections $(\mathcal{V}_{\leq j})^* \rightarrow (\mathcal{V}_{\leq i})^*$, thinking of it as a projective limit of bundles of finite rank.

The $\text{Aut } \mathcal{O}$ -bundle \widehat{X} on X above carries an action of the Lie algebra $\text{Der } \mathcal{O}$, which is compatible with the $\text{Aut } \mathcal{O}$ -action and simply transitive, i.e., $\text{Der } \mathcal{O} \widehat{\otimes} \mathcal{O}_{\widehat{X}} \simeq \Theta_{\widehat{X}}$. Since V is a $\text{Der } \mathcal{O}$ -module, by the general construction of §6.1 we obtain a flat connection $\nabla : \mathcal{V} \rightarrow \mathcal{V} \otimes \Omega$ on \mathcal{V} . Locally, ∇ can be written as $d + L_{-1} \otimes dz$, where d the deRham differential. The connection ∇ on \mathcal{V} gives us a connection ∇^* on the dual bundle \mathcal{V}^* . To be precise, \mathcal{V} is not an inductive limit of flat bundles, since $\nabla(\mathcal{V}_{\leq i}) \subset \mathcal{V}_{i+1} \otimes \Omega$, and likewise, \mathcal{V}^* is not a projective limit of flat bundles (it is a \mathcal{D} -module on X , which is not quasicohherent as an \mathcal{O} -module).

Now let us restrict \mathcal{V}^* to the punctured disc $D_x^\times = \text{Spec } \mathcal{K}_x$ around $x \in X$. We want to define an $\text{End } \mathcal{V}_x$ -valued meromorphic section \mathcal{Y}_x of \mathcal{V}^* on D_x^\times . Pick a formal coordinate z at x . With this choice, identify \mathcal{V}_x with V and trivialize $\mathcal{V}|_{D_x}$. We define our section

\mathcal{Y}_x in this trivialization through its matrix coefficients: to each triple $c \in V$, $c' \in V^*$ and a section $F : D_x \rightarrow V$ we need to assign a meromorphic function on the disc, which is \mathbb{C} -linear in c and c' and \mathcal{O} -linear in F . It suffices to assign such a function to triples c, c', F , where F is the constant section equal to $A \in V$ with respect to our trivialization. We set this function to be equal to $\langle c' | Y(A, z) | c \rangle$.

THEOREM 4.1. — *The section \mathcal{Y}_x defined this way is canonical, i.e., independent of the choice of formal coordinate z at x . Moreover, \mathcal{Y}_x is horizontal: $\nabla^* \mathcal{Y}_x = 0$.*

Thus, we see that a conformal vertex algebra gives rise to a vector bundle on any smooth curve and a canonical horizontal section of the restriction of this bundle to the neighborhood of each point with values in the endomorphisms of the corresponding twist of V . The fact that this section is canonical is equivalent to formula (16). Flatness follows from the formula $\partial_z Y(B, z) = Y(L_{-1}B, z)$, which holds in any vertex algebra.

REMARK 4.2. — Let $\mathcal{V}^{\boxtimes n}$ be the n -fold external tensor power of \mathcal{V} . Consider the n -point functions (10). According to Proposition 2.7, these are elements of

$$\mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}.$$

Trivializing the restriction of $\mathcal{V}^{\boxtimes n}$ to $D_x^n = \text{Spec } \mathbb{C}[[z_1, \dots, z_n]]$ using the coordinate z , we obtain a \mathcal{V}_x^* -valued section of $(\mathcal{V}^{\boxtimes n})^*|_{D_x^n}$ with poles on the diagonals. Then this section is canonical, i.e., independent of the choice of z , and horizontal.

4.3. Examples

If W is an $\text{Aut } \mathcal{O}$ -stable subspace of V , it gives rise to a subbundle \mathcal{W} of \mathcal{V} . The universal section \mathcal{Y}_x then gives rise to an $\text{End } \mathcal{V}_x$ -valued section of $\mathcal{W}^*|_{D_x^\times}$. This way one obtains many familiar geometric objects.

Suppose $A \in V_\Delta$ satisfies $L_n \cdot A = 0, L_0 A = \Delta A$. Then subspace $\mathbb{C}A \subset V$ is a one-dimensional $\text{Aut } \mathcal{O}$ submodule of V . The line bundle associated to this module is nothing but the line bundle $\Omega^{-\Delta}$ of Δ -differentials on X . Thus we obtain a line subbundle \mathcal{L}_A of \mathcal{V} . Our section \mathcal{Y}_x therefore gives us an $\text{End } \mathcal{V}_x$ -valued section of $\mathcal{L}_A^{-1} = \Omega^\Delta$. In other words, the $\text{End } \mathcal{V}_x$ -valued Δ -differential $Y(A, z)(dz)^\Delta$ does not depend on the choice of formal coordinate z at x .

Next, consider the Heisenberg vertex algebra $V = \pi$, with conformal structure given by the vector $\frac{1}{2}b_{-1}^2 + b_{-2}$. For each smooth curve X , π gives rise to a vector bundle that we denote by Π . The first piece π_0 of our filtration on $\pi = \mathbb{C}[b_n]_{n < 0}$ is the one-dimensional subspace spanned by the vacuum vector 1. This is a trivial representation of the group $\text{Aut } \mathcal{O}$. Hence it gives rise to a trivial subbundle $\mathcal{O}_X \subset \Pi$. The next piece in the filtration, $\pi_{\leq 1} = \mathbb{C}1 \oplus \mathbb{C}b_{-1}$, gives rise to a rank two subbundle of Π , which we denote by \mathcal{B} . Its dual bundle is an extension

$$(17) \quad 0 \rightarrow \Omega \rightarrow \mathcal{B}^* \rightarrow \mathcal{O}_X \rightarrow 0.$$

Our universal section \mathcal{Y}_x gives rise to an $\text{End } \Pi_x$ -valued section of $\mathcal{B}^*|_{D_x^\times}$, which projects onto the section $\text{End } \Pi_x$ -valued section of \mathcal{O}_X equal to Id (this follows from the vacuum axiom). Explicit computation shows that the space of sections of \mathcal{B}^* which project onto the section 1 of \mathcal{O}_X (equivalently, splittings of (17)) is canonically isomorphic to the space of *affine connections* (they may also be described as affine structures, see [Gu]). Thus, we conclude that the field $b(z)$ (or, more precisely, the expression $\partial_z + b(z)$) transforms as an $\text{End } \Pi_x$ -valued affine connection on D_x^\times . Similarly, we obtain connections on G -bundles from the subspace $V_k(\mathfrak{g})_{\leq 1}$ of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$.

Let V be a conformal vertex algebra with central charge c . It follows from the definition that the vector space $\mathbb{C}|0\rangle \oplus \mathbb{C}\omega$ is $\text{Aut } \mathcal{O}$ -stable. Hence it gives rise to a rank two subbundle \mathcal{T}_c of \mathcal{V} . Its dual bundle is an extension

$$(18) \quad 0 \rightarrow \Omega^2 \rightarrow \mathcal{T}_c^* \rightarrow \mathcal{O}_X \rightarrow 0.$$

The universal section \mathcal{Y}_x gives rise to an $\text{End } \mathcal{V}_x$ -valued section of \mathcal{T}_c^* over the punctured disc, which projects onto $\text{Id} \in \text{End } \mathcal{V}_x \otimes \mathcal{O}_x$. Explicit calculation shows that the space of sections of \mathcal{T}_c^* projecting onto $1 \in \Gamma(\mathcal{O}_X)$ (equivalently, splittings of (18)) is isomorphic to the space of self-adjoint second order differential operators $\rho : \Omega^{-\frac{1}{2}} \rightarrow \Omega^{\frac{3}{2}}$ with constant symbol $\frac{c}{6}$. Locally, such an operator can be written as $\frac{c}{6}\partial_z^2 + q(z)$. Operators of this form with symbol 1 are known as *projective connections* (they may also be described as projective structures, see [Gu]). Thus, we see that for $c \neq 0$ the field $T(z)$ (called the stress tensor in conformal field theory), or more precisely, the expression $\partial_z^2 + \frac{6}{c}T(z)$ transforms as an $\text{End } \mathcal{V}_x$ -valued projective connection on the punctured disc.

4.4. General twisting property

We have seen in §4.2 that the vertex operation Y is in some sense invariant under the $\text{Aut } \mathcal{O}$ -action (see formula (16)). Therefore Y gives rise to a well-defined operation on the twist of V by any $\text{Aut } \mathcal{O}$ -torsor. This “twisting property” follows from the fact that $\text{Aut } \mathcal{O}$ acts on V by “internal symmetries”, that is by exponentiation of Fourier coefficients of the vertex operator $Y(\omega, z)$. It turns out that vertex algebras exhibit a similar twisting property with respect to any group \mathcal{G} of internal symmetry, i.e., a group (or more generally, an ind-group) obtained by exponentiation of Fourier coefficients of vertex operators. Using associativity, one can obtain an analogue of formula (16) for the transformations of Y under the action of \mathcal{G} . This formula means that we get a well-defined operation on the twist of V by any \mathcal{G} -torsor.

For example, let V be a vertex algebra with a $\widehat{\mathfrak{g}}$ -structure of level $k \neq -h^\vee$, i.e., one equipped with a homomorphism $V_k(\mathfrak{g}) \rightarrow V$, whose image is not contained in $\mathbb{C}|0\rangle$. Recall that $V_k(\mathfrak{g})$ is a conformal vertex algebra if $k \neq -h^\vee$ (see §3.2). Given such a homomorphism $V_k(\mathfrak{g}) \rightarrow V$, the Fourier coefficients of the corresponding fields $Y(\omega, z)$ and $Y(J^a, z)$ generate an action of the Lie algebra $\text{Vir} \ltimes \widehat{\mathfrak{g}}$ on V . Suppose that the

action of its Lie subalgebra $\text{Der}_0 \mathcal{O} \times \mathfrak{g}(\mathcal{O})$ can be exponentiated to an action of the group $\text{Aut } \mathcal{O} \times G(\mathcal{O})$ on V (then we say that this group acts on V by internal symmetries). Let \mathcal{P} be a G -bundle on a smooth curve X . Denote by $\widehat{\mathcal{P}}$ the principal $\text{Aut } \mathcal{O} \times G(\mathcal{O})$ -bundle over X , whose fiber at x consists of pairs (z, s) , where z is a formal coordinate at x and s is a trivialization of $\mathcal{P}|_{D_x}$. Let $\mathcal{V}^{\mathcal{P}}$ be the $\widehat{\mathcal{P}}$ -twist of V . Then the vertex operation Y on V gives rise to a canonical section $\mathcal{Y}_x^{\mathcal{P}}$ of $(\mathcal{V}^{\mathcal{P}})^*|_{D_x^\times}$ with values in $\text{End } \mathcal{V}_x^{\mathcal{P}}$.

5. CONFORMAL BLOCKS

5.1. The results of the previous section allow us to assign to a conformal vertex algebra a vector bundle with connection on any smooth curve X , and a family of local structures on it. We can now associate to any compact curve X an invariant of the vertex algebra structure.

DEFINITION 5.1. — *Let V be a conformal vertex algebra, X a smooth projective curve, and x a point of X . A linear functional φ on \mathcal{V}_x is called a conformal block if $\varphi(\mathcal{Y}_x \cdot A) \in \Gamma(D_x^\times, \mathcal{V}^*)$ can be extended to a regular section of \mathcal{V}^* on $X \setminus x$ for all $A \in \mathcal{V}_x$.*

The set of conformal blocks is a vector subspace of \mathcal{V}_x^ , denoted by $C(X, x, V)$.*

Let $\varphi \in C(X, x, V)$, and $A \in \mathcal{V}_x$. Denote by φ_A the corresponding section of \mathcal{V}^* over $X \setminus x$. If W is an $\text{Aut } \mathcal{O}$ -stable subspace of V , then φ_A can be projected onto a section of W^* over $X \setminus x$. Note that according to the vacuum axiom, $\varphi_{|0\rangle}$ is actually a regular section of \mathcal{V}^* over the whole X , and so is its projection. In particular, taking as W the two-dimensional subspace of V spanned by $|0\rangle$ and ω , we assign to each conformal block a projective connection on X (if $c \neq 0$). Denote $\varphi(\cdot|0\rangle)$ as $\langle \cdot \rangle_\varphi$. If we choose a formal coordinate z at x , then we can write this projective connection as $\partial_z^2 + \frac{6}{c} \langle T(z) \rangle_\varphi$. If $c = 0$, then we obtain a quadratic differential $\langle T(z) \rangle_\varphi dz^2$.

REMARK 5.2. — Let V be a vertex algebra with $\widehat{\mathfrak{g}}$ -structure of level $k \neq -h^\vee$, and \mathcal{P} a G -bundle on X . A conformal block *twisted by \mathcal{P}* is by definition a linear functional φ on $\mathcal{V}_x^{\mathcal{P}}$, such that $\varphi(\mathcal{Y}_x^{\mathcal{P}} \cdot A)$ (see §4.4) can be extended to a regular section of $(\mathcal{V}^{\mathcal{P}})^*$ on $X \setminus x$ for all $A \in \mathcal{V}_x^{\mathcal{P}}$. We denote the corresponding space by $C^{\mathcal{P}}(X, x, V)$.

5.2. Examples

In the case of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$ the Definition 5.1 can be simplified. In this case the subspace $(\mathfrak{g} \otimes t^{-1})v_k$ is $\text{Aut } \mathcal{O}$ -invariant, and therefore we have a surjection $(\mathcal{V}_k(\mathfrak{g}))^* \rightarrow \mathfrak{g}^* \otimes \Omega$. Hence for each $\varphi \in C(X, x, \pi)$, $A \in \mathcal{V}_x$, we obtain a regular \mathfrak{g}^* -valued one-form on $X \setminus x$, whose restriction to D_x^\times is $\varphi(J(z) \cdot A)dz$, where $J(z) = \sum_a J_a \otimes J^a(z)$.

Now observe that any $f = \sum f_n z^n \in \mathfrak{g} \otimes \mathcal{K}_x$ gives rise to an operator on $\mathcal{V}_k(\mathfrak{g})_x$, $\widetilde{f} = \text{Res}_x(f, J(z))dz = \sum f_n b_n$, which does not depend on the choice of the coordinate z . Since by definition $\varphi(J(z) \cdot A)dz$ extends to a regular one-form on $X \setminus x$, we obtain from

the residue theorem that $\varphi(\tilde{f} \cdot A) = 0$ for all $f \in \mathfrak{g}_{x,\text{out}} = \mathfrak{g} \otimes \mathbb{C}[X \setminus x]$. Hence we obtain a map from $C(X, x, V_k(\mathfrak{g}))$ to the space of $\mathfrak{g}_{x,\text{out}}$ -invariant functionals on $\mathcal{V}_k(\mathfrak{g})_x$.

LEMMA 5.3. — The space of conformal blocks $C(X, x, V_k(\mathfrak{g}))$ is isomorphic to the space of $\mathfrak{g}_{x,\text{out}}$ -invariant functionals on $\mathcal{V}_k(\mathfrak{g})_x$.

Thus we recover the common definition of conformal blocks as $\mathfrak{g}_{x,\text{out}}$ -invariants. They have been extensively studied recently because of the relation to the moduli spaces of G -bundles on curves (see §6.4 below). More generally, we have: $C^{\mathcal{P}}(X, x, V_k(\mathfrak{g})) = \text{Hom}_{\mathfrak{g}_{x,\text{out}}^{\mathcal{P}}}(\mathcal{V}_k(\mathfrak{g})_x^{\mathcal{P}}, \mathbb{C})$, where $\mathfrak{g}_{x,\text{out}}^{\mathcal{P}}$ is the Lie algebra of sections of $\mathcal{P} \times \mathfrak{g}$ over $X \setminus x$.

The situation becomes more subtle in the case of the Virasoro vertex algebra Vir_c . The analogue of $\mathfrak{g}_{x,\text{out}}$ is then the Lie algebra $\text{Vect}(X \setminus x)$ of vector fields on $X \setminus x$. By analogy with the Kac-Moody case, we would like to define a homomorphism $\text{Vect}(X \setminus x) \rightarrow \text{End } \mathcal{V}_{c,x}$ sending $\xi(z)\partial_z \in \text{Vect}(X \setminus x)$ to $\text{Res}_x \xi(z)T(z)dz$. But it is not clear whether this map is independent of the choice of formal coordinate z , unless $c = 0$ (when the field $T(z)$ transforms as a quadratic differential). According to §4.3, if $c \neq 0$, then $\partial_z^2 + \frac{6}{c}T(z)$ is naturally a section of a rank two bundle \mathcal{T}_c^* . Hence it can be paired with sections of the rank two bundle $\mathcal{T}_c \otimes \Omega$ (we could split this bundle by choosing a projective structure on X , but we prefer not to do that). One can show that the sheaf $\widehat{\mathcal{T}}_c = (\mathcal{T}_c \otimes \Omega)/d\mathcal{O}$, which is an extension of Θ by $\Omega/d\mathcal{O}$, has a canonical Lie algebra structure (see [Wi, BS]), such that $\text{Vir}_x = \Gamma(D_x^\times, \widehat{\mathcal{T}}_c)$ is isomorphic to the Virasoro algebra and acts on $\mathcal{V}_{c,x}$. The map $\Gamma(X \setminus x, \widehat{\mathcal{T}}_c) \rightarrow \text{Vir}_x$ factors through $\text{Vect}(X \setminus x)$, and hence the above map $\text{Vect}(X \setminus x) \rightarrow \text{End } \mathcal{V}_{c,x}$ is well-defined. One can then show that the space of conformal blocks $C(X, x, \text{Vir}_c)$ is isomorphic to the space of $\text{Vect}(X \setminus x)$ -invariant functionals on $\mathcal{V}_{c,x}$.

5.3. General invariance condition

For general vertex algebras the analogue of the above invariance condition is defined as follows. Let $\mathcal{V} \xrightarrow{\nabla} \mathcal{V} \otimes \Omega$, where $\mathcal{V} \otimes \Omega$ is placed in degree 0, be the de Rham complex of the flat vector bundle \mathcal{V} , considered as a complex of sheaves on the curve X . Here ∇ is the connection operator defined in §4.2 (recall that $\nabla = d + L_{-1} \otimes dz$). For $\Sigma \subset X$, denote by $H_{\text{dR}}^0(\Sigma, \mathcal{V} \otimes \Omega)$ the degree 0 cohomology of the restriction of this complex to Σ . If Σ is affine or D_x^\times , then $H_{\text{dR}}^0(\Sigma, \mathcal{V} \otimes \Omega)$ is simply the quotient of $\Gamma(\Sigma, \mathcal{V} \otimes \Omega)$ by the image of ∇ .

There is a linear map $\gamma : \Gamma(D_x^\times, \mathcal{V} \otimes \Omega) \rightarrow \text{End } \mathcal{V}_x$ sending $\mu \in \Gamma(D_x^\times, \mathcal{V} \otimes \Omega)$ to a linear operator $O_\mu = \text{Res}_x \langle \mathcal{Y}_x, \mu \rangle$ on \mathcal{V}_x . Since the residue of a total derivative vanishes, γ factors through $U(\mathcal{V}_x) = H_{\text{dR}}^0(D_x^\times, \mathcal{V} \otimes \Omega)$. If we choose a formal coordinate z at x , then we can identify $U(\mathcal{V}_x)$ with $U(V)$, the completion of the span of all Fourier coefficients of all vertex operators $Y(A, z), A \in V$. According to formula (9), $U(\mathcal{V}_x)$ is a Lie algebra.

Note that in the case of Kac-Moody and Virasoro vertex algebras, $U(\mathcal{V}_x)$ contains $\widehat{\mathfrak{g}}_x$ and Vir_x , respectively, as Lie subalgebras.

Given an open $\Sigma \subset X$, we have a canonical map $H_{\text{dR}}^0(\Sigma, \mathcal{V} \otimes \Omega) \rightarrow U(\mathcal{V}_x)$. Denote its image by $U_\Sigma(\mathcal{V}_x)$. It follows from the Beilinson-Drinfeld chiral algebra formalism (see §7) that $U_\Sigma(\mathcal{V}_x)$ is a Lie subalgebra of $U(\mathcal{V}_x)$. The image of $U_\Sigma(\mathcal{V}_x)$ in $\text{End } \mathcal{V}_x$ is the span of all operators O_μ for $\mu \in \Gamma(\Sigma, \mathcal{V} \otimes \Omega)$. The residue theorem implies:

PROPOSITION 5.4. — *The space of conformal blocks $C(X, x, V)$ coincides with the space of $U_{X \setminus x}(\mathcal{V}_x)$ -invariant functionals on \mathcal{V}_x .*

Therefore the dual space to the space of conformal blocks is the *space of coinvariants* $H(X, x, V) = \mathcal{V}_x / U_{X \setminus x}(\mathcal{V}_x) \cdot \mathcal{V}_x$.

The definition of the spaces $C(X, x, V)$ and $H(X, x, V)$ can be generalized to the case of multiple (distinct) points x_1, \dots, x_n on the curve X , at which we “insert” arbitrary conformal V -modules M_1, \dots, M_n . A V -module M is called conformal if the gradation operator on M coincides, up to a shift, with the Fourier coefficient L_0^M of the field $Y_M(\omega, z)$. Assume for simplicity that the eigenvalues of L_0^M on all modules M_i are integers. Then each M_i is an $\text{Aut } \mathcal{O}$ -module, and so we can attach to it a vector bundle \mathcal{M}_i on X with a flat connection (a general conformal V -module gives rise to a vector bundle with connection on the space of pairs (x, τ_x) , where $x \in X$ and τ_x is a non-zero tangent vector to X at x).

The space of conformal blocks $C_V(X, (x_i), (M_i))_{i=1}^n$ is then by definition the space of linear functionals φ on $\mathcal{M}_{1, x_1} \otimes \dots \otimes \mathcal{M}_{n, x_n}$, such that for any $A_i \in \mathcal{M}_{i, x_i}, i = 1, \dots, n$, there exists a section of \mathcal{V}^* over $X \setminus \{x_1, \dots, x_n\}$, whose restriction to each $D_{x_i}^\times$ equals

$$\varphi(A_1 \otimes \dots \otimes \mathcal{Y}_{M_i} \cdot A_i \otimes \dots \otimes A_n).$$

Informally, one can say that the local sections of \mathcal{V}^* over the discs around the points, obtained by acting with vertex operators at those points, can be “glued together” into a single meromorphic section of \mathcal{V}^* . There is a canonical isomorphism

$$(19) \quad C_V(X, (x_1, \dots, x_n, y), (M_1, \dots, M_n, V)) \simeq C_V(X, (x_1, \dots, x_n), (M_1, \dots, M_n)),$$

given by $\varphi \mapsto \varphi|_{M_1 \otimes \dots \otimes M_n \otimes |0\rangle}$.

REMARK 5.5. — Let V be a rational vertex algebra (see §3.4). Then we obtain a functor from the category of pointed algebraic curves with insertions of simple V -modules at the points to the category of vector spaces, which sends $(X, (x_i), (M_i))$ to $C_V(X, (x_i), (M_i))_{i=1}^n$. This is a version of *modular functor* corresponding to V [Se, Ga, Z2]. It is known that in the case of rational vertex algebras $L_k(\mathfrak{g})$ or $L_{c(p,q)}$ the spaces of conformal blocks are finite-dimensional. Moreover, the functor can be extended to the category of pointed stable curves with insertions. It then satisfies a factorization property, which expresses the space of conformal blocks associated to a singular curve with a double point in terms of conformal blocks associated to its normalization. The same is expected to be true for general rational vertex algebras.

5.4. **Functional realization**

The spaces $C_V(X, (x_i), (M_i)_{i=1}^n)$ with varying points x_1, \dots, x_n can be organized into a vector bundle on $\overset{\circ}{X}^n := X^n \setminus \Delta$ (where Δ is the union of all diagonals) with a flat connection, which is a subbundle of $(\mathcal{M}_1 \boxtimes \dots \boxtimes \mathcal{M}_n)^*$. Consider for example the spaces $C_V(X, (x_i), (V)_{i=1}^n)$. By (19), all of them are canonically isomorphic to each other and to $C_V(X, x, V)$. Hence the corresponding bundles of conformal blocks are canonically trivialized. For $\varphi \in C_V(X, x, V)$, let φ_n denote the corresponding element of $C_V(X, (x_i), (V)_{i=1}^n)$. Let $A_i(x_i)$ be local sections of \mathcal{V} near x_i . Evaluating our conformal blocks on them, we obtain what physicists call the chiral correlation functions

$$(20) \quad \varphi_n(A_1(x_1) \otimes \dots \otimes A_n(x_n)) \sim \langle A_1(x_1) \dots A_n(x_n) \rangle_\varphi.$$

Moreover, we can explicitly describe these functions when x_i 's are very close to each other, in the neighborhood of a point $x \in X$: if we choose a formal coordinate z at x and denote the coordinate of the point x_i by z_i , then

$$(21) \quad \varphi_n(A_1(z_1) \otimes \dots \otimes A_n(z_n)) = \varphi(Y(A_1, z_1) \dots Y(A_n, z_n)|0).$$

We obtain that the restriction of the n -point chiral correlation function to $\overset{\circ}{D}_x^n = D_x^n \setminus \Delta := \text{Spec } \mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}$ coincides with the n -point function corresponding to the functional φ introduced in §2.7.

According to Remark 4.2, the matrix elements given by the right hand side of formula (21) give rise to a horizontal section of $(\mathcal{V}^{\boxtimes n})^*$ on $\overset{\circ}{D}_x^n$. Therefore by Proposition 2.7 we obtain an embedding $\mathcal{V}_x^* \hookrightarrow \Gamma_{\nabla}(\overset{\circ}{D}_x^n, (\mathcal{V}^{\boxtimes n})^*)$, where Γ_{∇} stands for the space of horizontal sections. If $\varphi \in \mathcal{V}_x^*$ is a conformal block, then Definition 5.1 implies that the image of $\varphi \in C_V(X, x, V)$ in $\Gamma_{\nabla}(\overset{\circ}{D}_x^n, (\mathcal{V}^{\boxtimes n})^*)$ extends to a horizontal section φ_n of $(\mathcal{V}^{\boxtimes n})^*$ over $\overset{\circ}{X}^n$. Thus, we obtain an embedding $C_V(X, x, V) \hookrightarrow \Gamma(\overset{\circ}{X}^n, (\mathcal{V}^{\boxtimes n})^*) = \Gamma_{\nabla}(X^n, j_*j^*(\mathcal{V}^{\boxtimes n})^*)$, where $j : \overset{\circ}{X}^n \hookrightarrow X^n$.

Imposing the bootstrap conditions on the diagonals from Proposition 2.7, we obtain a quotient \mathcal{V}_n^* of $j_*j^*(\mathcal{V}^{\boxtimes n})^*$ (the precise definition of \mathcal{V}_n^* uses the operation $\mathfrak{Y}^{(2)}$ introduced in Theorem 7.1). Since by construction our sections satisfy the bootstrap conditions, we have embeddings $\mathcal{V}_x^* \hookrightarrow \Gamma_{\nabla}(D_x^n, \mathcal{V}_n^*)$ and $C_V(X, x, V) \hookrightarrow \Gamma_{\nabla}(X^n, \mathcal{V}_n^*)$. A remarkable fact is that these maps are isomorphisms for all $n > 1$. Thus, using correlation functions we obtain *functional realizations* of the (dual space of) a vertex algebra and its spaces of conformal blocks.

REMARK 5.6. — Note that neither $(\mathcal{V}^{\boxtimes n})^*$ nor \mathcal{V}_n^* is quasicohherent as an \mathcal{O}_{X^n} -module. It is more convenient to work with the dual sheaves $\mathcal{V}^{\boxtimes n}$ and \mathcal{V}_n , which are quasicohherent. The sheaf \mathcal{V}_n on X^n is defined by Beilinson and Drinfeld in their construction of factorization algebra corresponding to \mathcal{V} (see §7). Note that the space $\Gamma_{\nabla}(X, \mathcal{V}_n^*)$ is dual to the top deRham cohomology $H_{\text{dR}}^n(X, \mathcal{V}_n \otimes \Omega_{X^n})$ (which is therefore isomorphic to $H(X, x, V)$).

In the case of Heisenberg algebra, the functional realization can be simplified, because we can use the sections of the exterior products of the canonical bundle Ω instead of the bundle Π^* (here we choose $\frac{1}{2}b_{-1}^2$ as the conformal vector, so that $b(z)dz$ transforms as a one-form). Namely, as in §2.7 we assign to $\varphi \in \Pi_x^*$ the collection of polydifferentials

$$\varphi(b(z_1) \dots b(z_n)|0\rangle) dz_1 \dots dz_n.$$

These are sections of $\Omega^{\boxtimes n}(2\Delta)$ over D_x^n , which are symmetric and satisfy the bootstrap condition (12). Let Ω_∞ be the sheaf on X , whose sections over $U \subset X$ are collections of sections of $\Omega^{\boxtimes n}(2\Delta)$ over U^n , for $n \geq 0$, satisfying these conditions (the bootstrap condition makes sense globally because of the identification $\Omega^{\boxtimes 2}(2\Delta)/\Omega^{\boxtimes 2}(\Delta) \simeq \mathcal{O}$). Then we have canonical isomorphisms $\Pi_x^* \simeq \Gamma(D_x, \Omega_\infty)$, $C(X, x, \pi) \simeq \Gamma(X, \Omega_\infty)$.

The same construction can be applied to the Kac-Moody vertex algebras, see [BD1].

REMARK 5.7. — If we consider conformal blocks with general V -module insertions, then the horizontal sections of the corresponding flat bundle of conformal blocks over $\overset{\circ}{X}^n$ will have non-trivial monodromy around the diagonals. For example, in the case of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$ and $X = \mathbb{P}^1$, these sections are solutions of the Knizhnik-Zamolodchikov equations, and the monodromy matrices are given by the R -matrices of the quantum group $U_q(\mathfrak{g})$, see [TK, SV2].

6. SHEAVES OF CONFORMAL BLOCKS ON MODULI SPACES

In the previous section we associated the space of conformal blocks to a vertex algebra V , a smooth curve X and a point x of X . Now we want to understand the behavior of these spaces as we move x along X and vary the complex structure on X . More precisely, we wish to organize the spaces of conformal blocks into a sheaf on the moduli space $\mathfrak{M}_{g,1}$ of smooth pointed curves of genus g (here and below by moduli space we mean the corresponding moduli stack in the smooth topology; \mathcal{D} -modules on algebraic stacks are defined in [BB, BD2]). Actually, for technical reasons we prefer to work with the spaces of coinvariants (also called covacua), which are dual to the spaces of conformal blocks. One can construct in a straightforward way a quasicoherent sheaf on $\mathfrak{M}_{g,1}$, whose fiber at (X, x) is the space of coinvariants attached to (X, x) . But in fact this sheaf also carries a structure of (twisted) \mathcal{D} -module on $\mathfrak{M}_{g,1}$, i.e., we can canonically identify the (projectivizations of) the spaces of coinvariants attached to infinitesimally nearby points of $\mathfrak{M}_{g,1}$ (actually, this \mathcal{D} -module always descends to \mathfrak{M}_g). The key fact used in the proof of this statement is the “Virasoro uniformization” of $\mathfrak{M}_{g,1}$. Namely, the Lie algebra $\text{Der } \mathcal{K}$ acts transitively on the moduli space of triples (X, x, z) , where (X, x) are as above and z is a formal coordinate at x ; this action is obtained by “gluing” X from D_x and $X \setminus x$ (see Theorem 6.1).

In the case when the vertex algebra V is rational, the corresponding twisted \mathcal{D} -module on \mathfrak{M}_g is believed to be coherent, i.e., isomorphic to the sheaf of sections of a vector bundle of finite rank equipped with a projectively flat connection (this picture was first suggested by Friedan and Shenker [FrS]). This is known to be true in the case of the Kac-Moody vertex algebra $L_k(\mathfrak{g})$ [TUY] (see also [So]) and the minimal models of the Virasoro algebra [BFM]. Moreover, in those cases the \mathcal{D} -module on \mathfrak{M}_g can be extended to a \mathcal{D} -module with regular singularities on the Deligne-Mumford compactification $\overline{\mathfrak{M}}_g$, and the dimensions of the fibers at the boundary are equal to those in \mathfrak{M}_g . This allows one to compute these dimensions by “Verlinde formula” from the dimensions of the spaces of coinvariants attached to \mathbb{P}^1 with three points. The latter numbers, called the fusion rules, can in turn be found from the matrix of the modular transformation $\tau \mapsto -1/\tau$ acting on the space of characters (see Theorem 3.2). The same pattern is believed to hold for other rational vertex algebras, but as far as I know this has not been proved in general.

The construction of the \mathcal{D} -module structure on the sheaf of coinvariants is a special case of the general formalism of localization of modules over Harish-Chandra pairs, due to Beilinson-Bernstein [BB], which we now briefly recall.

6.1. Generalities on localization

A *Harish-Chandra pair* is a pair (\mathfrak{g}, K) where \mathfrak{g} is a Lie algebra, K is a Lie group, such that $\mathfrak{k} = \text{Lie } K$ is embedded into \mathfrak{g} , and an action Ad of K on \mathfrak{g} compatible with the adjoint action of K on $\mathfrak{k} \subset \mathfrak{g}$ and the action of \mathfrak{k} on \mathfrak{g} .

A (\mathfrak{g}, K) -*action* on a scheme Z is the data of an action of \mathfrak{g} on Z (that is, a homomorphism $\rho : \mathfrak{g} \rightarrow \Theta_Z$), together with an action of K on Z , such that (1) the differential of the K -action is the restriction of the \mathfrak{g} -action to \mathfrak{k} , and (2) $\rho(Ad_k(a)) = k\rho(a)k^{-1}$. A (\mathfrak{g}, K) -action is called *transitive* if the map $\mathfrak{g} \otimes \mathcal{O}_Z \rightarrow \Theta_Z$ is surjective, and *simply transitive* if this map is an isomorphism.

A (\mathfrak{g}, K) -*structure* on a scheme S is a principal K -bundle $\pi : Z \rightarrow S$ together with a simply transitive (\mathfrak{g}, K) -action on Z which extends the fiberwise action of K .

An example of (\mathfrak{g}, K) -structure is the homogeneous space $S = G/K$, where G is a finite-dimensional Lie group, $\mathfrak{g} = \text{Lie } G$, K is a Lie subgroup of G , and we take the obvious right action of (\mathfrak{g}, K) on $\widehat{X} = G$. In the finite-dimensional setting, any space with a (\mathfrak{g}, K) -structure is locally of this form. However, in the infinite-dimensional setting this is no longer so. An example is the $(\text{Der } \mathcal{O}, \text{Aut } \mathcal{O})$ -structure $Z = \widehat{X} \rightarrow X$ on a smooth curve X , which we used above. This example can be generalized to the case when S is an arbitrary smooth scheme. Define \widehat{S} to be the scheme of pairs (x, \vec{t}_x) , where $x \in S$ and \vec{t}_x is a system of formal coordinates at x . Set $\mathfrak{g} = \text{Der } \mathbb{C}[[z_1, \dots, z_n]]$, $K = \text{Aut } \mathbb{C}[[z_1, \dots, z_n]]$. Then \widehat{S} is a (\mathfrak{g}, K) -structure on S , first considered by Gelfand, Kazhdan and Fuchs [GKF].

Let V be a (\mathfrak{g}, K) -module, i.e., a vector space together with actions of \mathfrak{g} and K satisfying the obvious compatibility condition. Then we define a flat vector bundle \mathcal{V} on any variety S with a (\mathfrak{g}, K) -structure Z . Namely, as a vector bundle, $\mathcal{V} = Z \times_K V$. Since by assumption the \mathfrak{g} -action on Z is simply transitive, it gives rise to a flat connection on the trivial vector bundle $Z \times V$ over Z , which descends to \mathcal{V} , because the actions of K and \mathfrak{g} are compatible. In the special case of the $(\text{Der } \mathcal{O}, \text{Aut } \mathcal{O})$ -structure \widehat{X} over a smooth curve X we obtain the flat bundle \mathcal{V} on X from §4.2.

Now consider the case when the \mathfrak{g} -action is transitive but not simply transitive (there are stabilizers). Then we can still construct a vector bundle \mathcal{V} on S equipped with a K -equivariant action of the Lie algebroid $\mathfrak{g}_Z = \mathfrak{g} \otimes \mathcal{O}_Z$, but we cannot obtain an action of Θ_Z on \mathcal{V} because the map $a : \mathfrak{g}_Z \rightarrow \Theta_Z$ is no longer an isomorphism. Nevertheless, Θ_Z will act on any sheaf, on which $\text{Ker } a$ acts by 0. In particular, the sheaf $V \otimes \mathcal{O}_Z / \text{Ker } a \cdot (V \otimes \mathcal{O}_Z) \simeq \mathcal{D}_Z \otimes_{U_{\mathfrak{g}}} V$ gives rise to a K -equivariant \mathcal{D} -module on Z , and hence to a \mathcal{D} -module on S , denoted $\Delta(V)$ and called the *localization* of V on S .

More generally, suppose that V is a module over a Lie algebra \mathfrak{l} , which contains \mathfrak{g} as a Lie subalgebra and carries a compatible K -action. Suppose also that we are given a K -equivariant Lie algebra subsheaf $\widetilde{\mathfrak{l}}$ of the constant sheaf of Lie algebras $\mathfrak{l} \otimes \mathcal{O}_Z$, which is preserved by the natural action of the Lie algebroid \mathfrak{g}_Z . Then if $\widetilde{\mathfrak{l}}$ contains $\text{Ker } a$, the sheaf $V \otimes \mathcal{O}_Z / \widetilde{\mathfrak{l}} \cdot (V \otimes \mathcal{O}_Z)$ is a K -equivariant \mathcal{D} -module on Z , which descends to a \mathcal{D} -module $\widetilde{\Delta}(V)$ on S .

Let us describe the fibers of $\widetilde{\Delta}(V)$ (considered as an \mathcal{O}_S -module). For $s \in S$, let Z_s be the fiber of Z at s , and $\mathfrak{l}_s = Z_s \times_K \mathfrak{l}$, $\mathcal{V}_s = Z_s \times_K V$. The fibers of $\widetilde{\mathfrak{l}}$ at the points of $Z_s \subset M$ give rise to a well-defined Lie subalgebra $\widetilde{\mathfrak{l}}_s$ of \mathfrak{l}_s . Then the fiber of $\widetilde{\Delta}(V)$ at $s \in S$ is canonically isomorphic to the space of coinvariants $\mathcal{V}_s / \widetilde{\mathfrak{l}}_s \cdot \mathcal{V}_s$.

6.2. Localization on the moduli space of curves

We apply the above formalism in the case when S is the moduli space $\mathfrak{M}_{g,1}$ of smooth pointed curves of genus $g > 1$, and Z is the moduli space $\widehat{\mathfrak{M}}_{g,1}$ of triples (X, x, z) , where $(X, x) \in \mathfrak{M}_{g,1}$ and z is a formal coordinate at x . Clearly, $\widehat{\mathfrak{M}}_{g,1}$ is an $\text{Aut } \mathcal{O}$ -bundle over $\mathfrak{M}_{g,1}$.

THEOREM 6.1 (cf. [ADKP, BS, Ko, TUY]). — $\widehat{\mathfrak{M}}_{g,1}$ carries a transitive action of $\text{Der } \mathcal{K}$ compatible with the $\text{Aut } \mathcal{O}$ -action along the fibers.

The action of the corresponding ind-group $\text{Aut } \mathcal{K}$ is defined by the “gluing” construction. If (X, x, z) is an R -point of $\widehat{\mathfrak{M}}_{g,1}$, and $\rho \in \text{Aut } \mathcal{K}(R)$, we construct a new R -point (X_ρ, x_ρ, z_ρ) of $\widehat{\mathfrak{M}}_{g,1}$ by “gluing” the formal neighborhood of x in X with $X \setminus x$ with a “twist” by ρ .

Now we are in the situation of §6.1, with $\mathfrak{g} = \text{Der } \mathcal{K}$ and $K = \text{Aut } \mathcal{O}$. Denote by \mathcal{A} the Lie algebroid $\text{Der } \mathcal{K} \widehat{\otimes} \mathcal{O}_{\widehat{\mathfrak{M}}_{g,1}}$ on $\widehat{\mathfrak{M}}_{g,1}$. By construction, the kernel of the corresponding

homomorphism $a : \mathcal{A} \rightarrow \Theta$ is the subsheaf \mathcal{A}_{out} of \mathcal{A} , whose fiber at $(X, x) \in \mathfrak{M}_{g,1}$ is $\text{Vect}(X \setminus x)$. Applying the construction of §6.1 to $V = \text{Vir}_0$ (the Virasoro vertex algebra with $c = 0$), we obtain a \mathcal{D} -module $\Delta(\text{Vir}_0)$ on $\mathfrak{M}_{g,1}$, whose fibers are the spaces of coinvariants $\text{Vir}_0 / \text{Vect}(X \setminus x) \cdot \text{Vir}_0$.

More generally, let V be an arbitrary conformal vertex algebra with central charge 0. Then it is a module over the Lie algebra $\mathfrak{l} = U(V)$ from §5.3. Let $\tilde{\mathfrak{l}} = \mathcal{U}(V)_{\text{out}}$ be the subsheaf of $U(V) \hat{\otimes} \mathcal{O}_{\widehat{\mathfrak{M}}_{g,1}}$, whose fiber at $(X, x, z) \in \mathfrak{M}_{g,1}$ equals $U_{X \setminus x}(V)$ (see §5.3). Note that $U(V)$ contains $\text{Der } \mathcal{K}$, and $\mathcal{U}(V)_{\text{out}}$ contains \mathcal{A}_{out} . Moreover, we have:

LEMMA 6.2. — $\mathcal{U}(V)_{\text{out}}$ is preserved by the action of the Lie algebroid \mathcal{A} .

This is equivalent to the statement that $U_{X \setminus x}(V)$ and $U_{X_\rho \setminus x_\rho}(V)$ are conjugate by ρ , which follows from the fact that formula (16) is valid for any $\rho \in \text{Aut } \mathcal{K}$.

So we are again in the situation of §6.1, and hence we obtain a \mathcal{D} -module $\tilde{\Delta}(V)$ on $\mathfrak{M}_{g,1}$, whose fiber at (X, x) is precisely the space of coinvariants $H(X, x, V) = \mathcal{V}_x / U_{X \setminus x}(\mathcal{V}_x) \cdot \mathcal{V}_x$ (see §5.3), which is what we wanted.

In the case of vertex algebras with non-zero central charge, we need to modify the general construction of §6.1 as follows. Suppose that \mathfrak{g} has a central extension $\widehat{\mathfrak{g}}$ which splits over \mathfrak{k} , and such that the extension

$$(22) \quad 0 \rightarrow \mathcal{O}_Z \rightarrow \widehat{\mathfrak{g}} \otimes \mathcal{O}_Z \rightarrow \mathfrak{g}_Z \rightarrow 0$$

splits over the kernel of $a : \mathfrak{g}_Z \rightarrow \Theta_Z$. Then the quotient \mathfrak{g}'_Z of $\widehat{\mathfrak{g}} \otimes \mathcal{O}_Z$ by $\text{Ker } a$ is an extension of Θ_Z by \mathcal{O}_Z , with a natural Lie algebroid structure. The enveloping algebra of this Lie algebroid is a sheaf \mathcal{D}'_Z of twisted differential operators (TDO) on Z , see [BB]. Further, \mathfrak{g}'_Z descends to a Lie algebroid on S , and gives rise to a sheaf of TDO \mathcal{D}'_S .

Now if V is a $(\widehat{\mathfrak{g}}, K)$ -module, we obtain a \mathcal{D}'_Z -module $\mathcal{D}'_Z \otimes_{U\widehat{\mathfrak{g}}} V$. By construction, it is K -equivariant, and hence descends to a \mathcal{D}'_S -module on S . More generally, suppose that $\widehat{\mathfrak{g}}$ is a Lie subalgebra of $\widehat{\mathfrak{l}}$, and we are given a subsheaf $\tilde{\mathfrak{l}}$ of $\widehat{\mathfrak{l}} \otimes \mathcal{O}_Z$, containing $\text{Ker } a$ and preserved by the action of $\widehat{\mathfrak{g}} \otimes \mathcal{O}_Z$. Then the sheaf $V \otimes \mathcal{O}_Z / \tilde{\mathfrak{l}} \cdot V \otimes \mathcal{O}_Z$ is a K -equivariant \mathcal{D}'_Z -module on Z , which descends to a \mathcal{D}'_S -module (still denoted by $\tilde{\Delta}(V)$). Its fibers are the coinvariants $\mathcal{V}_s / \tilde{\mathfrak{l}}_s \cdot \mathcal{V}_s$.

Let V be a conformal vertex algebra with an arbitrary central charge c . Then by the residue theorem the corresponding sequence (22) with $\widehat{\mathfrak{g}} = \text{Vir}$ splits over $\text{Ker } a = \mathcal{A}_{\text{out}}$. Hence there exists a TDO sheaf \mathcal{D}_c on $\mathfrak{M}_{g,1}$ and a \mathcal{D}_c -module $\tilde{\Delta}(V)$ on $\mathfrak{M}_{g,1}$ whose fiber at (X, x) is the space of coinvariants $H(X, x, V)$. Actually, $\tilde{\Delta}(V)$ descends to \mathfrak{M}_g . Explicit computation shows that the sheaf of differential operators on the determinant line bundle over \mathfrak{M}_g corresponding to the sheaf of relative λ -differentials on the universal curve is $\mathcal{D}_{c(\lambda)}$, where $c(\lambda) = -12\lambda^2 + 12\lambda - 2$ (see [BS, BFM]).

It is straightforward to generalize the above construction to the case of multiple points with arbitrary V -module insertions M_1, \dots, M_n . In that case we obtain a twisted \mathcal{D} -module on the moduli space $\mathfrak{M}'_{g,n}$ of n -pointed curves with non-zero tangent vectors at the points.

6.3. Localization on other moduli spaces

In the previous section we constructed twisted \mathcal{D} -modules on the moduli spaces of pointed curves using the interior action of the Harish-Chandra pair $(Vir, \text{Aut } \mathcal{O})$ on conformal vertex algebras and its modules. This construction can be generalized to the case of other moduli spaces if we consider interior actions of other Harish-Chandra pairs.

For example, consider the Kac-Moody vertex algebra $V_k(\mathfrak{g})$. It carries an action of the Harish-Chandra pair $(\widehat{\mathfrak{g}}, G(\mathcal{O}))$, where G is the simply-connected group with Lie algebra \mathfrak{g} . The corresponding moduli space is the moduli space $\mathfrak{M}_G(X)$ of G -bundles on a curve X . For $x \in X$, let $\widehat{\mathfrak{M}}_G(X)$ be the moduli spaces of pairs (\mathcal{P}, s) , where \mathcal{P} is a G -bundle on X , and s is a trivialization of $\mathcal{P}|_{D_x}$. Then $\widehat{\mathfrak{M}}_G(X)$ is a principal $G(\mathcal{O}_x)$ -bundle over $\mathfrak{M}_G(X)$, equipped with a compatible $\mathfrak{g}(\mathcal{K}_x)$ -action. The latter is given by the gluing construction similar to that explained in the proof of Theorem 6.1.

Now we can apply the construction of §6.1 to the $(\widehat{\mathfrak{g}}_x, G(\mathcal{O}_x))$ -module $\mathcal{V}_k(\mathfrak{g})_x$, and obtain a twisted \mathcal{D} -module on $\mathfrak{M}_G(X)$. Its fiber at $\mathcal{P} \in \mathfrak{M}_G(X)$ is the space of coinvariants $\mathcal{V}_k(\mathfrak{g})_x / \mathfrak{g}_{x,\text{out}}^{\mathcal{P}} \cdot \mathcal{V}_k(\mathfrak{g})_x$, where $\mathfrak{g}_{x,\text{out}}^{\mathcal{P}}$ is the Lie algebra of sections of $\mathcal{P} \times_{\widehat{\mathfrak{g}}_x} \mathfrak{g}$ over $X \setminus x$.

More generally, let V be a vertex algebra equipped with a $\widehat{\mathfrak{g}}$ -structure of level $k \neq -h^\vee$ (see §4.4). Then V carries an action of the Harish-Chandra pair $(\widehat{\mathfrak{g}}_x, G(\mathcal{O}_x))$. In the same way as in the case of the Virasoro algebra, we attach to V a twisted \mathcal{D} -module on $\mathfrak{M}_G(X)$, whose fiber at \mathcal{P} is the twisted space of coinvariants $H^{\mathcal{P}}(X, x, V)$, dual to the space $C^{\mathcal{P}}(X, x, V)$ defined in Remark 5.2.

For the Heisenberg vertex algebra π , the corresponding moduli space is the Jacobian $J(X)$ of X . The bundle $\widehat{J}(X)$ over $J(X)$, which parameterizes line bundles together with trivializations on D_x , carries a transitive action of the abelian Lie algebra \mathcal{K} . Applying the above construction, we assign to any conformal vertex algebra equipped with an embedding $\pi \rightarrow V$ a twisted \mathcal{D} -module on $J(X)$, whose fiber at $\mathcal{L} \in J(X)$ is the twisted space of coinvariants $H^{\mathcal{L}}(X, x, V)$. The corresponding TDO sheaf is the sheaf of differential operators acting on the theta line bundle on $J(X)$.

By considering the action of the semi-direct product $Vir \rtimes \widehat{\mathfrak{g}}$ on V we obtain a twisted \mathcal{D} -module on the moduli space of curves and G -bundles on them. We can further generalize the construction by inserting modules at points of the curve, etc.

6.4. Local and global structure of moduli spaces

In the simplest cases the twisted \mathcal{D} -modules obtained by the above localization construction are the twisted sheaves \mathcal{D} themselves (considered as left modules over themselves). For example, the sheaf $\Delta(V_k(\mathfrak{g}))$ on $\mathfrak{M}_G(X)$ is the sheaf \mathcal{D}_k of differential operators acting on the k th power of the determinant line bundle \mathcal{L}_G on $\mathfrak{M}_G(X)$ (the ample generator of the Picard group of $\mathfrak{M}_G(X)$). The dual space to the stalk of \mathcal{D}_k at $\mathcal{P} \in \mathfrak{M}_G(X)$ is canonically identified with the space of sections of $\mathcal{L}_G^{\otimes k}$ on the formal neighborhood of \mathcal{P} in $\mathfrak{M}_G(X)$. But according to our construction, this space is also isomorphic to the space of conformal blocks $C^{\mathcal{P}}(X, x, V_k(\mathfrak{g}))$. In particular, the coordinate ring of the formal deformation space of a G -bundle \mathcal{P} on a curve X is isomorphic to $C^{\mathcal{P}}(X, x, V_0(\mathfrak{g}))$. Thus, using the description of conformal blocks in terms of correlation functions (see §5.4), we obtain a realization of the formal deformation space and a line bundle on it in terms of polydifferentials on powers of the curve X satisfying bootstrap conditions on the diagonals (see [BG, BD2, Gi]). On the other hand, if we replace X by D_x , the corresponding space of polydifferentials becomes $V_k(\mathfrak{g})^*$ (more precisely, its twist by the torsor of formal coordinates at x). Therefore the vertex algebra $V_k(\mathfrak{g})$ may be viewed as the local object responsible for deformations of G -bundles on curves.

Similarly, the Virasoro vertex algebra Vir_c and the Heisenberg vertex algebra π may be viewed as the local objects responsible for deformations of curves and line bundles on curves, respectively.

Optimistically, one may hope that any one-parameter family of “Verma module type” vertex algebras (such as Vir_c or $V_k(\mathfrak{g})$) is related to a moduli space of curves with some additional structures. It would be interesting to identify the deformation problems related to the most interesting examples of such vertex algebras (for instance, the \mathcal{W} -algebras of §3.7), and to construct the corresponding global moduli spaces. One can attach a “formal moduli space” to any vertex algebra V by taking the double quotient of the ind-group, whose Lie algebra is $U(V)$, by its “in” and “out” subgroups. But this moduli space is very big, even in the familiar examples of Virasoro and Kac-Moody algebras. The question is to construct a smaller formal moduli space with a line bundle, whose space of sections is isomorphic to the space of conformal blocks $C(X, x, V)$.

While “Verma module type” vertex algebras are related to the local structure of moduli spaces, rational vertex algebras (see §3.4) can be used to describe the global structure.

For instance, the space of conformal blocks $C(X, x, L_k(\mathfrak{g})) \simeq \text{Hom}_{\mathfrak{g}_{\text{out}}}(L_k(\mathfrak{g}), \mathbb{C})$, is isomorphic to $\Gamma_k = \Gamma(\mathfrak{M}_G(X), \mathcal{L}_G^{\otimes k})$, see [BL, Fa, KNR]. The space $C(X, x, L_k(\mathfrak{g}))$ satisfies a factorization property: its dimension does not change under degenerations of the curve into curves with nodal singularities. This property allows one to find this dimension by computing the spaces of conformal blocks in the case of \mathbb{P}^1 and three points with $L_k(\mathfrak{g})$ -module insertions (fusion rules). The resulting formula for $\dim C(X, x, L_k) = \dim \Gamma_k$ is called the Verlinde formula, see [TUY, So].

Since \mathcal{L}_G is ample, we can recover the moduli space of semi-stable G -bundles on X as the Proj of the graded ring $\bigoplus_{k \geq 0} \Gamma_{Nk} = \bigoplus_{k \geq 0} C(X, x, L_{Nk}(\mathfrak{g}))$ for large N (the ring of “non-abelian theta functions”). Thus, if we could define the product on conformal blocks in a natural way, we would obtain a description of the moduli space. Using the correlation function description of conformal blocks Feigin and Stoyanovsky [FS] have identified the space $C(X, x, L_k(\mathfrak{g}))$ with the space of sections of a line bundle on the power of X satisfying certain conditions. For example, when $\mathfrak{g} = \mathfrak{sl}_2$ it is the space of sections of the line bundle $\Omega(2nx)^{\boxtimes nk}$ over X^{nk} , which are symmetric and vanish on the diagonals of codimension k (here $x \in X$, and $n \geq g$). In these terms the multiplication $\Gamma_k \times \Gamma_l \rightarrow \Gamma_{k+l}$ is just the composition of exterior tensor product and symmetrization of these sections. Thus we obtain an explicit description of the coordinate ring of the moduli space of semi-stable SL_2 -bundles. On the other hand, replacing in the above description X by D_x , we obtain a functional realization of $L_k(\mathfrak{sl}_2)^*$.

Similarly, the space of conformal blocks corresponding to the lattice vertex superalgebra $V_{\sqrt{N}\mathbb{Z}}$ (see §3.3) may be identified with the space of theta functions of order N on $J(X)$, so we obtain the standard “functional realization” of the Jacobian $J(X)$.

6.5. Critical level

When $k = -h^\vee$ (which is called the critical level), the vertex algebra $V_k(\mathfrak{g})$ is not conformal. Nevertheless, $V_{-h^\vee}(\mathfrak{g})$ still carries an action of $\text{Der } \mathcal{O}$ which satisfies the same properties as for non-critical levels. Therefore we can still attach to V_{-h^\vee} a twisted \mathcal{D} -module on $\mathfrak{M}_G(X)$ (but we cannot vary the curve). This \mathcal{D} -module is just the sheaf \mathcal{D}' of differential operators acting on the square root of the canonical bundle on $\mathfrak{M}_G(X)$ (see [BD2]).

Recall from §3.6 that the center of a vertex algebra V is the coset vertex algebra of the pair (V, V) . This is a commutative vertex subalgebra of V . The center of $V_k(\mathfrak{g})$ is simply its subspace of $\mathfrak{g}[[z]]$ -invariants. It is easy to show that this subspace equals $\mathbb{C}v_k$ if $k \neq -h^\vee$. In order to describe the center $\mathfrak{z}(\widehat{\mathfrak{g}})$ of $V_{-h^\vee}(\mathfrak{g})$ we need the concept of *opers*, due to Beilinson and Drinfeld [BD2].

Let G_{ad} be the adjoint group of \mathfrak{g} , and B_{ad} be its Borel subgroup. By definition, a \mathfrak{g} -oper on a smooth curve X is a G_{ad} -bundle on X with a (flat) connection ∇ and a reduction to B_{ad} , satisfying a certain transversality condition [BD2]. For example, an \mathfrak{sl}_2 -oper is the same as a projective connection. The set of \mathfrak{g} -opers on X is the set of points of an affine space, denoted $\text{Op}_{\mathfrak{g}}(X)$, which is a torsor over $\bigoplus_{i=1}^{\ell} \Gamma(X, \Omega^{d_i+1})$, where d_i 's are the exponents of \mathfrak{g} . Denote by $A_{\mathfrak{g}}(X)$ the ring of functions on $\text{Op}_{\mathfrak{g}}(X)$.

The above definition can also be applied when X is replaced by the disc $D = \text{Spec } \mathbb{C}[[z]]$ (see [DS]). The corresponding ring $A_{\mathfrak{g}}(D)$ has a natural $(\text{Der } \mathcal{O}, \text{Aut } \mathcal{O})$ -action, and is therefore a commutative vertex algebra (see §2.2). It is isomorphic to a limit of the \mathcal{W} -vertex algebra $\mathcal{W}_k(\mathfrak{g})$ as $k \rightarrow \infty$ (see §3.7), and because of that it is called the classical

W-algebra corresponding to \mathfrak{g} . For example, $A_{\mathfrak{sl}_2}(D)$ is a limit of Vir_c as $c \rightarrow \infty$. The following result was conjectured by Drinfeld and proved in [FF3].

THEOREM 6.3. — *The center $\mathfrak{z}(\widehat{\mathfrak{g}})$ of $V_{-h^\vee}(\mathfrak{g})$ is isomorphic, as a commutative vertex algebra with Der \mathcal{O} -action, to $A_{L\mathfrak{g}}(D)$, where $L\mathfrak{g}$ is the Langlands dual Lie algebra to \mathfrak{g} .*

In fact, more is true: both of these commutative vertex algebras carry what can be called Poisson vertex algebra (or coisson algebra, in the terminology of [BD3]) structures, which are preserved by this isomorphism.

The localization of $\mathfrak{z}(\widehat{\mathfrak{g}})$ on $\mathfrak{M}_G(X)$ is the “constant” \mathcal{D}' -module $H(X, x, \mathfrak{z}(\widehat{\mathfrak{g}})) \otimes \mathcal{O}_{\mathfrak{M}_G(X)}$. By functoriality, it is a commutative subalgebra of \mathcal{D}' . As shown in [BD3], the space of coinvariants of a commutative vertex algebra A is canonically isomorphic to the ring of functions on $\Gamma_{\nabla}(X, \text{Spec } \mathcal{A})$. In our case we obtain: $\Gamma_{\nabla}(X, \text{Spec } \mathcal{A}_{L\mathfrak{g}}) \simeq \text{Op}_{L\mathfrak{g}}(X)$, and so $H(X, x, A_{L\mathfrak{g}}(D)) \simeq A_{L\mathfrak{g}}(X)$. Theorem 6.3 then implies:

COROLLARY 6.4. — *There is a homomorphism of algebras $A_{L\mathfrak{g}}(X) \rightarrow \Gamma(\mathfrak{M}_G(X), \mathcal{D}')$.*

Beilinson and Drinfeld have shown in [BD2] that this map is an embedding, and if G is simply-connected, it is an isomorphism. In that case, each LG -oper ρ gives us a point of the spectrum of the commutative algebra $\Gamma(\mathfrak{M}_G(X), \mathcal{D}')$, and hence a homomorphism $\tilde{\rho}: \Gamma(\mathfrak{M}_G(X), \mathcal{D}') \rightarrow \mathbb{C}$. The \mathcal{D}' -module $\mathcal{D}'/(\mathcal{D}' \cdot m_{\tilde{\rho}})$ on $\mathfrak{M}_G(X)$, where $m_{\tilde{\rho}}$ is the kernel of $\tilde{\rho}$, is holonomic. It is shown in [BD2] that the corresponding untwisted \mathcal{D} -module is a Hecke eigensheaf attached to ρ by the geometric Langlands correspondence.

6.6. Chiral deRham complex

Another application of the general localization pattern of §6.1 is the recent construction, due to Malikov, Schechtman and Vaintrob [MSV], of a sheaf of vertex superalgebras on any smooth scheme S called the chiral deRham complex. Under certain conditions, this sheaf has a purely even counterpart, the sheaf of chiral differential operators. Here we give a brief review of the construction of these sheaves.

Let Γ_N be the Weyl algebra associated to the symplectic vector space $\mathbb{C}((t))^N \oplus (\mathbb{C}((t))dt)^N$. It has topological generators $a_{i,n}, a_{i,n}^*, i = 1, \dots, N; n \in \mathbb{Z}$, with relations

$$[a_{i,n}, a_{j,m}^*] = \delta_{i,j} \delta_{n,-m}, \quad [a_{i,n}, a_{j,m}] = [a_{i,n}^*, a_{j,m}^*] = 0.$$

Denote by H_N the Fock representation of Γ_N generated by the vector $|0\rangle$, satisfying $a_{i,n}|0\rangle = 0, n \geq 0; a_{i,m}^*|0\rangle = 0, m > 0$. This is a vertex algebra, such that

$$Y(a_{i,-1}|0\rangle, z) = a(z) = \sum_{n \in \mathbb{Z}} a_{i,n} z^{-n-1}, \quad Y(a_{i,0}^*|0\rangle, z) = a^*(z) = \sum_{n \in \mathbb{Z}} a_{i,n}^* z^{-n}.$$

Let \bigwedge_N^\bullet be the fermionic analogue of H_N defined in the same way as in §3.7. This is the Fock module over the Clifford algebra with generators $\psi_{i,n}, \psi_{i,n}^*, i = 1, \dots, N, n \in \mathbb{Z}$, equipped with an additional \mathbb{Z} -gradation. It is shown in [MSV] that the vertex algebra

structure on H_N may be extended to $\widehat{H}_N = H_N \otimes_{\mathbb{C}[[a_{i,0}^*]]} \mathbb{C}[[a_{i,0}^*]]$. Let $\widehat{\Omega}_N^\bullet = \widehat{H}_N \otimes \Lambda_N^\bullet$. It carries a differential $d = \text{Res} \sum_i a_i(z) \psi_i^*(z)$, which makes it into a complex.

Denote by W_N (resp., $\mathcal{O}_N, \Omega_N^1$) the topological Lie algebra of vector fields (resp., ring of functions, module of differentials) on $\text{Spec } \mathbb{C}[[t_1, \dots, t_N]]$. Define a map $W_N \rightarrow U(\widehat{\Omega}_N^\bullet)$ (the completion of the Lie algebra of Fourier coefficients of fields from $\widehat{\Omega}_N^\bullet$) sending

$$f(t_i) \partial_{t_j} \mapsto \text{Res} \left(: f(a_i^*(z)) a_j(z) : + \sum_{k=1}^N : (\partial_{t_k} f)(a_i^*(z)) \psi_k^*(z) \psi_j(z) : \right).$$

Explicit computation shows that it is a homomorphism of Lie algebras (this is an example of “supersymmetric cancellation of anomalies”) [FF2, MSV]. We obtain a W_N -action on $\widehat{\Omega}_N^\bullet$, which can be exponentiated to an action of the Harish-Chandra pair $(W_N, \text{Aut } \mathcal{O}_N)$. Furthermore, because the $(W_N, \text{Aut } \mathcal{O}_N)$ -action comes from the residues of vertex operators, it preserves the vertex algebra structure on V (so the group $\text{Aut } \mathcal{O}_N$ is an example of the group of internal symmetries from §4.4). Recall from §6.1 that any smooth scheme S of dimension N carries a canonical $(W_N, \text{Aut } \mathcal{O}_N)$ -structure \widehat{S} . Applying the construction of §6.1, we attach to $\widehat{\Omega}_N^\bullet$ a \mathcal{D} -module on S . The sheaf of horizontal sections of this \mathcal{D} -module is a sheaf of vertex superalgebras (with a structure of complex) on S . This is the *chiral deRham complex* of S , introduced in [MSV]. For its applications to mirror symmetry and elliptic cohomology, see [BoL, Bo1].

Now consider a purely bosonic analogue of this construction (i.e., replace $\widehat{\Omega}_N$ with \widehat{H}_N). The Lie algebra $U(\widehat{H}_N)$ of Fourier coefficients of the fields from \widehat{H}_N has a filtration $A_N^{\leq i}, i \geq 0$, by the order of $a_{i,n}$'s. Note that $A_N^{\leq 1}$ is a Lie subalgebra of $U(\widehat{H}_N)$ and A_N^0 is its (abelian) ideal. Denote by \mathcal{T}_N the Lie algebra $A_N^{\leq 1}/A_N^0$. Intuitively, $U(\widehat{H}_N)$ is the algebra of differential operators on the topological vector space $\mathbb{C}((t))^N$ with the standard filtration, $A^{(0)}$ (resp., \mathcal{T}_N) is the space of functions (resp., the Lie algebra of vector fields) on it, and \widehat{H}_N is the module of “delta-functions supported on $\mathbb{C}[[t]]^N$ ”.

In contrast to the finite-dimensional case, the extension

$$0 \rightarrow A_N^0 \rightarrow A_N^{\leq 1} \rightarrow \mathcal{T}_N \rightarrow 0$$

does not split [FF2]. Because of that, the naive map $W_N \rightarrow U(\widehat{H}_N)$ sending $f(t_i) \partial_{t_j}$ to $\text{Res} : f(a_i^*(z)) a_j(z) :$ is not a Lie algebra homomorphism. Instead, we obtain an extension

$$(23) \quad 0 \rightarrow \Omega_N^1/d\mathcal{O}_N \rightarrow \widetilde{W}_N \rightarrow W_N \rightarrow 0$$

of W_N by its module $\Omega_N^1/d\mathcal{O}_N$, which is embedded into $A_N^{\leq 1}$ by the formula $g(t_i) dt_j \mapsto \text{Res } g(a_i^*(z)) \partial_z a_j^*(z)$. If the sequence (23) were split, we would obtain a $(W_N, \text{Aut } \mathcal{O}_N)$ -action on \widehat{H}_N and associate to \widehat{H}_N a \mathcal{D} -module on S in the same way as above.

But in reality the extension (23) does not split for $N > 1$ (note that $\Omega_1^1/d\mathcal{O}_1 = 0$). Hence in general we only have an action on \widehat{H}_N of the Harish-Chandra pair $(\widetilde{W}_N, \widetilde{\text{Aut}} \mathcal{O}_N)$, where $\widetilde{\text{Aut}} \mathcal{O}_N$ is an extension of $\text{Aut } \mathcal{O}_N$ by the additive group $\Omega_N^1/d\mathcal{O}_N$. This action again preserves the vertex algebra structure on V . To apply the construction of §6.1, we need a

lifting of the $(W_N, \text{Aut } \mathcal{O}_N)$ -structure \widehat{S} on S to a $(\widetilde{W}_N, \widetilde{\text{Aut}} \mathcal{O}_N)$ -structure. Such liftings form a *gerbe* on S (in the analytic topology) with the lien $\Omega_S^1/d\mathcal{O}_S \simeq \Omega_{S,\text{cl}}^2$, the sheaf of closed holomorphic two-forms on S . The equivalence class of this gerbe in $H^2(S, \Omega_{S,\text{cl}}^2)$ is computed in [GMS]. If this class equals 0, then we can lift \widehat{S} to an $(\widetilde{W}_N, \widetilde{\text{Aut}} \mathcal{O}_N)$ -structure \widetilde{S} on S , and the set of isomorphism classes of such liftings is an $H^1(S, \Omega_{S,\text{cl}}^2)$ -torsor. Applying the construction of §6.1, we can then attach to \widehat{H}_N a \mathcal{D} -module on S . The sheaf of horizontal sections of this \mathcal{D} -module is a sheaf of vertex algebras on S . This is the sheaf of *chiral differential operators* on S (corresponding to \widetilde{S}) introduced in [MSV, GMS]. As shown in [GMS], the construction becomes more subtle in Zariski topology; in this case one naturally obtains a gerbe with the lien $\Omega_S^2 \rightarrow \Omega_{S,\text{cl}}^3$.

In the case when S is the flag manifold G/B of a simple Lie algebra \mathfrak{g} , one can construct the sheaf of chiral differential operators in a more direct way. The action of \mathfrak{g} on the flag manifold G/B gives rise to a homomorphism $\mathfrak{g}((t)) \rightarrow \mathcal{T}_N$, where $N = \dim G/B$. It is shown in [Wa, FF1, FF2] that it can be lifted to a homomorphism $\widehat{\mathfrak{g}} \rightarrow A_N^{\leq 1}$ of level $-h^\vee$ (Wakimoto realization). Moreover, it gives rise to a homomorphism of vertex algebras $V_{-h^\vee}(\mathfrak{g}) \rightarrow H_N$. Hence we obtain an embedding of the constant subalgebra \mathfrak{g} of $\widehat{\mathfrak{g}}$ into \widetilde{W}_N , and a (\mathfrak{g}, B) -action on \widehat{H}_N . On the other hand, G/B has a natural (\mathfrak{g}, B) -structure, namely G . Applying the localization construction to the Harish-Chandra pair (\mathfrak{g}, B) instead of $(\widetilde{W}_N, \widetilde{\text{Aut}} \mathcal{O}_N)$, we obtain the sheaf of chiral differential operators on G/B – note that it is uniquely defined in this case (see [MSV]).

7. CHIRAL ALGEBRAS

In §4.2, we gave a coordinate independent description of the vertex operation $Y(\cdot, z)$. The formalism of chiral algebras invented by Beilinson and Drinfeld [BD3] (see [G] for a review) is based on a coordinate-free realization of the operator product expansion (see §2.6). In the definition of \mathcal{Y}_x we acted by $Y(A, z)$ on B , placed at the fixed point $x \in X$. In order to define the OPE invariantly, we need to let x move along X as well. This is one as follows. Recall that to each conformal vertex algebra V we have assigned a vector bundle \mathcal{V} on any smooth curve X . Choose a formal coordinate z at $x \in X$, and use it to trivialize $\mathcal{V}|_{D_x}$. Then $\mathcal{V} \boxtimes \mathcal{V}(\infty\Delta)$ is a sheaf on $D_x^2 = \text{Spec } \mathbb{C}[[z, w]]$ associated to the $\mathbb{C}[[z, w]]$ -module $V \otimes V[[z, w]][(z-w)^{-1}]$. Let $\Delta_! \mathcal{V}$ be a sheaf on D_x^2 associated to the $\mathbb{C}[[z, w]]$ -module $V[[z, w]][(z-w)^{-1}]/V[[z, w]]$. Recall from §4.2 that we have a flat connection on \mathcal{V} . Hence \mathcal{V} is a vector bundle with a flat connection, i.e., a \mathcal{D} -module, on D_x . The sheaves $\mathcal{V} \boxtimes \mathcal{V}(\infty\Delta)$ and $\Delta_! \mathcal{V}$ are not vector bundles, but they have natural structures of \mathcal{D} -modules on D_x^2 (independent of the choice of z). Because of that, we switch to the language of \mathcal{D} -modules. The following result is proved in the same way as in Theorem 4.1 (see also [HL]).

THEOREM 7.1. — Define a map $\mathfrak{y}_x^{(2)} : \mathcal{V} \boxtimes \mathcal{V}(\infty\Delta) \rightarrow \Delta_! \mathcal{V}$ by the formula

$$\mathfrak{y}_x^{(2)}(f(z, w)A \boxtimes B) = f(z, w)Y(A, z - w) \cdot B \pmod{V[[z, w]]}.$$

Then $\mathfrak{y}_x^{(2)}$ is a homomorphism of \mathcal{D} -modules, which is independent of the choice of the coordinate z .

At this point it is convenient to pass to the right \mathcal{D} -module on X corresponding to the left \mathcal{D} -module \mathcal{V} . Recall that if \mathcal{F} is a left \mathcal{D} module on a smooth variety Z , then $\mathcal{F}^r := \mathcal{F} \otimes \Omega_Z$, where Ω_Z denotes the canonical sheaf on Z , is a right \mathcal{D} -module on Z . Denote $\Delta : X \hookrightarrow X^2, j : X^2 \setminus \Delta(X) \hookrightarrow X^2$.

COROLLARY 7.2. — The vertex algebra structure on V gives rise to a homomorphism of right \mathcal{D} -modules on X^2 , $\mu : j^* j_* \mathcal{V}^r \boxtimes \mathcal{V}^r \rightarrow \Delta_! \mathcal{V}^r$ satisfying the following conditions:

- (skew-symmetry) $\mu(f(x, y)A \boxtimes B) = -\sigma_{12} \circ \mu(f(y, x)B \boxtimes A)$;
- (Jacobi identity) $\mu(\mu(f(x, y, z)A \boxtimes B) \boxtimes C) + \sigma_{123} \circ \mu(\mu(f(y, z, x)B \boxtimes C) \boxtimes A) + \sigma_{123}^{-1} \circ \mu(\mu(f(z, x, y)C \boxtimes A) \boxtimes B) = 0$.
- (vacuum) we are given an embedding $\Omega \hookrightarrow \mathcal{V}^r$ compatible with the natural homomorphism $j_* j^* \Omega \boxtimes \Omega \rightarrow \Delta_! \Omega$.

Beilinson and Drinfeld define a *chiral algebra* on a curve X as a right \mathcal{D} -module \mathcal{A} on X equipped with a homomorphism $j_* j^* \mathcal{A} \boxtimes \mathcal{A} \rightarrow \Delta_! \mathcal{A}$ satisfying the conditions of Corollary 7.2 (see [BD3, G]).

The axioms of chiral algebra readily imply that for any chiral algebra \mathcal{A} , its deRham cohomology $H_{\text{dR}}^\bullet(\Sigma, \mathcal{A})$ is a (graded) Lie algebra for any open $\Sigma \subset X$. In particular, for an affine open $\Sigma \subset X$ we obtain the result mentioned in §5.3 that $H_{\text{dR}}^0(\Sigma, \mathcal{V}^r)$ is a Lie algebra.

Define the right \mathcal{D} -module \mathcal{V}_2^r on X^2 as the kernel of the homomorphism μ , and let $\mathcal{V}_2 = \mathcal{V}_2^r \otimes \Omega_{X^2}^{-1}$ be the corresponding left \mathcal{D} -module. There is a canonical isomorphism $H(X, x, V) \simeq H_{\text{dR}}^2(X^2, \mathcal{V}_2^r)$ (see [G], Prop. 5.1). It is dual to the isomorphism $C(X, x, V) \simeq \Gamma_{\nabla}(X^2, \mathcal{V}_2^*)$ mentioned in §5.4.

Beilinson and Drinfeld give a beautiful description of chiral algebras as *factorization algebras*. By definition, a factorization algebra is a collection of quasicoherent \mathcal{O} -modules \mathcal{F}_n on $X^n, n \geq 1$, satisfying a factorization condition, which essentially means that the fiber of \mathcal{F}_n at (x_1, \dots, x_n) is isomorphic to $\otimes_{s \in S} (\mathcal{F}_1)_s$, where $S = \{x_1, \dots, x_n\}$. For example, $j^* \mathcal{F}_2 = j^*(\mathcal{F}_1 \boxtimes \mathcal{F}_1)$ and $\Delta^* \mathcal{F}_2 = \mathcal{F}_1$. Intuitively, such a collection may be viewed as an \mathcal{O} -module on the *Ran space* $\mathcal{R}(X)$ of all finite non-empty subsets of X . In addition, it is required that \mathcal{F}_1 has a global section (unit) satisfying natural properties. It is proved in [BD3] that under these conditions each \mathcal{F}_n is automatically a left \mathcal{D} -module. Moreover, the right \mathcal{D}_X -module $\mathcal{F}_1^r = \mathcal{F}_1 \otimes \Omega_X$ acquires a canonical structure of chiral algebra, with μ given by the composition $j^* j_* \mathcal{F}_1^r \boxtimes \mathcal{F}_1^r = j^* j_* \mathcal{F}_2^r \rightarrow \Delta_! \Delta^! \mathcal{F}_2^r = \Delta_! \mathcal{F}_1^r$. In fact, there is

an equivalence between the categories of factorization algebras and chiral algebras on X [BD2].

For a chiral algebra \mathcal{V}^r , the corresponding sheaf \mathcal{F}_n on X^n is the sheaf \mathcal{V}_n mentioned in §5.4 (its dual is the sheaf of chiral correlation functions). Using these sheaves, Beilinson and Drinfeld define “chiral homology” $H_i^{\text{ch}}(X, \mathcal{V}^r)$, $i \geq 0$, of a chiral algebra \mathcal{V}^r . Intuitively, this is (up to a change of cohomological dimension) the deRham cohomology of the factorization algebra corresponding to \mathcal{V} , considered as a \mathcal{D} -module on the Ran space $\mathcal{R}(X)$. The 0th chiral homology of \mathcal{V}^r is isomorphic to $H_{\text{dR}}^n(X^n, \mathcal{V}_n \otimes \Omega_{X^n})$ for all $n > 1$ and hence to the space of coinvariants $H(X, x, V)$. The full chiral homology functor may therefore be viewed as a derived coinvariants functor.

As an example, we sketch the Beilinson-Drinfeld construction of the factorization algebra corresponding to the Kac-Moody vertex algebra $V_0(\mathfrak{g})$ [BD2] (see also [G]). Let Gr_n be the ind-scheme over X^n whose fiber Gr_{x_1, \dots, x_n} over $(x_1, \dots, x_n) \in X^n$ is the moduli space of pairs (\mathcal{P}, t) , where \mathcal{P} is a G -bundle on X , and t is its trivialization on $X \setminus \{x_1, \dots, x_n\}$ (e.g., the fiber of Gr_1 over any $x \in X$ is isomorphic to the affine Grassmannian $G(\mathcal{K})/G(\mathcal{O})$). Let e be the section of Gr_n corresponding to the trivial G -bundle, and $\mathcal{A}_{x_1, \dots, x_n}$ be the space of delta-functions on Gr_{x_1, \dots, x_n} supported at e . These are fibers of a left \mathcal{D} -module \mathcal{A}_n on X^n . Claim: $\{\mathcal{A}_n\}$ form a factorization algebra. The factorization property for $\{\mathcal{A}_n\}$ follows from the factorization property of the ind-schemes Gr_n : namely, $Gr_{x_1, \dots, x_n} = \prod_{s \in S} Gr_s$, where $S = \{x_1, \dots, x_n\}$. The chiral algebra corresponding to the factorization algebra $\{\mathcal{A}_n\}$ is $\mathcal{V}_0(\mathfrak{g})^r$. To obtain the factorization algebra corresponding to $\mathcal{V}_k(\mathfrak{g})^r$ with $k \neq 0$, one needs to make a twist by a line bundle on Gr_1 .

A spectacular application of the above formalism is the Beilinson-Drinfeld construction of the *chiral Hecke algebra* associated to a simple Lie algebra \mathfrak{g} and a negative integral level $k < -h^\vee$. The corresponding vertex algebra $S_k(\mathfrak{g})$ is a module over $\widehat{\mathfrak{g}} \times {}^L G$: $S_k(\mathfrak{g}) = \bigoplus_{\lambda \in {}^L P^+} M_{\lambda, k} \otimes V_\lambda^*$, where V_λ is a finite-dimensional ${}^L G$ -module with highest weight λ , and $M_{\lambda, k}$ is the irreducible highest weight $\widehat{\mathfrak{g}}$ -module of level k with highest weight $(k + h^\vee)\lambda$. Here ${}^L G$ is the group of adjoint type corresponding to ${}^L \mathfrak{g}$, and ${}^L P^+$ is the set of its dominant weights; we identify the dual \mathfrak{h}^* of the Cartan subalgebra of \mathfrak{g} with $\mathfrak{h} = {}^L \mathfrak{h}^*$ using the normalized invariant inner product on \mathfrak{g} . In particular, $S_k(\mathfrak{g})$ is a module over $V_k(\mathfrak{g}) = M_{0, k}$. Note that using results of [KL] one can identify the tensor category of ${}^L G$ -modules with a subcategory of the category \mathcal{O} of $\widehat{\mathfrak{g}}$ -modules of level k , so that V_λ corresponds to $M_{\lambda, k}$. The analogues of $S_k(\mathfrak{g})$ when \mathfrak{g} is replaced by an abelian Lie algebra with a non-degenerate inner product are lattice vertex algebras V_L from §3.4.

As a chiral algebra, the chiral Hecke algebra is the right \mathcal{D} -module $\mathcal{S}_k(\mathfrak{g})^r$ on X corresponding to the twist of $S_k(\mathfrak{g})$ by \widehat{X} . The corresponding factorization algebra is constructed analogously to the above construction of $V_k(\mathfrak{g})$ using certain irreducible \mathcal{D} -modules on Gr_1 corresponding to V_λ ; see [G] for the construction in the abelian case (it is also possible to construct a vertex algebra structure on $S_k(\mathfrak{g})$ directly, using results of

[KL]). Moreover, for any flat ${}^L G$ -bundle \mathcal{E} on X , the twist of $\mathcal{S}_k(\mathfrak{g})^r$ by \mathcal{E} with respect to the ${}^L G$ -action on $\mathcal{S}_k(\mathfrak{g})$ is also a chiral algebra on X . Conjecturally, the complex of sheaves on $\mathfrak{M}_G(X)$ obtained by localization of this chiral algebra (whose fibers are its chiral homologies) is closely related to the automorphic \mathcal{D} -module that should be attached to \mathcal{E} by the geometric Langlands correspondence.

REFERENCES

- [ADKP] E. ARBARELLO, C. DECONCINI, V. KAC and C. PROCESI – *Moduli spaces of curves and representation theory*, Comm. Math. Phys. **117** (1988) 1–36.
- [BL] A. BEAUVILLE and Y. LASZLO – *Conformal blocks and generalized theta functions*, Comm. Math. Phys. **164** (1994) 385–419.
- [BB] A. BEILINSON and J. BERNSTEIN – *A Proof of Jantzen Conjectures*, Advances in Soviet Mathematics **16**, Part 1, pp. 1–50, AMS 1993.
- [BD1] A. BEILINSON and V. DRINFELD – *Affine Kac–Moody algebras and polydifferentials*, Int. Math. Res. Notices **1** (1994) 1–11.
- [BD2] A. BEILINSON and V. DRINFELD – *Quantization of Hitchin’s Integrable System and Hecke eigensheaves*. Preprint.
- [BD3] A. BEILINSON and V. DRINFELD – *Chiral Algebras*. Preprint.
- [BFM] A. BEILINSON, B. FEIGIN and B. MAZUR – *Introduction to Algebraic Field Theory on Curves*. Preprint.
- [BG] A. BEILINSON and V. GINZBURG – *Infinitesimal structure of moduli spaces of G -bundles*, Duke Math. J. IMRN **4** (1992) 63–74.
- [BS] A. BEILINSON and V. SCHECHTMAN – *Determinant bundles and Virasoro algebras*, Comm. Math. Phys. **118** (1988) 651–701.
- [BPZ] A. BELAVIN, A. POLYAKOV and A. ZAMOLODCHIKOV – *Infinite conformal symmetries in two-dimensional quantum field theory*, Nucl. Phys. **B241** (1984) 333–380.
- [BF] D. BEN-ZVI and E. FRENKEL – *Vertex algebras and algebraic curves*, book in preparation.
- [B1] R. BORCHERDS – *Vertex algebras, Kac–Moody algebras and the monster*, Proc. Nat. Acad. Sci. USA **83** (1986) 3068–3071.
- [B2] R. BORCHERDS – *Monstrous moonshine and monstrous Lie superalgebras*, Invent. Math. **109** (1992) 405–444.
- [B3] R. BORCHERDS – *Quantum vertex algebras*, Preprint math.QA/9903038.
- [Bo1] L. BORISOV – *Introduction to the vertex algebra approach to mirror symmetry*, Preprint math.AG/9912195.

- [BoL] L. BORISOV, A. LIBGOBER – *Elliptic genera of toric varieties and applications to mirror symmetry*, Invent. Math. **140** (2000) 453–485.
- [dBT] J. de BOER and T. TJIN – *The relation between quantum \mathcal{W} -algebras and Lie algebras*, Comm. Math. Phys. **160** (1994) 317–332.
- [dFMS] P. di FRANCESCO, P. MATHIEU and D. SENECHAL – *Conformal Field Theory*. Springer–Verlag 1997.
- [D1] C. DONG – *Vertex algebras associated with even lattices*, J. Algebra **161** (1993) 245–265.
- [D2] C. DONG – *Representations of the moonshine module vertex operator algebra*, Contemp. Math. **175** (1994) 27–36.
- [DLM] C. DONG, H. LI and G. MASON – *Twisted representations of vertex operator algebras*, Math. Ann. **310** (1998) 571–600.
- [DS] V. DRINFELD and V. SOKOLOV – *Lie algebras and KdV type equations*, J. Sov. Math. **30** (1985) 1975–2036.
- [EK] P. ETINGOF and D. KAZHDAN – *Quantization of Lie bialgebras. V*, Preprint math.QA/9808121.
- [Fa] G. FALTINGS – *A proof of the Verlinde formula*, J. Alg. Geom. **3** (1994) 347–374.
- [FL] V. FATEEV and S. LUKYANOV – *The models of two-dimensional conformal quantum field theory with \mathbb{Z}_n symmetry*, Int. J. Mod. Phys. **A3** (1988), 507–520.
- [Fe1] B. FEIGIN – *The semi-infinite cohomology of Kac–Moody and Virasoro Lie algebras*, Russ. Math. Surv. **39**, No. 2 (1984) 155–156.
- [FF1] B. FEIGIN and E. FRENKEL – *A family of representations of affine Lie algebras*, Russ. Math. Surv. **43**, No. 5 (1988) 221–222.
- [FF2] B. FEIGIN and E. FRENKEL – *Affine Kac–Moody algebras and semi-infinite flag manifolds*, Comm. Math. Phys. **128** (1990) 161–189.
- [FF3] B. FEIGIN and E. FRENKEL – *Affine Kac–Moody algebras at the critical level and Gelfand–Dikii algebras*, Int. Jour. Mod. Phys. **A7**, Suppl. 1A (1992) 197–215.
- [FF4] B. FEIGIN and E. FRENKEL – *Integrals of Motion and Quantum Groups*, in Lect. Notes in Math. **1620**, pp. 349–418, Springer–Verlag 1996.
- [FS] B. FEIGIN and A. STOYANOVSKY – *Realization of a modular functor in the space of differentials, and geometric approximation of the moduli space of G -bundles*, Funct. Anal. Appl. **28** (1994) 257–275.
- [FKW] E. FRENKEL, V. KAC and M. WAKIMOTO – *Characters and fusion rules for \mathcal{W} -algebras via quantized Drinfeld–Sokolov reduction*, Comm. Math. Phys. **147** (1992), 295–328.
- [FKRW] E. FRENKEL, V. KAC, A. RADUL and W. WANG – $\mathcal{W}_{1+\infty}$ and \mathcal{W}_N with central charge N , Comm. Math. Phys. **170** (1995) 337–357.

- [FR] E. FRENKEL and N. RESHETIKHIN – *Towards deformed chiral algebras*, Preprint q-alg/9706023.
- [FK] I. FRENKEL and V. KAC – *Basic representations of affine Lie algebras and dual resonance models*, Invent. Math. **62** (1980) 23–66.
- [FGZ] I. FRENKEL, H. GARLAND and G. ZUCKERMAN – *Semi-infinite cohomology and string theory*, Proc. Nat. Acad. Sci. U.S.A. **83** (1986) 8442–8446.
- [FLM] I. FRENKEL, J. LEPOWSKY and A. MEURMAN – *Vertex Operator Algebras and the Monster*. Academic Press 1988.
- [FHL] I. FRENKEL, Y.-Z. HUANG and J. LEPOWSKY – *On axiomatic approaches to vertex operator algebras and modules*. Mem. Amer. Math. Soc. **104** (1993), no. 494.
- [FZ] I. FRENKEL and Y. ZHU – *Vertex operator algebras associated to representations of affine and Virasoro algebras*, Duke Math. J. **60** (1992) 123–168.
- [FrS] D. FRIEDAN and S. SHENKER – *The analytic geometry of two-dimensional conformal field theory*, Nucl. Phys. **B281** (1987) 509–545.
- [G] D. GAITSGORY – *Notes on 2D Conformal Field Theory and String Theory*, in Quantum fields and strings: a course for mathematicians, Vol. 2, pp. 1017–1089, AMS 1999.
- [Ga] K. GAWĘDZKI – *Conformal field theory*, Sémin. Bourbaki, Exp. 704, Asterisque **177-278** (1989) 95–126.
- [GKF] I.M. GELFAND, D.A. KAZHDAN and D.B. FUCHS – *The actions of infinite-dimensional Lie algebras*, Funct. Anal. Appl. **6** (1972) 9–13.
- [Gi] V. GINZBURG – *Resolution of diagonals and moduli spaces*, in The moduli space of curves, Progress in Math. **129**, pp. 231–266, Birkhäuser 1995.
- [Go] P. GODDARD – *Meromorphic conformal field theory*, in Infinite-dimensional Lie algebras and groups, V. Kac (ed.), pp. 556–587, World Scientific 1989.
- [GKO] P. GODDARD, A. KENT and D. OLIVE – *Unitary representations of the Virasoro and super-Virasoro algebras*, Comm. Math. Phys. **103** (1986) 105–119.
- [GMS] V. GORBOUNOV, F. MALIKOV and V. SCHECHTMAN – *Gerbes of chiral differential operators*. I, math.AG/9906116; II, math.AG/0003170.
- [Gu] R. GUNNING – *Lectures on Riemann Surfaces*. Princeton University Press 1966.
- [Hu] Y.-Z. HUANG – *Two-dimensional conformal geometry and vertex operator algebras*. Progress in Math. **148**. Birkhäuser 1997.
- [HL] Y.-Z. HUANG and J. LEPOWSKY – *On the \mathcal{D} -module and formal variable approaches to vertex algebras*, in Topics in geometry, pp. 175–202, Birkhäuser 1996.
- [K1] V. KAC – *Infinite-dimensional Lie algebras*, Third Edition. Cambridge University Press 1990.

- [K2] V. KAC – *Vertex Algebras for Beginners*, Second Edition. AMS 1998.
- [K3] V. KAC – *Formal distribution algebras and conformal algebras*, in Proc. XXIIth ICMP, Brisbane, 1994, pp. 80–96, International Press 1999.
- [KL] D. KAZHDAN and G. LUSZTIG – *Tensor structures arising from affine Lie algebras IV*, J. of AMS **7** (1993) 383–453.
- [Ko] M. KONTSEVICH – *The Virasoro algebra and Teichmüller spaces*, Funct. Anal. Appl. **21** (1987), no. 2, 156–157.
- [KNR] S. KUMAR, M.S. NARASIMHAN and A. RAMANATHAN – *Infinite Grassmannians and moduli spaces of G -bundles*, Math. Ann. **300** (1993) 395–423.
- [LW] J. LEPOWSKY and R.L. WILSON – *Construction of the affine Lie algebra $A_1^{(1)}$* , Comm. Math. Phys. **62** (1978) 43–53.
- [Li] H. LI – *Local systems of vertex operators, vertex superalgebras and modules*, J. Pure Appl. Alg. **109** (1996) 143–195.
- [LZ] B. LIAN and G. ZUCKERMAN – *New perspectives on the BRST-algebraic structure of string theory*, Comm. Math. Phys. **154** (1993) 613–646.
- [MSV] A. MALIKOV, V. SCHECHTMAN and A. VAINTROB – *Chiral deRham complex*, Comm. Math. Phys. **204** (1999) 439–473.
- [SV2] V. SCHECHTMAN and A. VARCHENKO – *Quantum groups and homology of local systems*, in Algebraic Geometry and Analytic Geometry, M. Kashiwara and T. Miwa (eds.), pp. 182–191, Springer-Verlag 1991.
- [Se] G. SEGAL – *The Definition of Conformal Field Theory*, unpublished manuscript.
- [So] C. SORGER – *La formule de Verlinde*, Sémin. Bourbaki, Exp. 793, Asterisque **237** (1996) 87–114.
- [TK] A. TSUCHIYA and Y. KANIE – *Vertex operators in conformal field theory on \mathbb{P}^1 and monodromy representations of the braid group*, in Adv. Stud. Pure Math **16**, pp. 297–372, Academic Press 1988.
- [TUY] A. TSUCHIYA, K. UENO and Y. YAMADA – *Conformal field theory on universal family of stable curves with gauge symmetries*, Adv. Stud. Pure Math. **19**, pp. 459–566, Academic Press 1989.
- [Wa] M. WAKIMOTO – *Fock representations of affine Lie algebra $A_1^{(1)}$* , Comm. Math. Phys. **104** (1986) 605–609.
- [Wa] W. WANG – *Rationality of Virasoro vertex operator algebras*, Duke Math. J. IMRN **7** (1993) 197–211.
- [Wi] E. WITTEN – *Quantum field theory, Grassmannians and algebraic curves*, Comm. Math. Phys **113** (1988) 529–600.
- [Z1] Y. ZHU – *Modular invariance of characters of vertex operator algebras*, J. AMS **9** (1996) 237–302.

- [Z2] Y. ZHU – *Global vertex operators on Riemann surfaces*, Comm. Math. Phys. **165** (1994) 485–531.

Edward FRENKEL

Department of Mathematics

University of California

Berkeley, CA 94720, USA

E-mail: frenkel@math.berkeley.edu

**DERIVATIVES OF EISENSTEIN SERIES
AND GENERATING FUNCTIONS FOR ARITHMETIC CYCLES**

by **Stephen S. KUDLA**

The classical formula of Siegel and Weil identifies the values of Siegel–Eisenstein series at certain critical points as integrals of theta functions. When the critical point is the center of symmetry for the functional equation, the Fourier coefficients of the values of the ‘even’ Siegel–Eisenstein series thus contain arithmetic information about the representations of quadratic forms. It is natural to ask for an arithmetic interpretation of the *derivative* of the ‘odd’ series at their center of symmetry.

I would like to report on my work on a family of identities relating the Fourier expansions of the derivatives of certain Siegel–Eisenstein series at their center of symmetry, on one side, and generating functions for the degrees of 0–cycles on moduli schemes for abelian varieties, on the other. On the one hand, such identities can be viewed as generalizations of the Siegel–Weil formula to the case of the derivative. On the other hand, the identities imply that the generating functions in question, which are given as power series in q with coefficients arising from arithmetical algebraic geometry, are in fact the q -expansions of modular forms. This work grows out of results obtained in collaboration with Steve Rallis [18], [19], [20] and with John Millson [15], [16], [17]. More recent progress has been made in collaboration with Michael Rapoport [21], [22], [23] and Tonghai Yang [24], [25]. At present, the identities have been fully established only in certain special cases as explained below. Nonetheless, these examples, together with partial results in higher dimensions, suggest the outline of a more extensive theory.

An additional origin of the investigation described here was the study of the triple product L-function at the center of the critical strip, in collaboration with Michael Harris [9] and with Benedict Gross [7]. In particular, a Siegel–Eisenstein series is a key ingredient in the Rankin–Selberg integral representation of this L-function. Thus the occurrence of arithmetic geometric quantities in the central derivatives of the Eisenstein series should reflect their appearance in the central derivative of the L-function, and hence should provide a relation to the Gross–Zagier formula [8].

Support of NSF Grant DMS 9970506 is gratefully acknowledged.

Section 1 contains two examples, one recalling the work of Hirzebruch and Zagier on the modular generating functions for curves on a Hilbert–Blumenthal surface and the second illustrating a generating function in the simplest arithmetic case, involving the derivative of a classical Eisenstein series of weight 1. In section 2, the incoherent Siegel–Eisenstein series, which should be related to arithmetic generating series, are defined in general. Section 3 reviews the results of [15],[16],[17] on generating functions for cycles on locally symmetric spaces. These results suggest what one should hope for in the arithmetic case. In section 4, the generating function for 0–cycles on an arithmetic surface attached to a Shimura curve is defined, and Conjecture 4.7 relates it to the central derivative of an incoherent Eisenstein series of weight $\frac{3}{2}$ and genus 2. The comparison of the nonsingular Fourier coefficients of the two objects is discussed in sections 5 and 6. Section 7 contains a brief survey of results in higher dimensional cases as well as a second look at the simplest example of section 1. Some speculations about further developments are made in section 8.

1. Two examples
2. Central derivatives of Siegel–Eisenstein series
3. Generating functions in the geometric case
4. Generating functions for arithmetic 0–cycles; the case of Shimura curves
5. Non-singular Fourier coefficients
6. Green’s functions and Whittaker functions
7. Further results
8. Final remarks

I would like to warmly thank J.-B. Bost, G. Henniart and M. Rapoport for detailed comments and advice on the original draft of this report.

0.1. Notation

$\mathbb{Q}, \mathbb{A}, \mathbb{A}_f, \mathbb{A}^\times$ denote the rational numbers, the adèles, the finite adèles, and the idèles of \mathbb{Q} respectively.

$\hat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}$, and, for any \mathbb{Z} -module M , $\hat{M} = M \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$.

\mathbb{F}_q denotes the finite field with q elements, and $\bar{\mathbb{F}}_q$ denotes an algebraic closure of it.

$(\ , \)_p$ (resp. $(\ , \)_{\mathbb{A}}$) denotes the quadratic Hilbert symbol for \mathbb{Q}_p (resp. \mathbb{A}).

ψ is a fixed nontrivial character of \mathbb{A}/\mathbb{Q} .

$\mathfrak{H}_g = \{\tau = u + iv \in \text{Sym}_g(\mathbb{C}) \mid v > 0\}$ is the Siegel space of genus g ;

$\mathfrak{H} = \mathfrak{H}_1$ is the upper half–plane.

$e(x) = e^{2\pi ix}$

$\text{Sym}_n(R) = \{x \in M_n(R) \mid {}^t x = x\} =$ the space of $n \times n$ symmetric matrices.

$$\text{diag}(a_1, \dots, a_n) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \in \text{Sym}_n(R).$$

1. TWO EXAMPLES

To fix ideas, it may be useful to consider two examples in classical language. The first of these illustrates the construction of generating functions for curves on a complex surface. More precisely, it gives a compact quotient version of the Hirzebruch–Zagier generating function for curves on a Hilbert–Blumenthal surface and a similar generating function for 0–cycles on such a surface. The second example illustrates the arithmetic case where cycles on moduli spaces for abelian varieties are defined by imposing extra endomorphisms. The example involves CM elliptic curves and the generating function is identified as the central derivative of an Eisenstein series of weight 1.

1.1. The case of a complex surface.

The results of this section are special cases of joint work with John Millson [15], [16], [17]. Let $V, (\cdot, \cdot)$ be a 4–dimensional rational vector space with a symmetric bilinear form of signature $(2, 2)$. Let $Q(x) = \frac{1}{2}(x, x)$ be the associated quadratic form. Fix a lattice L in V on which the form is \mathbb{Z} –valued and let

$$\Gamma \subset \{\gamma \in SO(V)(\mathbb{Q})^+ \mid \gamma L = L\}$$

be a subgroup of finite index, where $SO(V)(\mathbb{R})^+$ is the identity component of the Lie group $SO(V)(\mathbb{R})$ and $SO(V)(\mathbb{Q})^+ = SO(V)(\mathbb{Q}) \cap SO(V)(\mathbb{R})^+$. The space D of negative 2–planes in $V(\mathbb{R})$ is isomorphic to the product $\mathfrak{H} \times \mathfrak{H}$ of two copies of the upper half–plane \mathfrak{H} , and the quotient $S = \Gamma \backslash D$ is (the complex points of) a quasi–projective variety. Now assume that the space V is anisotropic so that S is projective. This assumption eliminates complications coming from the compactification of the cusps, which are a significant issue in the Hilbert–Blumenthal case considered by Hirzebruch and Zagier.

For a vector $x \in V(\mathbb{Q})$ with $Q(x) > 0$, let

$$D_x = \{z \in D \mid z \perp x\}$$

be the set of negative 2–planes orthogonal to x , so that $D_x \subset D$ is isomorphic to an embedded upper half–plane $\mathfrak{H} \subset \mathfrak{H} \times \mathfrak{H}$. The image $Z(x, \Gamma)$ of D_x in the quotient S is a compact curve, the image of the quotient $\Gamma_x \backslash D_x$, where Γ_x is the stabilizer of x . Note that the curve $Z(x, \Gamma)$ depends only on the Γ orbit of x . Associated to a positive integer t , there is a finite sum of such curves

$$(1.1) \quad Z(t, L) = \sum_{\substack{x \in L, Q(x) = t \\ \text{mod } \Gamma}} Z(x, \Gamma),$$

parametrized by the Γ –orbits in the set of lattice vectors of length t . This is the analogue of the Hirzebruch–Zagier curve T_N on a Hilbert–Blumenthal surface [11]. Let

$[Z(t, L)] \in H^2(S, \mathbb{Q})$ be the cohomology class of $Z(t, L)$, and let $[Z(0, L)] \in H^2(S, \mathbb{Q})$ be the cohomology class of the invariant form*

$$(1.2) \quad \omega = -\frac{1}{4\pi} (\operatorname{Im}(z_1))^{-2} dz_1 \wedge d\bar{z}_1 + \operatorname{Im}(z_2)^{-2} dz_2 \wedge d\bar{z}_2$$

on $\mathfrak{H} \times \mathfrak{H}$. As in [11], one can form a generating function

$$(1.3) \quad \phi_1(\tau, L) = [Z(0, L)] + \sum_{t \in \mathbb{Z}_{>0}} [Z(t, L)] q^t,$$

where $q = e(\tau)$, for an auxiliary variable $\tau = u + iv \in \mathfrak{H}$. The analogue of the result of Hirzebruch–Zagier [11] is a special case of [15], [16], [17], see Theorem 3.1 below.

THEOREM 1.1. — *The generating function $\phi_1(\tau, L)$ is an elliptic modular form of weight 2, valued in $H^2(S, \mathbb{C})$.*

Taking the intersection product with the cohomology class of an arbitrary curve C on S one obtains:

COROLLARY 1.2. — *For a curve C on S , the generating function for intersection numbers*

$$\phi_1(\tau, L) \cdot [C] = \operatorname{vol}(C) + \sum_{t > 0} [Z(t, L)] \cdot [C] q^t$$

is a modular form of weight 2. Here $\operatorname{vol}(C) = \int_C \omega$.

One can also define a generating function for 0-cycles on S as follows. For a pair of vectors $x = [x_1, x_2] \in V(\mathbb{Q})^2$ with matrix of inner products $Q(x) = \frac{1}{2}((x_i, x_j))_{i,j} \in \operatorname{Sym}_2(\mathbb{Q})$, there is an associated cycle $D_x = D_{x_1} \cap D_{x_2} \subset D$. If $Q(x)$ is nonsingular, then D_x is a point when $Q(x)$ is positive definite and is empty otherwise. If $Q(x)$ has rank 1, then $D_x = D_{x_1} = D_{x_2}$ is a curve when $Q(x) \geq 0$ (the components x_1 and x_2 are colinear since V is anisotropic) and is empty otherwise. Finally, if $Q(x) = 0$, then $x = 0$ and $D_x = D$. Let $Z(x, \Gamma)$ be the image of D_x in the quotient S ; again, this depends only on the Γ -orbit of x . For $T \in \operatorname{Sym}_2(\mathbb{Z})$, let

$$(1.4) \quad Z(T, L) = \sum_{\substack{x \in L^2, Q(x) = T \\ \text{mod } \Gamma}} Z(x, \Gamma).$$

If $T > 0$ is positive definite, $Z(T, L)$ is either empty or a finite sum of points on S . If $T \geq 0$, has rank 1, then $Z(T, L)$ is either empty or a finite sum of curves on S , and if $T = 0$, then $Z(0, L) = S$. Let $[Z(T, L)] \in H^{2r(T)}(S, \mathbb{Q})$ be the cohomology class of $Z(T, L)$, where $r(T) = \operatorname{rank}(T)$. The generating function in this case is

$$(1.5) \quad \phi_2(\tau, L) = \sum_{T \in \operatorname{Sym}_2(\mathbb{Z})_{\geq 0}} [Z(T, L)] \cup [\omega]^{2-r(T)} q^T,$$

*This is twice the form used in [11], p.104, since the $Z(t, L)$'s of (1.1) are twice the corresponding cycles in [11].

where $\tau \in \mathfrak{H}_2$, the Siegel space of genus 2, and $q^T = e(\text{tr}(T\tau))$. Note that the terms for singular T 's are obtained by shifting by suitable powers of $[\omega]$. The coefficients of this generating function lie in $H^4(S, \mathbb{C})$, and [15], [16], [17] yield the following result.

THEOREM 1.3. — *The generating function $\phi_2(\tau, L)$ is a Siegel modular form of weight 2 and genus 2 valued in $H^4(S, \mathbb{C})$.*

Applying the degree map $H^4(S, \mathbb{C}) \rightarrow \mathbb{C}$, one obtains a scalar valued Siegel modular form.

COROLLARY 1.4. — *The generating function*

$$\deg(\phi_2(\tau, L)) = \text{vol}(S) + \sum_{\substack{T \in \text{Sym}_2(\mathbb{Z})_{\geq 0} \\ r(T)=1}} \text{vol}(Z(T, L)) q^T + \sum_{T \in \text{Sym}_2(\mathbb{Z})_{>0}} \deg(Z(T, L)) q^T$$

is a Siegel modular form of weight 2 and genus 2.

In particular, the positive definite Fourier coefficients of $\deg(\phi_2(\tau, L))$ are the degrees of the 0-cycles $Z(T, L)$ on the surface S . The volumes of curves on S are taken with respect to the restriction of the invariant (1,1)-form ω of (1.2) and

$$\text{vol}(S) = \int_S \omega^2.$$

Theorems 1.1 and 1.3 are proved by constructing a theta function $\theta_1(\tau, L)$ for $\tau \in \mathfrak{H}$, resp. $\theta_2(\tau, L)$ for $\tau \in \mathfrak{H}_2$, valued in *closed* (1, 1)-forms, resp. closed (2, 2)-forms on S . The generating function is the cohomology class of this theta function, i.e., $\phi_i(\tau, L) = [\theta_i(\tau, L)]$ for $i = 1, 2$, and hence is modular. In addition, the generating function of Corollary 1.2 resp. Corollary 1.4 is obtained as the integral of $\theta_1(\tau, L)$ over the curve C , resp. $\theta_2(\tau, L)$ over S . For suitable Γ 's, this last integral over S is a constant multiple of the group theoretic integral of the theta function which occurs in the Siegel–Weil formula, and hence coincides with a special value of a Siegel–Eisenstein series of genus 2 at the point $s = \frac{1}{2}$, [13] and Proposition 3.2 below.

COROLLARY 1.5. — *There is a nonzero constant c such that*

$$\deg(\phi_2(\tau, L)) = c \cdot E(\tau, \frac{1}{2}, L)$$

for a suitable Siegel–Eisenstein series $E(\tau, s, L)$ of genus 2 and weight 2.

In the case in which S is a product of modular curves, such a geometric interpretation of the Fourier coefficients of a Siegel–Eisenstein series was observed by Gross and Keating [6].

1.2. Interlude.

More general results of this type, [15], [16], [17], and [13], concerning generating functions for cycles of codimension n on Shimura varieties X defined by rational quadratic forms of signature $(m - 2, 2)$ are discussed in section 3 below. Note that the complex dimension of X is $m - 2$. The main aim of this report is to explain the first steps in establishing a similar theory in the arithmetic case. Roughly speaking, this means the following. First, one wants to consider cycles of codimension n on integral models \mathfrak{X} of the Shimura varieties X and their classes in the arithmetic Chow groups $\widehat{CH}^r(\mathfrak{X})$, [4]. For $2 \leq m \leq 4$, integral models can be obtained as moduli spaces of suitable abelian varieties, and cycles can be defined as the loci where the abelian varieties in question are equipped with additional endomorphisms of a certain type. The theta functions valued in the de-Rham complex are not available in this context, and so it is not clear at present how to define generating functions for cycles of arbitrary codimension. If one considers 0-cycles, however, one may apply the arithmetic degree map $\widehat{\deg} : \widehat{CH}^{m-1}(\mathfrak{X}) \rightarrow \mathbb{R}$. One may then look for an analogue of Corollary 1.5 and Proposition 3.2, where the Siegel–Eisenstein series will now have genus $n = m - 1$ and the critical point will be $s_0 = \frac{m}{2} - \frac{n+1}{2} = 0$, i.e., the central point on the real axis for the functional equation of the Eisenstein series. Moreover, it turns out that the ‘correct’ Eisenstein series of genus n and weight $\frac{n+1}{2}$ will have a zero at this point, so that one should look at its first derivative $E'(\tau, 0, L)$. The case of the arithmetic surfaces attached to Shimura curves is discussed at length below in sections 4–6, and the analogue of Corollary 1.5 is given in Conjecture 4.7. The simplest example, however, occurs for $m = 2$. This case involves only classical objects, e.g., elliptic curves with complex multiplication and Eisenstein series of weight 1 for SL_2 .

1.3. Derivatives of Eisenstein series of weight 1.

This section describes the simplest case in which the derivative at $s = 0$ of an Eisenstein series can be identified with a generating function for the arithmetic degrees of 0-cycles on a moduli scheme. These results are joint work with Michael Rapoport and Tonghai Yang [24]. Fix a prime $d \equiv 3 \pmod{4}$ with $d > 3$, and let $\mathbf{k} = \mathbb{Q}(\sqrt{-d})$ be the corresponding imaginary quadratic field with ring of integers $\mathcal{O}_{\mathbf{k}}$ and associated Dirichlet character χ_d . Let

$$\Lambda(s, \chi_d) = \pi^{-\frac{s+1}{2}} \Gamma\left(\frac{s+1}{2}\right) L(s, \chi_d)$$

where $L(s, \chi_d)$ is the Dirichlet L -series of χ_d . For a nonzero integer $n \in \mathbb{Z}$, let $\rho(n)$ be the number of ideals in $\mathcal{O}_{\mathbf{k}}$ of norm n . For example, note that, for a prime p , $\rho(p) = 0$ if p is inert in \mathbf{k} , $\rho(p) = 2$ if p is split in \mathbf{k} , and $\rho(d) = 1$.

There are two normalized Eisenstein series of weight 1 for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ attached to \mathfrak{k} . For $\tau = u + iv \in \mathfrak{H}$, and $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, let

$$E_{\pm}^*(\tau, s) = v^{\frac{s}{2}} d^{\frac{s+1}{2}} \Lambda(s+1, \chi_d) \sum_{\gamma \in \Gamma_{\infty} \setminus \Gamma} (c\tau + d)^{-1} |c\tau + d|^{-s} \Phi_{\pm}(\gamma),$$

where, for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

$$\Phi_{\pm}(\gamma) = \begin{cases} \chi_d(a) & \text{if } c \equiv 0 \pmod{d} \\ \pm i d^{-\frac{1}{2}} \chi_d(c) & \text{if } c \text{ is prime to } d. \end{cases}$$

The entire analytic continuation of these series in s satisfy the functional equations

$$E_{\pm}^*(\tau, -s) = \pm E_{\pm}^*(\tau, s).$$

A general construction of series of this sort is described in section 2 below. A case of the Siegel–Weil formula due to Hecke describes the value at $s = 0$ of the even series:

$$E_{+}^*(\tau, 0) = 2h_{\mathfrak{k}} + 4 \sum_{t=1}^{\infty} \rho(t) q^t = 2 \sum_{\mathfrak{a}} \vartheta(\tau, \mathfrak{a}),$$

where $h_{\mathfrak{k}}$ is the class number of \mathfrak{k} , the ideal \mathfrak{a} runs over representatives of the ideal classes of \mathfrak{k} , and $\vartheta(\tau, \mathfrak{a})$ is the binary theta series attached to \mathfrak{a} . For the odd series, $E_{-}^*(\tau, 0) = 0$, and the function of interest is the (negative of the) leading term

$$\phi(\tau, d) = -\frac{\partial}{\partial s} \{E_{-}^*(\tau, s)\} \Big|_{s=0}.$$

THEOREM 1.6. — *The modular form $\phi(\tau, d)$ of weight 1 has Fourier expansion*

$$\phi(\tau, d) = a_0(v) + \sum_{t < 0} a_t(v) q^t + \sum_{t=1}^{\infty} a_t q^t,$$

where, for $t > 0$,

$$a_t = 2 \log(d) (\mathrm{ord}_d(t) + 1) \rho(t) + 2 \sum_{p \neq d} \log(p) (\mathrm{ord}_p(t) + 1) \rho(t/p),$$

where the sum runs over primes p inert in \mathfrak{k} ,

$$a_0 = -h_{\mathfrak{k}} \left(\log(d) + \log(v) + 2 \frac{\Lambda'(1, \chi_d)}{\Lambda(1, \chi_d)} \right),$$

and, for $t < 0$,

$$a_t(v) = -2 \mathrm{Ei}(-4\pi|t|v) \rho(-t).$$

Here

$$-\mathrm{Ei}(-x) = \int_1^{\infty} u^{-1} e^{-ux} du$$

is the exponential integral.

The idea now is to give an interpretation of these coefficients as the degrees in the sense of arithmetic geometry of certain 0-cycles on the coarse moduli scheme \mathcal{M} for elliptic curves (E, ι) with complex multiplication $\iota : \mathcal{O}_{\mathfrak{k}} \hookrightarrow \text{End}(E)$ by $\mathcal{O}_{\mathfrak{k}}$. This scheme over $\mathcal{O}_{\mathfrak{k}}$ can be identified with $\text{Spec}(\mathcal{O}_H)$, where H is the Hilbert class field of \mathfrak{k} . For such a curve (E, ι) , the space of special endomorphisms is the \mathbb{Z} -module

$$(1.6) \quad V(E, \iota) = \{x \in \text{End}(E) \mid x \iota(a) = \iota(\bar{a}) x \text{ for all } a \in \mathcal{O}_{\mathfrak{k}} \}.$$

This space has a \mathbb{Z} -valued quadratic form Q defined by $x^2 = -Q(x) \cdot \text{id}_E$. For $t \in \mathbb{Z}$, let $Z(t)$ be the coarse moduli scheme whose points over an algebraically closed field correspond to triples (E, ι, x) where $x \in V(E, \iota)$ with $Q(x) = t$. The scheme $Z(t) \rightarrow \mathcal{M}$ is the locus of (E, ι) 's with an extra multiplication, anticommuting with the action of $\mathcal{O}_{\mathfrak{k}}$. Such extra endomorphisms can only exist for a supersingular curve E in characteristic p for a prime p which is not split in \mathfrak{k} . Then $Z(t) = \text{Spec}(R(t))$ where $R(t)$ is an Artin algebra in which p is nilpotent. Let

$$\widehat{\text{deg}}(Z(t)) = \log |R(t)|.$$

The second main result of [24] is a calculation of this degree; this calculation depends in an essential way on the results of Gross [5].

THEOREM 1.7. — *For $t > 0$,*

$$\widehat{\text{deg}}(Z(t)) = a_t,$$

and hence

$$\phi(\tau, d) = \sum_{t>0} \widehat{\text{deg}}(Z(t)) q^t + a_0(v) + \sum_{t<0} a_t(v) q^t.$$

A sort of geometric interpretation of the remaining terms will be discussed in section 7 below, cf. also [24].

2. CENTRAL DERIVATIVES OF SIEGEL–EISENSTEIN SERIES

A general construction of ‘even’ and ‘odd’ Siegel–Eisenstein series is best described in representation theoretic language, and is connected with the Siegel–Weil formula at the central critical point. These series, which will be called coherent and incoherent series respectively, for reasons explained below, will have integral or half-integral weight depending on the parity of the dimension of the relevant quadratic spaces. Hence it is necessary to work on the metaplectic group.

Let $G = \text{Sp}_{2n}$ be the symplectic group of rank n over \mathbb{Q} and let $P = MN$ be the maximal parabolic with Levi factor $M \simeq \text{GL}_n$ and unipotent radical $N \simeq \text{Sym}_n$. Let

$$G_{\mathbb{A}} = \begin{cases} \text{Mp}_{2n}(\mathbb{A}) & \text{if } n \text{ is even} \\ \text{Sp}_{2n}(\mathbb{A}) & \text{if } n \text{ is odd,} \end{cases}$$

where $Mp_{2n}(\mathbb{A})$ is the twofold metaplectic cover of $Sp_{2n}(\mathbb{A})$, and let $P_{\mathbb{A}}$ and $M_{\mathbb{A}}$ be the subgroups of $G_{\mathbb{A}}$ corresponding to P and M . Let $G_{\mathbb{Q}}$ be $Sp_{2n}(\mathbb{Q})$ for n odd resp. the image of this group in $Mp_{2n}(\mathbb{A})$ under the canonical splitting, if n is even. For each place $p \leq \infty$ of \mathbb{Q} , there are groups G_p , P_p and M_p , defined analogously.

A quadratic character χ of $\mathbb{A}^{\times}/\mathbb{Q}^{\times}$ determines a character $\chi = \chi^{\psi}$ of $M_{\mathbb{A}}$, trivial on $M_{\mathbb{Q}} = M_{\mathbb{A}} \cap G_{\mathbb{Q}}$, and for $s \in \mathbb{C}$, one has the degenerate principal series representation

$$I(s, \chi) = \text{Ind}_{P_{\mathbb{A}}}^{G_{\mathbb{A}}}(\chi | \cdot|^s)$$

of $G_{\mathbb{A}}$. (The character χ^{ψ} depends on the fixed choice of the nontrivial additive character ψ of \mathbb{A}/\mathbb{Q} in the metaplectic case.)

For $\Phi(s) \in I(s, \chi)$, the Siegel-Eisenstein series is defined for $\text{Re}(s) > \frac{n+1}{2}$ by

$$E(g, s, \Phi) = \sum_{\gamma \in P_{\mathbb{Q}} \backslash G_{\mathbb{Q}}} \Phi(\gamma g, s).$$

From the standard theory of Eisenstein series one knows that this function has a meromorphic analytic continuation to the whole s -plane and satisfies a functional equation relating s and $-s$. In addition, it has no poles on the line $\text{Re}(s) = 0$ (unitary axis). In particular, there is an intertwining map

$$(2.1) \quad E(0) : I(0, \chi) \longrightarrow \mathcal{A}(G), \quad \Phi(0) \mapsto E(0, \Phi)$$

from the degenerate principal series at $s = 0$ to the space $\mathcal{A}(G)$ of automorphic forms on $G_{\mathbb{A}}$.

The image and kernel of this map can be described in terms of representations associated to quadratic forms as follows.

A rational vector space V of dimension m with a nondegenerate quadratic form Q determines a quadratic character χ_V of $\mathbb{A}^{\times}/\mathbb{Q}^{\times}$ by

$$\chi_V(x) = (x, (-1)^{\frac{m(m-1)}{2}} \det(Q))_{\mathbb{A}},$$

where $(\ , \)_{\mathbb{A}}$ is the global quadratic Hilbert symbol. For such a space (V, Q) , there is a Weil representation ω_V of $G_{\mathbb{A}}$ on the Schwartz space $S(V(\mathbb{A})^n)$, determined by ψ . This gives rise to a $G_{\mathbb{A}}$ -intertwining map

$$\lambda_V : S(V(\mathbb{A})^n) \longrightarrow I(s_0, \chi_V), \quad \lambda_V(\varphi)(g) = (\omega_V(g)\varphi)(0),$$

where $s_0 = \frac{m}{2} - \frac{n+1}{2}$. Specializing to the case $\chi_V = \chi$ and $m = n + 1$, one obtains an irreducible constituent $\Pi(V) = \lambda_V(S(V(\mathbb{A})^n))$ of $I(0, \chi)$.

Similarly, for each place $p \leq \infty$, there is an analogous local construction which yields an irreducible constituent $\Pi_p(V_p)$ of the local induced representation $I_p(0, \chi_{V_p})$ of G_p associated to each quadratic space V_p over \mathbb{Q}_p of dimension $n + 1$ and character χ_{V_p} . Then, for a global space V , one has

$$\Pi(V) \simeq \otimes'_{p \leq \infty} \Pi_p(V_p) \subset I(0, \chi) = \otimes'_{p \leq \infty} I_p(0, \chi_p)$$

where $V_p = V \otimes_{\mathbb{Q}} \mathbb{Q}_p$ and the primes indicate the restricted tensor products. For a finite prime p , there are precisely two possible quadratic spaces V_p^+ and V_p^- over \mathbb{Q}_p , for a fixed n and character χ_p . They are distinguished by their Hasse invariants $\epsilon_p(V_p^\pm) = \pm 1$, and, in fact, [19],

$$I_p(0, \chi_p) = \Pi_p(V_p^+) \oplus \Pi_p(V_p^-).$$

For $p = \infty$, the quadratic spaces of dimension $n + 1$ and character χ_∞ are determined by their signature, and fall into two groups according to their Hasse invariant. The local induced representation $I_\infty(0, \chi_\infty)$ is the direct sum of the corresponding $\Pi_\infty(V_\infty)$'s, [18]. For example, for $n = 4$ the quadratic spaces over \mathbb{R} of signatures $(5, 0)$, and $(1, 4)$ have Hasse invariant $+1$, while that of signature $(3, 2)$ has Hasse invariant -1 , and

$$I_\infty(0, \chi_\infty) = \Pi(5, 0) \oplus \Pi(3, 2) \oplus \Pi(1, 4),$$

in the obvious notation. Here $\chi_\infty = 1$.

If a collection of local quadratic spaces $\mathcal{C} = \{V_p\}$ is the set of localizations of a global space V , then the product formula for the Hasse invariants asserts that

$$\epsilon(\mathcal{C}) := \prod_{p \leq \infty} \epsilon_p(V_p) = 1.$$

Such a collection and the Eisenstein series associated to it will be called *coherent*. On the other hand, the collection \mathcal{C} of local quadratic spaces obtained by choosing one prime p_0 (e.g., $p_0 = \infty$) and switching the space V_{p_0} to a space V'_{p_0} with the opposite Hasse invariant has

$$\epsilon(\mathcal{C}) := \epsilon_{p_0}(V'_{p_0}) \prod_{\substack{p \leq \infty \\ p \neq p_0}} \epsilon_p(V_p) = -1,$$

so that such a collection cannot be the set of localizations of any global quadratic space. In this case, the collection \mathcal{C} and the Eisenstein series associated to it will be called *incoherent*. The irreducible admissible representation

$$\Pi(\mathcal{C}) = \Pi_{p_0}(V'_{p_0}) \otimes \left(\bigotimes_{\substack{p \leq \infty \\ p \neq p_0}} \Pi_p(V_p) \right)$$

of $G_{\mathbb{A}}$ is also a constituent of $I(0, \chi)$. Then there is a direct sum decomposition

$$I(0, \chi) = \left(\bigoplus_V \Pi(V) \right) \oplus \left(\bigoplus_{\mathcal{C}} \Pi(\mathcal{C}) \right)$$

where V runs over all global quadratic spaces of dimension $n + 1$ and character χ , and \mathcal{C} runs over all incoherent collections, where $\epsilon(\mathcal{C}) = -1$, as just described. One then obtains a description of the kernel and image of the map $E(0)$ in terms of the $\Pi(V)$'s and $\Pi(\mathcal{C})$'s, [14], [20], [9].

THEOREM 2.1. — (i)

$$\ker(E(0)) = \bigoplus_{\epsilon(\mathcal{C}) = -1} \Pi(\mathcal{C}).$$

(ii) Each automorphic representation $\Pi(V)$ in the image

$$\mathrm{Im}(E(0)) \simeq \bigoplus_V \Pi(V)$$

coincides with space of (regularized) theta integrals

$$(2.2) \quad I(g, \varphi) = \int_{O(V)(\mathbb{Q}) \backslash O(V)(\mathbb{A})} \theta(g, h; \varphi) dh,$$

attached to the global quadratic space V . Here, for $\varphi \in S(V(\mathbb{A})^n)$, $g \in G_{\mathbb{A}}$ and $h \in O(V)(\mathbb{A})$,

$$\theta(g, h; \varphi) = \sum_{x \in V(\mathbb{Q})} (\omega(g)\varphi)(h^{-1}x).$$

The integral $I(g, \varphi)$ must be defined by a regularization procedure [20] whenever V is isotropic.

Part (ii) of Theorem 1.2 is essentially the Siegel–Weil formula in the present context, [20], [33]. Note that the theta functions involve global arithmetic, e.g., the number of solutions of diophantine equations of the form $Q(x) = T$ for $T \in \mathrm{Sym}_n(\mathbb{Z})$ and $x \in L^n$ for a lattice $L \subset V(\mathbb{Q})$, whereas the Eisenstein series is constructed from local data.

PROBLEM. — What is the arithmetic content of the first derivative $E'(g, 0, \Phi)$ when $\Phi \in \Pi(\mathcal{C})$ with $\epsilon(\mathcal{C}) = -1$?

REMARK 2.2. — The results to be discussed in the remainder of the talk suggest an answer to this question, at least for the following particular case:

Let V be a rational quadratic space, as above, with signature $(n-1, 2)$, and let \mathcal{C} be the collection of local quadratic spaces obtained from $\{V_p\}$ by replacing V_∞ by the space V'_∞ of signature $(n+1, 0)$. Let $\varphi'_\infty \in S((V'_\infty)^n)$ be the Gaussian $\varphi'_\infty(x) = \exp(-\pi Q'(x))$, and let $\Phi_\infty^{(n+1)/2}(s) \in I_\infty(s, \chi_\infty)$ be the corresponding section; it is the unique eigenvector for K_∞ of weight $\frac{n+1}{2}$. For any $\varphi \in S(V(\mathbb{A}_f)^n)$, with corresponding section $\Phi_f(s) \in I_f(s, \chi_f)$,

$$\Phi(0) = \Phi_\infty^{\frac{n+1}{2}}(0) \otimes \Phi_f(0) \in \Pi(\mathcal{C}).$$

Then the central derivative $E'(g, 0, \Phi)$ should be related to a generating function for the degrees of 0-cycles on an integral model of the Shimura variety associated to the group $\mathrm{GSpin}(V)$.

REMARK 2.3. — For comparison with the generating functions considered below, it is convenient to write the Eisenstein series in a more classical language. For $\tau = u + iv \in \mathfrak{H}_n$, the Siegel space of genus n , and for $\varphi \in S(V(\mathbb{A}_f)^n)$, let

$$(2.3) \quad \begin{aligned} E(\tau, s, \varphi) &:= \det(v)^{-\frac{1}{2}(s + \frac{n+1}{2})} E(g_\tau, s, \Phi) \\ &= \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \det(c\tau + d)^{-\frac{n+1}{2}} |\det(c\tau + d)|^{-s} (\omega_f(\gamma)\varphi)(0), \end{aligned}$$

where $\Gamma = \mathrm{Sp}_{2n}(\mathbb{Z})$, $\Gamma_\infty = \Gamma \cap P(\mathbb{Q})$, ω_f denotes the action of $G_{\mathbb{A}_f}$ on $S(V(\mathbb{A}_f)^n)$ via the Weil representation, and $g_\tau = \begin{pmatrix} v^{\frac{1}{2}} & uv^{-\frac{1}{2}} \\ & v^{-\frac{1}{2}} \end{pmatrix}$.

3. GENERATING FUNCTIONS IN THE GEOMETRIC CASE

This section describes the results of [15], [16], [17], and [13] on generating functions for the cohomology classes of special cycles in the case of $O(n-1, 2)$, special cases of which were described in section 1.1. These results suggest what one might hope to prove in the arithmetic case.

For a rational quadratic space V of signature $(n-1, 2)$, let $H = \mathrm{GSpin}(V)$ and let D be the space of oriented negative 2-planes in $V(\mathbb{R})$. The space D is isomorphic to two copies of a bounded domain of type IV in \mathbb{C}^{n-1} , [10], [28], and the group $H(\mathbb{R})$ acts on it by holomorphic automorphisms. For a compact open subgroup $K \subset H(\mathbb{A}_f)$, the orbit space

$$X_K(\mathbb{C}) = H(\mathbb{Q}) \backslash D \times H(\mathbb{A}_f) / K$$

is the set of complex points of a quasi-projective variety X_K defined over \mathbb{Q} , the Shimura variety attached to H , D and K . The variety X_K is in fact projective if V is anisotropic and smooth if K is sufficiently small.

Fix $r \in \mathbb{Z}$ with $1 \leq r \leq n-1$. For $x \in V(\mathbb{R})^r$, let

$$D_x = \{z \in D \mid x \perp z\}$$

be the set of z 's which are orthogonal to all components of x . If the matrix $T = Q(x) = \frac{1}{2}((x_i, x_j))$ is positive definite, i.e., if the components of x span a positive r -plane, then D_x has complex codimension r in D . If, in addition, $x \in V(\mathbb{Q})^r$, then x^\perp is a rational quadratic space of signature $(n-r-1, 2)$, the stabilizer H_x of x in H is isomorphic to $\mathrm{GSpin}(x^\perp)$, and there is a natural map of Shimura varieties

$$Z(x, K) : H_x(\mathbb{Q}) \backslash D_x \times H_x(\mathbb{A}_f) / H_x(\mathbb{A}_f) \cap K \longrightarrow X_K(\mathbb{C}),$$

giving a cycle of codimension r on $X_K(\mathbb{C})$. Given a function $\varphi \in S(V(\mathbb{A}_f)^r)^K$ and $T \in \mathrm{Sym}_r(\mathbb{Q})_{>0}$, there is a weighted linear combination $Z(T, \varphi)$ of such cycles [13], and the resulting cohomology classes $[Z(T, \varphi)] \in H^{2r}(X_K)$, where $H^*(X_K)$ is the cohomology of $X_K(\mathbb{C})$ with complex coefficients [17]. Examples are given by (1.1) and (1.4) in section 1.1, where φ is the characteristic function of $(\hat{L})^r \subset V(\mathbb{A}_f)^r$, for $\hat{L} = L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$. If T is only positive semi-definite with $\mathrm{rank}(T) = r(T)$, the associated cycles have codimension $r(T)$ and their cohomology classes lie in $H^{2r(T)}(X_K)$.

To form a generating function, let $\tau = u + iv \in \mathfrak{H}_r$, the Siegel space of genus r , and let $q^T = e(\mathrm{tr}(\tau T))$.

THEOREM 3.1. — [17]. For $\varphi \in S(V(\mathbb{A}_f)^r)^K$, and for a suitable choice of a Kähler form ω on $X_K(\mathbb{C})$, the generating series

$$\phi_r(\tau, \varphi) = \sum_{T \in \text{Sym}_r(\mathbb{Q})_{\geq 0}} [Z(T, \varphi)] \cdot [\omega]^{r-r(T)} q^T$$

is the q -expansion of a holomorphic Siegel modular form of weight $\frac{n+1}{2}$ and genus r valued in $H^{2r}(X_K)$.

The proof of this result depends on a construction of a theta function taking values in the space of closed $2r$ forms on $X_K(\mathbb{C})$. This method is quite general and applies to the locally symmetric spaces associated to $O(p, q)$, $U(p, q)$ and $Sp(p, q)$, cf. [15], [16], and [17].

Specializing to the case where V is anisotropic and $r = n - 1$ and applying $\text{deg} : H^{2(n-1)}(X_K) \rightarrow \mathbb{C}$, one obtains a holomorphic Siegel modular form of genus $n - 1$ and weight $\frac{n+1}{2}$,

$$(3.1) \quad \phi_{\text{deg}}(\tau, \varphi) = \text{deg}(\phi_{n-1}(\tau, \varphi)) = \sum_{T \in \text{Sym}_{n-1}(\mathbb{Q})_{\geq 0}} \text{deg}([Z(T, \varphi)] \cdot [\omega]^{n-1-r(T)}) q^T.$$

By the Siegel–Weil formula [20], this form is, in turn, a value of an Eisenstein series. More precisely, let

$$G'_{\mathbb{A}} = \begin{cases} \text{Mp}_{2(n-1)}(\mathbb{A}) & \text{if } n \text{ is even} \\ \text{Sp}_{2(n-1)}(\mathbb{A}) & \text{if } n \text{ is odd.} \end{cases}$$

The machinery described in section 2 carries over, except that the map $\lambda_V : S(V(\mathbb{A})) \rightarrow I(s_0, \chi)$ now takes values in the induced representation at the point $s_0 = \frac{1}{2}$. Let $E(g, s, \Phi)$ be the Siegel–Eisenstein series associated to the section $\Phi_{\infty}^{\frac{n+1}{2}}(s) \otimes \Phi_f(s)$, where $\Phi_f(\frac{1}{2}) = \lambda_V(\varphi)$ for the weight function $\varphi \in S(V(\mathbb{A}_f)^{n-1})^K$. The Siegel–Weil formula for the anisotropic space V then implies the following generalization of Corollary 1.5 above [13]:

PROPOSITION 3.2. — For $\tau = u+iv \in \mathfrak{H}_{n-1}$, and for a weight function $\varphi \in S(V(\mathbb{A}_f)^{n-1})^K$,

$$\phi_{\text{deg}}(\tau, \varphi) = \text{vol}(X_K(\mathbb{C})) \det(v)^{-\frac{n+1}{4}} E(g_{\tau}, \frac{1}{2}, \Phi).$$

where

$$\text{vol}(X_K(\mathbb{C})) = \int_{X_K(\mathbb{C})} \omega^{n-1}.$$

In particular, the positive definite Fourier coefficients of this Eisenstein series are the degrees of the (weighted) 0-cycles $Z(T, \varphi)$ on X_K .

4. GENERATING FUNCTIONS FOR ARITHMETIC 0-CYCLES: THE CASE OF SHIMURA CURVES

This section describes the generating function for the arithmetic degrees of 0-cycles on the arithmetic surface \mathfrak{X} associated to a Shimura curve X over \mathbb{Q} . The series of interest will be analogous to the series $\deg(\phi_2(\tau, L))$ for a complex surface given in Corollary 1.4 and will have the form

$$(4.1) \quad \hat{\phi}_{\deg}(\tau) = \sum_{T \in \text{Sym}_2(\mathbb{Z})} \widehat{\deg}(\widehat{\mathfrak{Z}}(T, v)) q^T,$$

where $\tau = u + iv \in \mathfrak{H}_2$ and the $\widehat{\mathfrak{Z}}(T, v)$'s are certain classes in $\widehat{CH}^2(\mathfrak{X})$, the top arithmetic Chow group of the arithmetic surface \mathfrak{X} , [4]. As in the second example of section 1, the definition of the relevant cycles will depend on a modular interpretation. To give a more detailed explanation, it is convenient to begin with the geometric situation of section 3.

Fix an indefinite division quaternion algebra B over \mathbb{Q} . The space $V = \{x \in B \mid \text{tr}(x) = 0\}$, equipped with the restriction of the reduced norm of B , is a three dimensional quadratic space over \mathbb{Q} of signature $(1, 2)$. The group $H = B^\times \simeq \text{GSpin}(V)$ acts on V by conjugation. The Shimura curve X_K associated to a compact open subgroup $K \subset H(\mathbb{A}_f)$, is a moduli space for abelian surfaces with \mathcal{O}_B -action and level structure, where \mathcal{O}_B is a maximal order in B . For example, suppose that $K = (\mathcal{O}_B \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})^\times$. Then $X = X_K$ is the (coarse) moduli scheme over \mathbb{Q} for pairs (A, ι) consisting of an abelian surface A together with an action $\iota : \mathcal{O}_B \hookrightarrow \text{End}(A)$ of \mathcal{O}_B , and $X(\mathbb{C}) \simeq \Gamma_B \backslash \mathfrak{H}$, where $\Gamma_B = (\mathcal{O}_B)_+^\times$ is the group of norm 1 units on \mathcal{O}_B .

The 0-cycles on X defined in section 3 can also be defined by specifying additional endomorphisms.

DEFINITION 4.1. — *For an abelian surface (A, ι) with \mathcal{O}_B -action, the space of special endomorphisms is*

$$V(A, \iota) = \{x \in \text{End}(A) \mid \iota(b)x = x\iota(b) \text{ and } \text{tr}(x) = 0\}.$$

This space is equipped with a \mathbb{Z} -valued quadratic form defined, for a special endomorphism $x \in V(A, \iota)$, by $x^2 = -Q(x) \cdot 1_A$.

DEFINITION 4.2. — *For $t \in \mathbb{Z}_{>0}$, the special cycle $Z(t)$ is the locus of triples (A, ι, x) where $x \in V(A, \iota)$ with $Q(x) = t$.*

In fact, one then has $Z(t) = Z(t, \varphi)$ where $Z(t, \varphi)$ is the 0-cycle on $X = X_K$ defined in section 3 for the weight function $\varphi \in S(V(\mathbb{A}_f))^K$, the characteristic function of the set $V(\mathbb{A}_f) \cap (\mathcal{O}_B \otimes_{\mathbb{Z}} \hat{\mathbb{Z}})$.

The 0-cycle $Z(t)$ on X is rational over \mathbb{Q} and is analogous to the set of CM points on the modular curve associated to the imaginary quadratic field $\mathbb{Q}(\sqrt{-t})$. For $\tau = u + iv \in \mathfrak{H}$,

the upper halfplane and $q = e(\tau)$, the degree generating function (3.1) in the present case

$$(4.2) \quad \phi_{\text{deg}}(\tau) = \text{vol}(X) + \sum_{t>0} \text{deg}(Z(t)) q^t$$

is the value at $s = \frac{1}{2}$ of an Eisenstein series of weight $\frac{3}{2}$ for a congruence subgroup of $SL_2(\mathbb{Z})$, cf. Proposition 3.2. Here $\text{vol}(X)$ is the volume of $X(\mathbb{C}) \simeq \Gamma_B \backslash \mathfrak{H}$ with respect to $-\frac{1}{2\pi} y^{-2} dx \wedge dy$.

Now consider the arithmetic case. Let \mathfrak{X} be the coarse moduli scheme over $S = \text{Spec}(\mathbb{Z})$ for abelian surfaces (A, ι) with \mathcal{O}_B -action satisfying Drinfeld’s ‘special’ condition [2]. The arithmetic surface \mathfrak{X} has generic fiber $\mathfrak{X}_{\mathbb{Q}} = X$, the canonical model of the Shimura curve; \mathfrak{X} has good reduction at all primes $p \nmid D(B)$, where $D(B)$ is the product of the primes p such that $B_p = B \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is a division algebra. The points of \mathfrak{X} over an algebraically closed field k correspond to isomorphism classes of (A, ι) ’s over k .

For an abelian scheme (A, ι) over a connected base, the space of special endomorphisms $V(A, \iota)$ with its \mathbb{Z} -valued quadratic form Q is defined as before.

DEFINITION 4.3. — For $T \in \text{Sym}_2(\mathbb{Z})_{>0}$, the arithmetic special cycle $\mathfrak{Z}(T)$ is the locus of triples (A, ι, x) , where $x = [x_1, x_2] \in V(A, \iota)^2$ is a pair of special endomorphisms $x_i \in V(A, \iota)$ with $Q(x) = \frac{1}{2}((x_i, x_j))_{i,j} = T$.

PROPOSITION 4.4. — [23]. The cycle $\mathfrak{Z}(T)$ is either empty or is supported in the set of supersingular points of a single fiber \mathfrak{X}_p , where p is determined by T , as described in Lemma 4.5 below. If $p \nmid D(B)$, or if $p \mid D(B)$ but $p \nmid T$, then $\mathfrak{Z}(T)$ is a 0-cycle on \mathfrak{X}_p . If $p \mid D(B)$ and $p \mid T$, then $\mathfrak{Z}(T)$ is a union, with multiplicities, of components of the fiber \mathfrak{X}_p (and some additional embedded components).

SKETCH OF PROOF. — The proof of this result illustrates the way in which the basic structure of the cycle $\mathfrak{Z}(T)$ is determined by the space $V(A, \iota)$. Observe that, for a geometric point (A, ι) of \mathfrak{X} and viewing A up to isogeny,

$$\text{End}^0(A) = \text{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq \begin{cases} B & \text{if } A \text{ is simple,} \\ M_2(C) & \text{if } A \simeq E \times E \text{ with } E \text{ ordinary,} \\ M_2(\mathbb{B}) & \text{if } A \simeq E \times E \text{ with } E \text{ supersingular,} \end{cases}$$

for an elliptic curve E . In the second case, $\text{End}^0(E) \simeq C$ is an imaginary quadratic field which splits B , and, in the third case, which occurs only in characteristic $p > 0$, $\text{End}^0(E) \simeq \mathbb{B}$ is the quaternion algebra over \mathbb{Q} ramified at ∞ and p . In this case, write $M_2(\mathbb{B}) \simeq B \otimes B^{(p)}$ where $B^{(p)}$ is the definite quaternion algebra over \mathbb{Q} whose local invariants differ from those of B precisely at ∞ and p . Then, in the three cases,

$\text{End}^0(A, \iota) \simeq \mathbb{Q}$, C , and $B^{(p)}$ respectively, and

$$V^0(A, \iota) = V(A, \iota) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq \begin{cases} 0 \\ C^0 & = \{x \in C \mid \text{tr}(x) = 0\}, \\ V^{(p)} & = \{x \in B^{(p)} \mid \text{tr}(x) = 0\}. \end{cases}$$

It follows that $\mathfrak{Z}(T)_{\mathbb{Q}} = \emptyset$, i.e., the cycle $\mathfrak{Z}(T)$ has no points in characteristic 0. If $\mathfrak{Z}(T)$ meets the fiber \mathfrak{X}_p at p , then $\mathfrak{Z}(T) \cap \mathfrak{X}_p$ is contained in the supersingular locus of \mathfrak{X}_p , and the rational quadratic space $V^{(p)}$ represents T , i.e., there exists a pair of vectors $x = [x_1, x_2] \in V^{(p)}(\mathbb{Q})$ such that $Q^{(p)}(x) = T$. This last condition implies that $\mathfrak{Z}(T)$ is supported in a single fiber, due to the following simple observation.

LEMMA 4.5. — (i) The quadratic spaces $V^{(p)}$ have the same determinant as V , i.e., $\det(V^{(p)}) = \det(V) \in \mathbb{Q}^{\times}/\mathbb{Q}^{\times,2}$.

(ii) For a given $T \in \text{Sym}_2(\mathbb{Q})$ with $\det(T) \neq 0$, there is a unique three dimensional rational quadratic space V_T with $\det(V_T) = \det(V)$ which represents T ; the quadratic form on V_T has matrix

$$(4.3) \quad \begin{pmatrix} T & \\ & \det(V)/\det(T) \end{pmatrix}.$$

Thus, if V_T is not isomorphic to any of the $V^{(p)}$'s then $\mathfrak{Z}(T)$ is empty, while if $V_T \simeq V^{(p)}$ for some p , then $\mathfrak{Z}(T)$ is supported in the supersingular locus of \mathfrak{X}_p .

DEFINITION 4.6. — A matrix $T \in \text{Sym}_2(\mathbb{Z})_{>0}$ will be called irregular if $V_T \simeq V^{(p)}$ where $p|D(B)$ and $p|T$. Otherwise T will be called regular.

The assertions about the irregular case require a more detailed argument, using the p -adic uniformization of the formal completion of \mathfrak{X}_p , [23]. □

The arithmetic Chow group $\widehat{CH}^2(\mathfrak{X})$ is generated by pairs (\mathfrak{Z}, g) where \mathfrak{Z} is a 0-cycle on \mathfrak{X} and g is a smooth $(1, 1)$ -form on $X(\mathbb{C})$, modulo a suitable equivalence, [4], [32]. There is a degree map $\widehat{\text{deg}} : \widehat{CH}^2(\mathfrak{X}) \rightarrow \mathbb{R}$.

For $T \in \text{Sym}_2(\mathbb{Z})_{>0}$ regular, let

$$(4.4) \quad \widehat{\mathfrak{Z}}(T, v) = (\mathfrak{Z}(T), 0) \in \widehat{CH}^2(\mathfrak{X}).$$

Then $\mathfrak{Z}(T) = \text{Spec}(R(T))$ for an Artin ring $R(T)$ in which p is nilpotent, and the corresponding positive definite coefficients of the generating function (4.1) are given by

$$(4.5) \quad \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T, v)) = \log |R(T)|.$$

For nonsingular $T \in \text{Sym}_2(\mathbb{Z})$ with signature $(1, 1)$ or $(0, 2)$, set

$$(4.6) \quad \widehat{\mathfrak{Z}}(T, v) = (0, \mathfrak{g}(T, v)) \in \widehat{CH}^2(\mathfrak{X}),$$

where $\mathfrak{g}(T, v)$ is a smooth $(1, 1)$ -form, depending on T and v , which is described in (6.3) of section 6.

Finally, if $T \in \text{Sym}_2(\mathbb{Z})_{\geq 0}$ has rank 1, then, in effect, only one special endomorphism had been imposed, so that there is an associated divisor $\mathfrak{Z}(T)$ on \mathfrak{X} . A Green's function $\Xi(T, v)$ for this divisor is constructed in section 6, below. This function continues to make sense when $T \leq 0$. There is a resulting class

$$(4.7) \quad \widehat{\mathfrak{Z}}(T, v) = (\mathfrak{Z}(T), \Xi(T, v)) \in \widehat{CH}^1(\mathfrak{X}),$$

where $\mathfrak{Z}(T)$ is empty when $T \leq 0$. Let $\widehat{\omega}_{\mathfrak{X}}$ be the relative dualizing sheaf of \mathfrak{X} over $\text{Spec}(\mathbb{Z})$, with metric coming from the uniformization of $X(\mathbb{C})$ by D , viewed as an element of $\widehat{CH}^1(\mathfrak{X})$ via the isomorphism $\widehat{\text{Pic}}(\mathfrak{X}) \simeq \widehat{CH}^1(\mathfrak{X})$, and let

$$(4.8) \quad \widehat{\omega}_{\mathfrak{X}}(v) = \widehat{\omega}_{\mathfrak{X}} + (0, \log \det(v)) \in \widehat{CH}^1(\mathfrak{X}).$$

This class plays a role analogous to that of $[\omega]$, the Kähler class, in section 3 above.

Using the arithmetic intersection pairing $\widehat{CH}^1(\mathfrak{X}) \times \widehat{CH}^1(\mathfrak{X}) \rightarrow \widehat{CH}^2(\mathfrak{X})$, the full arithmetic generating function (4.1), analogous to ϕ_{deg} given in Corollary 1.4 and (3.1) in the geometric case, is then

$$(4.9) \quad \widehat{\phi}_{\text{deg}}(\tau) = \widehat{\text{deg}}(\widehat{\omega}_{\mathfrak{X}}(v)^2) + \sum_{\text{rank}(T)=1} \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T, v) \cdot \widehat{\omega}_{\mathfrak{X}}(v)) q^T + \sum_{\substack{T \\ \det(T) \neq 0}} \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T, v)) q^T.$$

where suitable terms have been added[†] for irregular T .

In the present situation, the construction of Remark 2.3 yields an incoherent Eisenstein series $E(\tau, s, \varphi)$ of weight $\frac{3}{2}$ associated to $\varphi \in S(V(\mathbb{A}_f)^2)$, the characteristic function of $(\widehat{\mathcal{O}}_B \cap V(\mathbb{A}_f))^2$.

CONJECTURE 4.7. — *The generating function $\widehat{\phi}_{\text{deg}}(\tau)$ is a Siegel modular form of weight $\frac{3}{2}$, more precisely*

$$(4.10) \quad \widehat{\phi}_{\text{deg}}(\tau) = \text{vol}(X(\mathbb{C})) \cdot E'(\tau, 0, \varphi).$$

The main results in this direction assert that many of the Fourier coefficients of the two series coincide. Recall that $D(B)$ is the product of the primes p at which B_p is division. Also put $V^{(\infty)} = V$, so that there is a rational quadratic space associated to each place of \mathbb{Q} . By Lemma 4.5, a given nonsingular $T \in \text{Sym}_2(\mathbb{Q})$ is represented by at most one of these spaces. If T is positive definite, this space, if it exists, must be one of the $V^{(p)}$'s for a finite prime p , while if T has signature $(1, 1)$ or $(0, 2)$, then this space can only be $V = V^{(\infty)}$.

[†]by a still not very satisfactory procedure

THEOREM 4.8. — [14]. *Suppose that $T \in \text{Sym}_2(\mathbb{Z})$ is nonsingular.*

(i) *If T is not represented by any $V^{(p)}$, then the T -th Fourier coefficient of both sides of (4.10) vanish.*

(ii) *If T is represented by $V^{(p)}$ with $p \nmid 2D(B)$, including $p = \infty$, then the T -th Fourier coefficients of the two sides of (4.10) agree, i.e.,*

$$(4.11) \quad \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T, v)) q^T = \text{vol}(X(\mathbb{C})) \cdot E'_T(\tau, 0, \varphi).$$

In fact, work in progress [25] should extend (4.11) to all T of rank 1.

The proof of (ii) is similar to the proof of the main identity at the heart of the Gross–Zagier formula [8]; it amounts to an explicit computation of the two quantities, one from arithmetic geometry, the other from automorphic forms. A sketch of the proof is given in the next two sections.

5. NON-SINGULAR FOURIER COEFFICIENTS

The Fourier coefficients of the central derivative of an incoherent Eisenstein series $E(\tau, s, \varphi)$ associated to a rational quadratic space V of signature $(n-1, 2)$, as defined as in Remark 2.3, have an interesting structure.

For each prime $p \leq \infty$, define a quadratic space $V^{(p)}$ of dimension $n+1$ and the same determinant as that of V as follows. Let $V^{(\infty)} = V$. For a finite prime p , $V^{(p)}$ has signature $(n+1, 0)$ and local Hasse invariants at finite primes ℓ determined by

$$(5.1) \quad \epsilon_\ell(V^{(p)} \otimes_{\mathbb{Q}} \mathbb{Q}_\ell) = \begin{cases} \epsilon_\ell(V) & \text{if } \ell \neq p, \text{ and} \\ -\epsilon_p(V_p) & \text{if } \ell = p. \end{cases}$$

For each nonsingular $T \in \text{Sym}_n(\mathbb{Q})$, there is a rational quadratic space V_T of dimension $n+1$ and the same determinant as V defined by equation (4.3).

It is well known that if the function $\varphi = \otimes_{p < \infty} \varphi_p \in S(V(\mathbb{A}_f)^n)$ is factorizable and $\text{Re}(s) > \frac{n+1}{2}$, then each nonsingular Fourier coefficient of $E(\tau, s, \varphi)$ is given by a product

$$E_T(\tau, s, \varphi) = W_{T, \infty}^{\frac{n+1}{2}}(\tau, s) \cdot \prod_{p < \infty} W_{T, p}(s, \varphi_p)$$

of local (degenerate) Whittaker functions, [20]. In classical language, the archimedean factor is a confluent hypergeometric function of a matrix argument studied by Shimura [30], while the product over the finite primes is the Siegel series.

PROPOSITION 5.1. — [14], section 6. (i) *If V_T is not isomorphic to any $V^{(p)}$, then*

$$E'_T(\tau, 0, \varphi) = 0.$$

(ii) *If $V_T \simeq V^{(p)}$ for a finite prime, then $W_{T, p}(0, \varphi_p) = 0$ and*

$$E'_T(\tau, 0, \varphi) = q^T \cdot W_{T, p}(s, \varphi_p)' \Big|_{s=0} \cdot A_T^{(p)}(\varphi),$$

where $A_T^{(p)}(\varphi)$ is, up to a factor at p , the Fourier coefficient of a theta integral (cf. (2.2)) attached to $V^{(p)}$.

(iii) If $V_T \simeq V^{(\infty)} = V$,

$$E'_T(\tau, 0, \varphi) = W_{T, \infty}^{\frac{n+1}{2}}(\tau, s)'|_{s=0} \cdot A_T^{(\infty)}(\varphi),$$

where $A_T^{(\infty)}(\varphi)$ is the T -th Fourier coefficient of a theta integral attached to V .

Idea of proof of Theorem 4.8 for $p < \infty$. Restricting to the case $n = 2$ as in section 4, suppose that $T \in \text{Sym}_2(\mathbb{Z})$ is nonsingular with $V_T \simeq V^{(p)}$ for a finite prime $p \nmid 2D(B)$. In this case, using Proposition 5.1, the identity to be proved in (ii) of Theorem 4.8 amounts to

$$(5.2) \quad \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T, v)) = \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T), 0) = \text{vol}(X(\mathbb{C})) \cdot W_{T, p}(s, \varphi_p)'|_{s=0} \cdot A_T^{(p)}(\varphi).$$

This identity, which is of the same nature as the identities between heights and Fourier coefficients involved in the Gross-Zagier formula, is proved by computing the two sides explicitly.

On the geometric side, since T is regular, $\mathfrak{Z}(T)$ is a collection of supersingular points on \mathfrak{X}_p , each counted with a certain multiplicity. This multiplicity is the length of the local Artin ring associated to the deformations of the triple (A, ι, x) , where x is a pair of special endomorphisms. By the Serre–Tate Theorem, one can pass to the deformations of $(A(p), \iota, x)$ where $A(p)$ is corresponding p -divisible group with the action ι of $(\mathcal{O}_B)_p = \mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p \simeq M_2(\mathbb{Z}_p)$ and a pair of special endomorphisms x . Since A is isogenous to $E \times E$ for a supersingular elliptic curve E , one reduces, via the idempotents in $(\mathcal{O}_B)_p$, to the problem of deforming $(E(p), x)$ for a pair $x = [x_1, x_2]$ of endomorphisms of the p -divisible group of such a curve. Note that $E(p)$ is a p -divisible formal group of dimension 1 and height 2. The length of the associated Artin ring is then obtained by specializing a beautiful result of Gross and Keating [6].

They consider the deformations of a collection $(\mathbb{X}, \mathbb{X}', y)$ where \mathbb{X} and \mathbb{X}' are formal groups of dimension 1 and height 2 over $\overline{\mathbb{F}}_p$ and $y = [y_1, y_2, y_3]$ is a triple of nonzero isogenies $y_i : \mathbb{X} \rightarrow \mathbb{X}'$. Note that the universal deformation ring of the pair $(\mathbb{X}, \mathbb{X}')$ is $W[[t, t']]$, where $W = W(\overline{\mathbb{F}}_p)$ is the ring of Witt vectors of $\overline{\mathbb{F}}_p$.

PROPOSITION 5.2. — (*Gross-Keating*, [6], *Proposition 5.4*) *Suppose that the matrix Q of inner products of the triple $y = [y_1, y_2, y_3]$ with respect to the degree quadratic form on $\text{Hom}(\mathbb{X}, \mathbb{X}')$ has invariants $a_1 \leq a_2 \leq a_3$. For p odd, this means that $Q \in \text{Sym}_3(\mathbb{Z}_p)$ is $GL_3(\mathbb{Z}_p)$ equivalent to $\text{diag}(\epsilon_1 p^{a_1}, \epsilon_2 p^{a_2}, \epsilon_3 p^{a_3})$ for units $\epsilon_i \in \mathbb{Z}_p^\times$. Then the length of the*

deformation ring of $(\mathbb{X}, \mathbb{X}', y)$ is given by:

$$(5.3) \quad \sum_{i=0}^{a_1-1} (i+1)(a_1+a_2+a_3-3i)p^i + \sum_{i=a_1}^{(a_1+a_2-2)/2} (a_1+1)(2a_1+a_2+a_3-4i)p^i + \frac{1}{2}(a_1+1)(a_3-a_2+1)p^{(a_1+a_2)/2}.$$

if $a_1 + a_2$ is even, and

$$\sum_{i=0}^{a_1-1} (i+1)(a_1+a_2+a_3-3i)p^i + \sum_{i=a_1}^{(a_1+a_2-1)/2} (a_1+1)(2a_1+a_2+a_3-4i)p^i$$

if $a_1 + a_2$ is odd.

Specialized to the case where y_1 is an isomorphism, so that $a_1 = 0$, one obtains an explicit formula for the multiplicity $\delta_p(T)$ of a point in $\mathfrak{Z}(T)$, and, in particular, observes that this multiplicity depends only on the $GL_2(\mathbb{Z}_p)$ -equivalence class of T , and not on the particular point. Thus, the left hand side of (5.2) has the form $(\delta_p(T) \cdot (\# \text{ points in } \mathfrak{Z}(T)))$.

On the other hand, for $p \neq 2$, the quantity $W_{T,p}(s, \varphi_p)'|_{s=0}$ on the analytic side of (5.2) can be computed from the result of Kitaoka [12]. More precisely, the quadratic form on the lattice $L_p = (\mathcal{O}_B \otimes \mathbb{Z}_p) \cap V_p$ has matrix $S_0 = \text{diag}(1, 1, -1)$ for a suitable basis. Let

$$S_r = \text{diag}(1, 1, -1, \underbrace{1, \dots, 1}_r, \underbrace{-1, \dots, -1}_r)$$

be the quadratic form obtained from S_0 by adding r hyperbolic planes. Then

$$W_{T,p}(r, \varphi_p) = \gamma_p \cdot \alpha_p(S_r, T),$$

for a root of unity γ_p , where

$$\alpha_p(S_r, T) = \lim_{t \rightarrow \infty} p^{-t(3+4r)} |\{x \in M_{3+2r,2}(\mathbb{Z}_p/p^t\mathbb{Z}_p) \mid {}^t x S_r x \equiv T \pmod{p^t}\}|$$

is the classical representation density of T by S_r as defined by Siegel, [12]. This quantity has the form $\alpha_p(S_r, T) = A_p(q^{-r}, T)$ for a polynomial $A_p(X, T)$, which was first calculated in this case by Kitaoka [12][‡]. When T is such that $W_{T,p}(0, \varphi_p) = 0$, and $T \in \text{Sym}_2(\mathbb{Z}_p)$ is $GL_2(\mathbb{Z}_p)$ -equivalent to $\text{diag}(\epsilon_1 p^a, \epsilon_2 p^b)$, then Kitaoka's formula yields

$$(5.4) \quad \begin{aligned} W_{T,p}(s, \varphi_p)'|_{s=0} &= -\log(p) \gamma_p \frac{\partial}{\partial X} \{A_p(X, T)\}|_{X=1} \\ &= \log(p) \gamma_p (1 - p^{-2}) \\ &\quad \times \begin{cases} \sum_{j=0}^{(a-1)/2} (a+b-4j)p^j & \text{if } a \text{ is odd} \\ \sum_{j=0}^{(a-2)/2} (a+b-4j)p^j + \frac{1}{2}(b-a+1)p^{a/2} & \text{if } a \text{ is even} \\ & \text{and } b \text{ is odd} \end{cases} \end{aligned}$$

[‡]Recently, it has been calculated in general, for $p \neq 2$, by F. Sato and Y. Hironaka, [29]

Remarkably,

$$(5.5) \quad W_{T,p}(s, \varphi_p)' \Big|_{s=0} = \log(p) \gamma_p (1 - p^{-2}) \delta_p(T),$$

where $\delta_p(T)$ is the multiplicity computed via the Gross–Keating formula! Finally, by a straightforward counting argument, the number of points in $\mathfrak{Z}(T)$ is given by $A_T^{(p)}(\varphi)$, up to simple factors, independent of T , which compensate for the extra $p^2 - 1$ and the factor $\text{vol}(X(\mathbb{C}))$. \square

6. GREEN’S FUNCTIONS AND WHITTAKER FUNCTIONS

The classes $\widehat{\mathfrak{Z}}(T, v)$ in the arithmetic Chow groups of integral models of Shimura curves involve the Green’s functions defined in [14], section 11. This section describes the construction of such functions. First, suppose that V is a rational quadratic space of signature $(n - 1, 2)$, as in section 3, and recall that for $t \in \mathbb{Q}_{>0}$ and for $\varphi \in S(V(\mathbb{A}_f))^K$, there is a divisor $Z(t, \varphi)$ on X_K , given as a weighted sum of the images in X_K of the divisors D_x in D . A Green’s function of logarithmic type for $Z(t, \varphi)$ in the sense of Gillet–Soulé [4] can be constructed by averaging rapidly decreasing Green’s functions for D_x ’s. Next, in the case of a Shimura curve ($n = 2$), the smooth $(1, 1)$ –form $\mathfrak{g}(T, v)$ used in the construction of the terms for indefinite T ’s in the generating function of arithmetic degrees is defined via the star product. The notation is that of section 3.

For an oriented negative 2–plane $z \in D$, let pr_z be the projection to z with respect to the orthogonal decomposition $V(\mathbb{R}) = z + z^\perp$. For $x \in V(\mathbb{R})$, $x \neq 0$, the quantity

$$R(x, z) = -(\text{pr}_z(x), \text{pr}_z(x))$$

is nonnegative and vanishes if and only if $z \in D_x$. Let

$$\xi(x, z) = -\text{Ei}(-2\pi R(x, z)),$$

where, for $r > 0$,

$$-\text{Ei}(-r) = \int_1^\infty \frac{e^{-rt}}{t} dt$$

is the exponential integral. Since $-\text{Ei}(-r)$ decays exponentially as r goes to infinity and behaves like $-\log(r) + O(1)$ as r goes to 0, the function $\xi(x)$ has a logarithmic singularity along the (possibly empty) divisor D_x and decays very rapidly away from D_x . In addition, it satisfies the Green’s equation:

$$dd^c \xi(x) + \delta_{D_x} = \mu(x)$$

for a smooth $(1, 1)$ –form $\mu(x)$ on D , and hence defines a Green’s form of log type for D_x , in the sense of Gillet–Soulé.

Note that $R(hx, hz) = R(x, z)$ for $h \in \text{GSpin}(V)(\mathbb{R})$, so that $\xi(hx, hz) = \xi(x, z)$ as well. For $v > 0$ and $t \in \mathbb{Q}_{>0}$, and a weight function $\varphi \in S(V(\mathbb{A}_f))^K$, the sum

$$(6.1) \quad \mathfrak{g}(t, v, \varphi)(z, h) = \sum_{\substack{x \in V(\mathbb{Q}) \\ Q(x)=t}} \varphi(h^{-1}x) \xi(x, z)$$

depends only on the orbit $H(\mathbb{Q})(z, h)K$ of the point $(z, h) \in D \times H(\mathbb{A}_f)$. Thus, $\mathfrak{g}(t, v, \varphi)$ defines a Green's function of logarithmic type for the divisor $Z(t, \varphi)$, defined in section 3 (for $r = 1$), on X_K .

In the case $n = 2$, for the space V considered in section 4, and a pair of vectors $x = [x_1, x_2] \in V(\mathbb{R})^2$ such that $\det(Q(x)) \neq 0$, the integral

$$\Lambda(x) = \int_D \xi(x_1) * \xi(x_2)$$

of the $*$ -product of the associated Green's functions [4] is well defined and satisfies $\Lambda(hx) = \Lambda(x)$ for all $h \in \text{GSpin}(V)(\mathbb{R})$. This implies that $\Lambda(x)$ actually only depends on the matrix of inner products $Q(x)$, i.e.,

$$(6.2) \quad \Lambda(x) =: \Lambda_0(Q(x)).$$

One can view $\Lambda(x)$ as the 'archimedean height pairing' of the 0-cycles D_{x_1} and D_{x_2} in D . In fact this quantity has the following rather surprising additional invariance:

THEOREM 6.1. — [14], section 13. *For any $k \in SO(2)$, and any x ,*

$$\Lambda(xk) = \Lambda(x).$$

Note that, even when $Q(x_1) > 0$ and $Q(x_2) > 0$, so that one initially has a pair of 0-cycles D_{x_1} and D_{x_2} in D , eventually, after rotation, one encounters a pair of vectors x'_1 and x'_2 with $Q(x'_2) < 0$, so that the cycle $D_{x'_2}$ has vanished (!) and is replaced, in some sense, by the geodesic arc $\{z \in D \mid x'_2 \in z\}$. Nonetheless, the 'archimedean height pairing' $\Lambda(x)$ is invariant under such a deformation.

It follows that, writing $v \in \text{Sym}_2(\mathbb{R})_{>0}$ as $v = a^t a$ for $a \in \text{GL}_2(\mathbb{R})^+$, the quantity $\Lambda(xa)$ depends only on v and not on the choice of a . Then, for a nonsingular $T \in \text{Sym}_2(\mathbb{Z})$, one has a smooth $(1, 1)$ -form

$$(6.3) \quad \mathfrak{g}(T, v) = \left(\sum_{\substack{x \in V(\mathbb{Q})^2 \\ Q(x) = T \\ \text{mod } \Gamma_B}} \varphi(x) \Lambda(xa) \right) \mu$$

on X , where $\mu = y^{-2} dx \wedge dy$, φ is the characteristic function of the set $(\widehat{\mathcal{O}}_B^2 \cap V(\mathbb{A}_f)^2)$, and $\Gamma_B = \mathcal{O}_B^\times$. This is the form used in the definition (4.6) of $\widehat{\mathfrak{Z}}(T, v)$ when T has signature $(1, 1)$ or $(0, 2)$. Note that the sum is nonempty precisely when V represents T .

Idea of proof of Theorem 4.8 for $p = \infty$. Again using Proposition 5.1, the identity to be proved in this case is

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T, v)) q^T = \text{vol}(X(\mathbb{C})) \cdot W_{T, \infty}^{\frac{n+1}{2}}(\tau, s)' \Big|_{s=0} \cdot A_T^{(\infty)}(\varphi).$$

The left hand side is simply

$$(6.4) \quad q^T \cdot \Lambda_0({}^t a T a) \cdot \sum_{\substack{x \in V(\mathbb{Q})^2 \\ Q(x) = T \\ \text{mod } \Gamma_B}} \varphi(x),$$

where $v = a^t a$, as before and Λ_0 is given by (6.2). Then some transformations of the integral representations of the matrix argument confluent hypergeometric function of Shimura’s paper [30] together with a manipulation of the integral defining Λ shows that

$$W_{T, \infty}^{\frac{n+1}{2}}(\tau, s)' \Big|_{s=0} = c_\infty \cdot \Lambda_0({}^t a T a) \cdot q^T,$$

where c_∞ is an innocuous constant. Again, the sum in (6.4) counts pairs of lattice vectors modulo Γ_B and coincides with $A_T^{(\infty)}$, the Fourier coefficient of the theta integral, up to a constant which absorbs c_∞ and provides the required $\text{vol}(X(\mathbb{C}))$ factor. \square

7. FURTHER RESULTS

One would like to identify the central derivative $E'(\tau, 0, \varphi)$ of the incoherent Eisenstein series (2.3) as a generating function for arithmetic degrees in the general case. In the series of papers [24], [22], [21] the cases $n = 1, 3$ and 4 are considered. In each of these cases (and also for $n = 5$), the Shimura variety X associated to $H = \text{GSpin}(V)$ is of PEL type, i.e., can be interpreted as a moduli space for abelian varieties (A, ι) with a specified endomorphism ring, due to the existence of an accidental isomorphism. This allows one to give a modular definition of an integral model \mathfrak{X}_K of X_K , at least over $\text{Spec}(\mathbb{Z}[N^{-1}])$ for a suitable N depending on the compact open subgroup K . For each abelian scheme (A, ι) , there is a space of special endomorphisms $V(A, \iota)$, equipped with a \mathbb{Z} -valued quadratic form. Special cycles on \mathfrak{X} are then defined by imposing collections of such endomorphisms, as in section 4 and the example of section 1.3 above.

The case $n = 1$. This case is considered in [24] and is described in classical language in section 1.3 above. From the point of view now developed it amounts to the following. An imaginary quadratic field $\mathfrak{k} = \mathbb{Q}(\sqrt{-d})$ with quadratic form $Q(x) = -N_{\mathfrak{k}/\mathbb{Q}}(x)$ gives a rational quadratic space (V, Q) of signature $(0, 2)$. The group $\text{GSpin}(V)$ is then the torus over \mathbb{Q} with $\text{GSpin}(V)(\mathbb{Q}) \simeq \mathfrak{k}^\times$, and $\mathfrak{X} \simeq \text{Spec}(\mathcal{O}_H)$ is the restriction of scalars to $\text{Spec}(\mathbb{Z})$ of the coarse moduli space over $\mathcal{O}_{\mathfrak{k}}$ of elliptic curves (A, ι) with complex multiplication

by $\mathcal{O}_{\mathfrak{k}}$. In this case the space of special endomorphisms (1.6) is

$$V(A, \iota) = \{x \in \text{End}(A) \mid x\iota(b) = \iota(\bar{b})x\}.$$

This space is zero unless A is a supersingular elliptic curve in characteristic p , where p is not split in \mathfrak{k} , in which case, $V(A, \iota) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq V^{(p)}$, the rational quadratic space given by $V^{(p)} = \mathfrak{k}$ with quadratic form $Q^{(p)}(x) = pN_{\mathfrak{k}/\mathbb{Q}}(x)$. For $\tau = u + iv \in \mathfrak{H}$, the generating function is then given by

$$(7.1) \quad \hat{\phi}_{\text{deg}}(\tau) = \sum_{t \in \mathbb{Z}} \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(t, v)) q^t.$$

Here, for $t > 0$, $\widehat{\mathfrak{Z}}(t, v) = (\mathfrak{Z}(t), 0) \in \widehat{CH}^1(\mathfrak{X})$ where $\mathfrak{Z}(t)$ is the locus of (A, ι, x) 's where $x \in V(A, \iota)$ with $Q(x) = -x^2 = t$. For $t < 0$, $\widehat{\mathfrak{Z}}(t, v) = (0, \mathfrak{g}(t, v)) \in \widehat{CH}^1(\mathfrak{X})$, where $\mathfrak{g}(t, v)$ is the function on $X(\mathbb{C}) \simeq H(\mathbb{Q}) \backslash D \times H(\mathbb{A}_f) / K$ given by (6.1). Here $K \simeq \widehat{\mathcal{O}}_{\mathfrak{k}}^{\times}$. Note that in this case, D consists of two points (via the two possible orientations of $V(\mathbb{R})$), and $R(x, z) = 2N_{\mathfrak{k}/\mathbb{Q}}(x)$, so that $\xi(xa, z) = -\text{Ei}(-4\pi v N_{\mathfrak{k}/\mathbb{Q}}(x))$, precisely as in [24]. This gives an improved version of Theorem 1.7:

THEOREM 7.1. — [24]. *When $d \equiv 3 \pmod{4}$ is a prime and for a suitable definition of the constant term $\widehat{\text{deg}}(\widehat{\mathfrak{Z}}(0, v))$,*

$$\hat{\phi}_{\text{deg}}(\tau) = \text{vol}(X(\mathbb{C})) \cdot E'(\tau, 0, \varphi),$$

where φ is the characteristic function of $\widehat{\mathcal{O}}_{\mathfrak{k}} \subset V(\mathbb{A}_f)$ and $\text{vol}(X(\mathbb{C})) = h_{\mathfrak{k}}$ is the class number of \mathfrak{k} .

The result is proved by a direct calculation of both sides, using the results of Gross [5] to compute multiplicities on the moduli space. The restriction to prime discriminant is only made to streamline the calculations.

The higher dimensional cases treated so far exhibit some new phenomena.

The case $n = 3$. This case is considered in [22]. The incoherent Eisenstein series (2.2) associated to a rational quadratic space of signature $(2, 2)$ will have weight 2 and genus 3. To define the relevant moduli problem and generating function for arithmetic degrees, let $C(V) = C^+(V) \oplus C^-(V)$ be the Clifford algebra of V with its \mathbb{Z}_2 -grading. Then the center of $C^+(V)$ is a real quadratic field \mathfrak{k} (possibly $\mathfrak{k} = \mathbb{Q} \oplus \mathbb{Q}$), and $C^+(V)$ has the form $B_0 \otimes_{\mathbb{Q}} \mathfrak{k}$ for an indefinite quaternion algebra B_0 over \mathbb{Q} . The associated Shimura variety X is a surface whose complex points parametrize polarized abelian varieties (A, ι) of dimension 8 with an action of a maximal order in $C(V) \otimes \mathfrak{k}$. Included among the X 's are products of modular curves ($\mathfrak{k} = \mathbb{Q} \oplus \mathbb{Q}$, $B_0 = M_2(\mathbb{Q})$), products of Shimura curves ($\mathfrak{k} = \mathbb{Q} \oplus \mathbb{Q}$, B_0 division), Hilbert-Blumenthal surfaces (\mathfrak{k} a field, $B_0 = M_2(\mathbb{Q})$) and their twisted (quaternionic) analogues (\mathfrak{k} a field, B_0 division). An integral model \mathfrak{X} of X over $\text{Spec}(\mathbb{Z}[N^{-1}])$, is defined as the moduli space of polarized abelian schemes with such an

action, level structure, etc. The space of special endomorphisms of a given (A, ι) is then

$$V(A, \iota) = \{x \in \text{End}(A) \mid x \iota(c \otimes a) = \iota(c \otimes \bar{a}) x, \text{ and } x^* = x\}$$

where $*$ denotes the Rosati involution of A . As before, this space has a \mathbb{Z} -valued quadratic form defined by $x^2 = Q(x) \cdot 1_A$. Again, a key point is that, for (A, ι) supersingular, the space $V(A, \iota) \otimes \mathbb{Q} \simeq V^{(p)}$ is the 4 dimensional rational quadratic space defined by (5.1) above. For $T \in \text{Sym}_3(\mathbb{Z})_{>0}$, the special cycle $\mathfrak{Z}(T)$ is the locus of (A, ι, x) 's where $x \in V(A, \iota)^3$ with $Q(x) = T$. This cycle is either empty or is supported in the supersingular locus $\mathfrak{X}_p^{\text{s.s.}}$ of the fiber at p for the unique prime p for which $V_T \simeq V^{(p)}$. Here one assumes that $p \nmid N$, so that p is a prime of good reduction; in particular, p is not ramified in \mathbf{k} .

If a prime $p \nmid N$ splits in \mathbf{k} , then the supersingular locus $\mathfrak{X}_p^{\text{s.s.}}$ consists of a finite set of points. If p is inert in \mathbf{k} , then $\mathfrak{X}_p^{\text{s.s.}}$ has dimension 1 and is, in fact, a union of \mathbb{P}^1 's [31], [22], section 4, crossing at \mathbb{F}_{p^2} rational points.

Let $\varphi \in S(V(\mathbb{A}_f)^3)$ be the characteristic function of $L \otimes \hat{\mathbb{Z}}$ for a lattice $L \subset V(\mathbb{Q})$ such that L_p is self dual for all $p \nmid N$, and let $E(\tau, s, \varphi)$ be the associated incoherent Eisenstein series (2.3) of weight 2.

THEOREM 7.2. — [22]. *Suppose that $T \in \text{Sym}_3(\mathbb{Z})_{>0}$.*

(i) *If V_T is not isomorphic to any $V^{(p)}$, then $\mathfrak{Z}(T)$ is empty and $E'_T(\tau, 0, \varphi) = 0$.*

(ii) *If $V_T \simeq V^{(p)}$ for a prime $p \nmid N$ which splits in \mathbf{k} , then there is a class $\widehat{\mathfrak{Z}}(T) = (\mathfrak{Z}(T), 0) \in \widehat{CH}^3(\mathfrak{X})$ and*

$$(7.2) \quad \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T)) \cdot q^T = C \cdot E'_T(\tau, 0, \varphi),$$

for a constant C independent of T .

(iii) *Suppose that $V_T \simeq V^{(p)}$ for a prime $p \nmid N$ which is inert in \mathbf{k} . If $p \nmid T$, then $\mathfrak{Z}(T)$ consists of a finite number of points, there is a class $\widehat{\mathfrak{Z}}(T) = (\mathfrak{Z}(T), 0) \in \widehat{CH}^3(\mathfrak{X})$ and, again,*

$$(7.3) \quad \widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T)) \cdot q^T = C \cdot E'_T(\tau, 0, \varphi).$$

If $p \mid T$, then $\mathfrak{Z}(T)$ is a union of components of the supersingular locus $\mathfrak{X}_p^{\text{s.s.}}$.

Here, $T \in \text{Sym}_3(\mathbb{Z})_{>0}$ will be called irregular if $V_T \simeq V^{(p)}$ with $p \nmid N$ inert in \mathbf{k} and $p \mid T$. The situation for $p \mid N$, e.g., p ramified in \mathbf{k} , has not yet been studied.

Here, as in (4.5), $\widehat{\text{deg}}(\widehat{\mathfrak{Z}}(T)) = \log |R(T)|$ for the Artin ring $R(T)$ defining $\mathfrak{Z}(T)$, so that the fact that \mathfrak{X} is only defined over $\text{Spec}(\mathbb{Z}[N^{-1}])$ and need not be proper will not be important. These issues will be essential, however, if one wants to define the full generating function $\hat{\phi}_{\text{deg}}(\tau)$ and compare it to $E'(\tau, 0, \varphi)$. The proof of (7.2) and (7.3), where $\mathfrak{Z}(T)$ is a 0-cycle, again comes down to a relation between derivatives of representation densities for quadratic forms and the result of Gross and Keating described in Proposition 5.2 above, now in its full generality. Indeed, that result was obtained in connection with the

study of the derivative of a Siegel–Eisenstein series of weight 2, due to the connection of such a series with the triple product L -function [3], [27], [9], [7].

The case $n = 4$. This case is considered in [21]. Here the Shimura varieties are (twisted) Siegel 3-folds, and the pattern is similar to that for $n = 3$. The one new point is that the ‘regularity’ condition on $T \in \text{Sym}_4(\mathbb{Z})$ required to obtain a 0-cycle in the supersingular locus (which is again a curve with \mathbb{P}^1 components) becomes: $V_T \simeq V^{(p)}$ and T represents 1 over \mathbb{Z}_p . One then obtains an analogue of Theorem 6.2, with the comparison again based on [6].

8. FINAL REMARKS

Beyond the range of the accidental isomorphisms, i.e., for $n \geq 6$, the Shimura varieties associated to $\text{GSpin}(V)$ for rational quadratic spaces V of signature $(n - 1, 2)$ are no longer of PEL type, so that the modular interpretation of points, special endomorphisms, and other tools used before are no longer available. Instead, it will be necessary to work with integral models defined by suitable types of Hodge classes, etc., [26]. Presumably there is a good notion of special endomorphisms or special Hodge classes in this situation which cut out the required cycles. This theory remains to be established.

Even in the range $2 \leq n \leq 5$, many difficulties lie in the way of a full treatment of the generating function for $\hat{\phi}_{\text{deg}}(\tau)$. For example, there is the problem of the contribution of irregular T 's, which occur even in the case of good reduction. The contribution of the singular T 's is ‘global’ and will presumably involve models over \mathbb{Z} , detailed information about the fibers of bad reduction, etc. Much work remains to be done.

In addition, there is the question of defining modular arithmetic generating functions $\hat{\phi}_r(\tau)$, analogous to the series $\phi_r(\tau)$ of Theorem 3.1 above, valued in arithmetic Chow groups $\widehat{CH}^r(\mathfrak{X})$ for arbitrary codimension. Of course the image of such series under the cycle class map should be the generating series discussed in section 3. Recent work of Borcherds [1], which gives a generating function involving the classes of the divisors $Z(t, \varphi)$ on $X_K(\mathbb{C})$ in the Chow group $CH^1(X_K(\mathbb{C}))$, rather than in cohomology, will be relevant here.

In all cases, it remains to work out the modifications required in the case of non-compact quotient.

REFERENCES

- [1] R. E. Borcherds, *The Gross-Kohnen-Zagier Theorem in higher dimensions*, Duke Math. J. **97** (1999), pp. 219–233.

- [2] J.-F. Boutot and H. Carayol, *Uniformisation p -adiques des courbes de Shimura*, Astérisque **196–197** (1991), 45–158.
- [3] P. Garrett, *Decomposition of Eisenstein series: Rankin triple products*, Annals of Math. **125** (1987), 209–235.
- [4] H. Gillet and C. Soulé, *Arithmetic intersection theory*, Jour of IHES, **72** (1990), 94–174.
- [5] B. H. Gross, *On canonical and quasi-canonical liftings*, Inventiones Math. **84** (1986), 321–326.
- [6] B. H. Gross and K. Keating, *On the intersection of modular correspondences*, Invent. Math. **112** (1993), 225–245.
- [7] B. H. Gross and S. Kudla, *Heights and the central critical values of triple product L -functions*, Compositio Math., **81** (1992), 143–209.
- [8] B. H. Gross and D. Zagier, *Heegner points and the derivatives of L -series*, Inventiones math. **84** (1986), 225–320.
- [9] M. Harris and S. Kudla, *The central critical value of a triple product L -function*, Annals of Math. **133** (1991), 605–672.
- [10] S. Helgason, *Differential Geometry and Symmetric Spaces*, Academic Press, 1962.
- [11] F. Hirzebruch and D. Zagier, *Intersection numbers of curves on Hilbert modular surfaces and modular forms of Nebentypus*, Inventiones Math. **36** (1976), 57–113.
- [12] Y. Kitaoka, *A note on local densities of quadratic forms*, Nagoya Math. J. **92** (1983), 145–152.
- [13] S. Kudla, *Algebraic cycles on Shimura varieties of orthogonal type*, Duke Math. J. **86** (1997), 39–78.
- [14] S. Kudla, *Central derivatives of Eisenstein series and height pairings*, Annals of Math. **146** (1997), 545–646.
- [15] S. Kudla and J. Millson, *The theta correspondence and harmonic forms I*, Math. Annalen, **274** (1986), 353–378.
- [16] S. Kudla and J. Millson, *The theta correspondence and harmonic forms II*, Math. Annalen, **277** (1987), 267–314.
- [17] S. Kudla and J. Millson, *Intersection numbers of cycles on locally symmetric spaces and Fourier coefficients of holomorphic modular forms in several complex variables*, IHES **71** (1990), 121–172.
- [18] S. Kudla and S. Rallis, *Degenerate principal series and invariant distributions*, Israel J. Math., **69** (1990), 25–45
- [19] S. Kudla and S. Rallis, *Ramified degenerate principal series representations for $Sp(n)$* , Israel J. Math., **78** (1992), 209–256.
- [20] S. Kudla and S. Rallis, *A regularized Siegel-Weil formula: the first term identity*, Annals of Math. **140** (1994), 1–80.

- [21] S. Kudla and M. Rapoport, *Cycles on Siegel 3-folds and derivatives of Eisenstein series*, Ann. École Norm. Sup. (to appear).
- [22] S. Kudla and M. Rapoport, *Arithmetic Hirzebruch–Zagier cycles*, J. reine angew. Math. **515** (1999), 155–244.
- [23] S. Kudla and M. Rapoport, *Heights on Shimura curves and p -adic uniformization*, Inventiones Math. (to appear)
- [24] S. Kudla, M. Rapoport and T. Yang, *On the derivative of an Eisenstein series of weight 1*, Int. Math. Res. Notices No.7 (1999), 347–385.
- [25] S. Kudla, M. Rapoport and T. Yang (in preparation).
- [26] B. Moonen, *Models of Shimura varieties in mixed characteristics*, in Galois Representations in Arithmetic Algebraic Geometry, A. J. Scholl and R. L. Taylor, eds., London Math. Society Lecture Note Series **254**, Cambridge Univ. Press, 1998.
- [27] I.I. Piatetski-Shapiro and S. Rallis, *Rankin triple L -functions*, Compositio Math. **64** (1987), 31–115.
- [28] I. Satake, *Algebraic Structures of Symmetric Domains*, Publ. of the Math. Soc. of Japan, **14**, Iwanami Shoten and Princeton Univ. Press, 1980.
- [29] F. Sato and Y. Hironaka, *Local densities of representations of quadratic forms over p -adic integers (the non-dyadic case)*, J. Number Theory **83** (2000), 106–136.
- [30] G. Shimura, *Confluent hypergeometric functions on tube domains*, Math. Annalen **260** (1982), 269–302.
- [31] H. Stamm, *On the reduction of the Hilbert-Blumenthal-moduli scheme with $\Gamma_0(p)$ -level structure*, Forum Math. **9** (1997), 405–455.
- [32] C. Soulé, D. Abramovich, J.-F. Burnol, and J.Kramer, *Lectures on Arakelov Geometry*, Cambridge Stud. Adv. Math., vol 33, Cambridge U. Press, 1992.
- [33] W. J. Sweet, *The metaplectic case of the Weil–Siegel formula*, thesis, Univ. of Maryland, 1990.
- [34] T. Yang, *An explicit formula for local densities of quadratic forms*, J. Number Theory **72** (1998), 309–356.

Stephen S. KUDLA

Department of Mathematics

University of Maryland

College Park, MD 20742

USA

E-mail : ssk@math.umd.edu

POLYNOMIAL FUNCTORS OVER FINITE FIELDS
[after Franjou, Friedlander, Henn, Lannes, Schwartz, Suslin]

by **Teimuraz PIRASHVILI**

INTRODUCTION

Let K be a field and let \mathbf{Vect} be the category of vector spaces over K . Let \mathcal{V} be the full subcategory of \mathbf{Vect} consisting of finite dimensional vector spaces and let \mathcal{F} be the category of all covariant functors $\mathcal{V} \rightarrow \mathbf{Vect}$. A functor $F : \mathcal{V} \rightarrow \mathcal{V}$ is called a *polynomial functor* if the maps $\mathrm{Hom}(V, W) \rightarrow \mathrm{Hom}(FV, FW)$ given by $f \mapsto F(f)$ are polynomial for all V and W . If these maps are homogeneous of degree d , then F is called *homogeneous of degree d* . Let us recall that a map $g : X \rightarrow Y$ between finite dimensional vector spaces is called *polynomial* if after choosing bases it is given by m polynomials with n variables, where $n = \dim X$ and $m = \dim Y$. Over finite fields there is an essential difference between polynomials and maps obtained by polynomials. This yields another version of polynomial functors – strict polynomial functors, which was recently proposed by Friedlander and Suslin [14]. There is a functor $\mathcal{P} \rightarrow \mathcal{F}$ from the category of strict polynomial functors to the category of functors, which is no longer an embedding. The category \mathcal{F} is closely related to the category of unstable modules over the Steenrod algebra thanks to the remarkable result of Henn, Lannes and Schwartz [16], while the category \mathcal{P} is related to the theory of polynomial representations of the general linear group.

The aim of this talk is to describe some important properties of the categories \mathcal{F} and \mathcal{P} . We start with some general remarks on the category \mathcal{F} and then we describe relationship between the category \mathcal{F} and the category of unstable modules over the Steenrod algebra. In the section 2 we deal with the category \mathcal{P} and its relations with the category of polynomial representation of GL_n . In the next section we discuss the theorem of Franjou-Friedlander-Scorichenko-Suslin [13] on the comparison between $\mathrm{Ext}_{\mathcal{P}}$ and $\mathrm{Ext}_{\mathcal{F}}$. In the last section we discuss the Betley-Pirashvili conjecture [2] involving Waldhausen's stable K -theory. Recently this conjecture was proved for finite fields by Betley [1] and Suslin (see Appendix of [13]) independently.

1. THE CATEGORY \mathcal{F}

1.1. General Properties

In what follows \otimes_K and Hom_K are denoted by \otimes and Hom respectively. For any vector space V one denotes the dual vector space by V^\diamond .

The category \mathcal{F} of all functors from the category \mathcal{V} of finite dimensional vector spaces to the category of all vector spaces is an abelian category. For any $V \in \mathcal{V}$ one defines $P_V := K[\text{Hom}(V, -)]$. Here $K[S]$ denotes the free vector space generated by a set S . For any $F \in \mathcal{F}$ one has a natural isomorphism $\text{Hom}_{\mathcal{F}}(P_V, F) \cong F(V)$ thanks to the Yoneda lemma and therefore P_V is a projective object in \mathcal{F} .

For an object $F \in \mathcal{F}(K)$ one defines the *dual* DF of F by $V \mapsto (F(V^\diamond))^\diamond$. Let $J_W = DP_W$. Then $\text{Hom}_{\mathcal{F}}(F, J_V) \cong (DF)(V)$ for all $F \in \mathcal{F}$. Hence J_W is injective. Moreover P_V and J_V , $V \in \mathcal{V}$ are respectively projective and injective generators of the category \mathcal{F} . Clearly $P_V \otimes P_W \cong P_{V \oplus W}$ and $J_V \otimes J_W \cong J_{V \oplus W}$.

Let us recall the definition of the degree of functors due to Eilenberg and Mac Lane [10]. Let $F \in \mathcal{F}$ be a functor with $F(0) = 0$. The *second cross-effect* of F is given by

$$F(X | Y) := \ker(F(X \oplus Y) \longrightarrow F(X) \oplus F(Y)).$$

Clearly $F(X \oplus Y) \cong F(X) \oplus F(Y) \oplus F(X | Y)$. In order to define the third cross-effect of F , one can consider the second cross-effect $F(X | Y)$ as a functor on X and then take the second cross-effect of it. In this way one gets all higher cross-effects. One easily shows that the n -th cross-effect $F(X_1 | \dots | X_n)$ of F is the same as the kernel of the natural homomorphism:

$$F(X_1 \oplus \dots \oplus X_n) \rightarrow \bigoplus_{i=1}^n F(X_1 \oplus \dots \oplus \hat{X}_i \oplus \dots \oplus X_n).$$

A functor F is called of *degree* n if the $(n+1)$ -st cross-effect of F vanishes and the n -th cross-effect is nonzero. In this case we write $\text{deg}(F) = n$. We let \mathcal{F}_n be the category of all functors $F \in \mathcal{F}$ of degree $\leq n$. The examples of functors of degree n are functors T^n , S^n , Λ^n , Γ^n . Here $T^n(V)$ (resp. $S^n V$, $\Lambda^n V$, $\Gamma^n V$) is the the n -th tensor (resp. symmetric, exterior, divided) power of V . Clearly $DS^n \cong \Gamma^n$, while $D\Lambda^n \cong \Lambda^n$ and $DT^n \cong T^n$.

A functor F is called *analytic* if it is an union of subfunctors of finite degrees. We let \mathcal{F}_ω be the category of all analytic functors. The functors $T^*(V) = \bigoplus_{d \geq 0} T^d(V)$, $\Lambda^*(V)$, $\Gamma^*(V)$ are examples of analytic functors.

If K is a finite field, then any map between finite dimensional vector spaces is polynomial. A consequence of this phenomenon is the fact that over such fields the functor J_V is analytic for all $V \in \mathcal{V}$. Indeed, since $J_V \cong J_K^{\otimes(\dim V)}$ it is sufficient to consider only J_K . Clearly $J_K(V) \cong \text{Maps}(V^\diamond, K)$ and therefore it has a natural commutative algebra structure. Moreover if q is the number of elements in K , then the equality $a^q = a$ holds

in $J_K(V)$. Hence the natural embedding $V \subset \text{Maps}(V^\diamond, K)$ yields a homomorphism of algebras $S_q^*(V) \rightarrow J_K(V)$. It is not difficult to prove that this homomorphism is an isomorphism. Here $S_q^*(V)$ is the quotient of the symmetric algebra $S^*(V)$ by the ideal generated by $v^q - v, v \in V$. Since the quotient of an analytic functor is analytic, J_V is analytic too. Thus for finite fields the category \mathcal{F}_ω is an abelian category with sufficiently many injective objects.

1.2. The quotient category $\mathcal{F}_d/\mathcal{F}_{d-1}$

In this section we assume for simplicity that K is a prime field, although almost everything works for general rings (see [26]).

Let $F \in \mathcal{F}_n$. Since $F(X_1 | \cdots | X_n)$ is additive with respect to each variable we see that

$$F(X_1 | \cdots | X_n) \cong (cr_n F) \otimes (X_1 \otimes \cdots \otimes X_n),$$

where $cr_n F := F(K | \cdots | K) \subset F(\bigoplus_{i=1}^n K)$. The action of the symmetric group Σ_n on $\bigoplus_{i=1}^n K$ by permutation of summands yields an action of Σ_n on $cr_n F$. In this way one obtains an exact functor $cr_n : \mathcal{F}_n \rightarrow \Sigma_n\text{-mod}$. Clearly $cr_n F = 0$ if and only if $F \in \mathcal{F}_{n-1}$. For any $M \in \Sigma_n\text{-mod}$, one denotes by M^\flat and M^\sharp the functors given by $M^\flat(X) = (M \otimes X^{\otimes n})_{\Sigma_n}$ and $M^\sharp(X) := (M \otimes X^{\otimes n})^{\Sigma_n}$. Here for a Σ_n -module A , we let A_{Σ_n} and A^{Σ_n} be the coinvariants and invariants under the action of Σ_n . These functors are related by the *norm* homomorphism $N : M^\flat \rightarrow M^\sharp$, which is induced by the action of $\sum_{\sigma \in \Sigma_n} \sigma$.

PROPOSITION 1.1 ([26]). — *The functors $M \mapsto M^\flat$ and $M \mapsto M^\sharp$ define left and right adjoints to the functor $cr_n : \mathcal{F}_n \rightarrow \Sigma_n\text{-mod}$. Moreover $cr_n M^\flat \cong M \cong cr_n M^\sharp$. Thus the kernels and cokernels of the natural maps $(cr_n F)^\flat \rightarrow F$, $F \rightarrow (cr_n F)^\sharp$ are functors of degree $< n$. For any $F \in \mathcal{F}_n$ the composition $(cr_n F)^\flat \rightarrow F \rightarrow (cr_n F)^\sharp$ coincides with the norm homomorphism.*

Since the norm homomorphism is an isomorphism over the rationals it follows that in this case for any functor F of degree d one has an isomorphism $F \cong cr_d(F)^\flat \oplus F'$, where $\text{deg}(F') \leq d-1$. Using this fact it is not difficult to prove that any functor of finite degree has a unique decomposition $\bigoplus_d F_d$, where F_d is of the form M^\flat and M is a representation of Σ_d . Moreover any natural transformation between functors F_d and F_l for $d \neq l$ vanishes. Hence over the rationals one has an equivalence of categories $\mathcal{F}_n \cong \prod_{i=1}^n (\Sigma_i\text{-mod})$ (see also Appendix A of [24]).

For $\text{char}(K) > 0$ we only have an equivalence of categories $\mathcal{F}_n/\mathcal{F}_{n-1} \cong \Sigma_n\text{-mod}$.

If M is a simple representation of Σ_n then the image of the norm homomorphism $M^\flat \rightarrow M^\sharp$ is a simple object in \mathcal{F} and any simple object of \mathcal{F} is of this form. This is a formal consequence of Proposition 1.1 (compare with Proposition 4.7 of [21]). Recently Piriou and Schwartz [30] proved that for any simple functor $F \in \mathcal{F}$ one has $\text{Ext}_{\mathcal{F}}^1(F, F) = 0$, unlike to the situation with representations of symmetric groups.

1.3. Unstable modules over the Steenrod algebra and \mathcal{F}

In the 90's Henn, Lannes and Schwartz [16] discovered a remarkable link between the category $\mathcal{F}(\mathbf{F}_p)$ and representations of the Steenrod algebra. For simplicity we consider only the case $K = \mathbf{F}_2$.

We let \mathcal{A} denote the mod-2 Steenrod algebra. Let us recall that it is generated by the elements Sq^i of degree i , $i > 0$ and these generators satisfy the Adem relations (see [S1]). A (graded) module $M^* = \bigoplus_{i \geq 0} M^i$ over \mathcal{A} is said to be *unstable* if for every $x \in M$ one has $Sq^i x = 0$ if $i > |x|$. Here $|x|$ denotes the degree of x . Note that for any space X , the mod-2 cohomology $H^*(X)$ is an example of an unstable \mathcal{A} -module. Let \mathcal{U} be the category of unstable \mathcal{A} -modules. One observes that for any $n \geq 0$ there exists a unique (up to an isomorphism) unstable \mathcal{A} -module $F(n)$ such that for any $M^* \in \mathcal{U}$ one has a natural isomorphism

$$\mathrm{Hom}_{\mathcal{U}}(F(n), M^*) \cong M^n.$$

An unstable \mathcal{A} -module M^* is called *nilpotent* if for any $x \in M$ there exists $k > 0$ such that $(Sq_0)^k x = 0$. Here $Sq_0 : M^n \rightarrow M^{2n}$ is given by $m \mapsto Sq^n m$ and $(Sq_0)^k$ denotes the k -th iteration of Sq_0 . Let us notice that $H^*(X)$ is nilpotent as an unstable module if and only if $H^*(X)$ is nilpotent as a commutative algebra.

If V is a finite dimensional vector space, then by $H_* V$ and $H^* V$ we denote the homology and cohomology of V considered as an abelian group. Let us recall that one has the following isomorphisms of functors

$$H^*(V) \cong S^*(V^\diamond[1]), \quad H_*(V) \cong \Gamma^*(V[1]).$$

Here $V[1]$ indicates the fact that V is concentrated in degree 1. Since $H^* V$ is an unstable module, any $\theta \in \mathcal{A}$ yields a morphism $S^n \rightarrow S^{n+|\theta|}$ and therefore it induces by duality a natural transformation $\Gamma^{n+|\theta|} \rightarrow \Gamma^n$. Hence for any $F \in \mathcal{F}$ the graded vector space $\mathbf{m}(F) = \bigoplus_n \mathrm{Hom}_{\mathcal{F}}(\Gamma^n, F)$ has a natural \mathcal{A} -module structure and $\mathbf{m}(F) \in \mathcal{U}$. In this way one obtains a functor $\mathbf{m} : \mathcal{F} \rightarrow \mathcal{U}$. It follows from the very definition that

$$\mathbf{m}J_V = \bigoplus_n \mathrm{Hom}_{\mathcal{F}}(\Gamma^n, J_V) \cong \bigoplus_n (\Gamma^n V)^\diamond = S^* V.$$

The functor \mathbf{m} has a left adjoint $\mathbf{f} : \mathcal{U} \rightarrow \mathcal{F}$. So one has an isomorphism

$$\mathrm{Hom}_{\mathcal{F}}(\mathbf{f}M^*, F) \cong \mathrm{Hom}_{\mathcal{U}}(M^*, \mathbf{m}F).$$

Putting $F = J_V$ one obtains $(\mathbf{f}(M^*)(V))^\diamond \cong \mathrm{Hom}_{\mathcal{U}}(M^*, S^* V)$. One observes that if M^* is a finitely generated \mathcal{A} -module then $\mathrm{Hom}_{\mathcal{U}}(M^*, S^* V)$ is a finite dimensional vector space. Hence in this case $\mathbf{f}(M^*)(V)$ is isomorphic to $(\mathrm{Hom}_{\mathcal{U}}(M^*, S^* V))^\diamond$. In general M^* is a union of finitely generated \mathcal{A} -modules and therefore $\mathrm{Hom}_{\mathcal{U}}(M^*, S^* V)$ has a structure of a profinite vector space, whose continuous dual is denoted by $\mathrm{Hom}_{\mathcal{U}}(M^*, S^* V)'$. Since \mathbf{f} preserves colimits one has the following isomorphism

$$\mathbf{f}(M^*)(V) \cong \mathrm{Hom}_{\mathcal{U}}(M^*, S^* V)'$$

For example, $\mathbf{f}(F(n)) = \Gamma^n$.

THEOREM 1.2 ([16]). — *The functor \mathbf{f} is exact and preserves tensor products. For any $V \in \mathcal{V}$ one has a natural isomorphism $\mathbf{f}(S^*V) \cong J_V$. Furthermore the values of \mathbf{f} are analytic functors and for any analytic F the natural map $\mathbf{f}(\mathbf{m}(F)) \rightarrow F$ is an isomorphism. An object $M^* \in \mathcal{U}$ is nilpotent if and only if $\mathbf{f}(M^*) = 0$.*

Let \mathcal{N} be the category of nilpotent unstable modules. Obviously it is a Serre subcategory in \mathcal{U} . Let us note that the theorem implies the equivalence of categories $\mathcal{U}/\mathcal{N} \cong \mathcal{F}_\omega$.

The original proof of this theorem used deep structural results on injective objects in the category of unstable modules due to Carlsson, Miller, Lannes, Zarati and others. Kuhn [20] gave a reasonably simple proof of Theorem 1.2. His proof can be divided in several steps. The first one is the following important result.

THEOREM 1.3 ([20]). — *Let $F \in \mathcal{F}$ be a finite functor. Then one can embed F in a sum of the form $\bigoplus_{i=1}^k S^{n_i}$.*

Here a functor $T \in \mathcal{F}$ is called *finite* if it is of finite degree and has values in \mathcal{V} . The proof of Theorem 1.3 consists of several reductions. One observes that 2nd power map and multiplication in the symmetric algebra yield a natural transformation: $S^i \otimes S^j \rightarrow S^{i2^r+j}$, which is a monomorphism for $2^r > j$. Hence the class of functors for which Theorem 1.3 is true is closed with respect to tensor products. One observes also that F embeds in a finite sum of J_V , because $J_V, V \in \mathcal{V}$ are injective cogenerators. Since $J_V = J_K^{\otimes(\dim V)}$ and F has finite degree it suffices to consider the case when F is the image of S^m into $J_K, m \geq 1$ under the projection $S^*(V) \rightarrow S_2^*(V)$, because any sub-functor of finite degree of J_K is of such type. Using a combinatorial argument Kuhn was able to give an explicit embedding in this case (see [20]).

The second step in the proof of Theorem 1.2 is the following

PROPOSITION 1.4. — *The full subcategory of \mathcal{F} , whose objects are $\Gamma^n, n \geq 0$, and the full subcategory of \mathcal{U} , whose objects are $F(n), n \geq 0$ are isomorphic categories. The isomorphism takes Γ^n to $F(n)$.*

Sketch of the proof. — Let $a \in \mathcal{A}$ be a homogeneous element of degree m . It yields a natural transformation $a^* : S^n \rightarrow S^{n+m}$, thanks to the isomorphism $S^*(V[1]) \cong H^*(V^\diamond)$. It is not too difficult to show that in this way we get a homomorphism

$$\text{Hom}_{\mathcal{A}}(F(n+m), F(n)) \rightarrow \text{Hom}_{\mathcal{F}}(S^n, S^{n+m})$$

and counting argument shows that this map is an isomorphism. □

The next step is the following variant of the Gabriel-Popescu theorem on abelian categories. Let us recall that a *ringoid* is a small category with abelian group structure on Hom-sets, such that the composition is biadditive. Clearly a ringoid with one object is

the same ring considered as a category with one object. Moreover any small subcategory of an additive category is a ringoid. A *right module* over a ringoid \mathbf{C} is a contravariant functor $M : \mathbf{C} \rightarrow Ab$ with the property that the maps $\text{Hom}_{\mathbf{C}}(c, d) \rightarrow \text{Hom}(M(d), M(c))$ given by $f \mapsto M(f)$ are homomorphisms of abelian groups, $c, d \in \mathbf{C}$. For a ringoid \mathbf{C} we let $\text{mod-}\mathbf{C}$ be the category of all right \mathbf{C} -modules.

THEOREM 1.5 ([20]). — *Let \mathbf{A} be an abelian category with exact directed colimits and let $\mathbf{C} \subset \mathbf{A}$ be a small subcategory, such that any object in \mathbf{A} is a quotient of an object of the form $\bigoplus c_i$, where $c_i \in \mathbf{C}$. Let $r : \mathbf{A} \rightarrow \text{mod-}\mathbf{C}$ be the functor given by $r(a)(c) = \text{Hom}_{\mathbf{A}}(c, a)$. Then*

- i) *The functor r has a left adjoint $l : \text{mod-}\mathbf{C} \rightarrow \mathbf{A}$.*
- ii) *The functor l is exact, r is fully faithful and $lra \rightarrow a$ is an isomorphism for any $a \in \mathbf{A}$. Hence in this case $\text{mod-}\mathbf{C}/\text{Ker}(l) \cong \mathbf{A}$.*
- iii) *\mathbf{A} has enough injectives and r preserves injectives. Moreover if I and J are injective in \mathbf{A} then r yields an isomorphism*

$$\text{Hom}_{\mathbf{A}}(I, J) \cong \text{Hom}_{\text{mod-}\mathbf{C}}(rI, rJ).$$

If additionally all objects from \mathbf{C} are small and projective in \mathbf{A} then r (and hence l) is an equivalence of categories.

Now we are in the position to give a sketch of the proof of Theorem 1.2. First one takes $\mathbf{A} = \mathcal{U}$ and $\mathbf{C} = \{F(n)\}_{n \geq 0}$. Here $\{F(n)\}_{n \geq 0}$ is the full subcategory of \mathcal{U} whose objects are $F(n)$, $n \geq 0$. The last statement of Theorem 1.5 shows that $\mathcal{U} \cong \text{mod-}\{F(n)\}_{n \geq 0}$. Then we put $\mathbf{A} = \mathcal{F}_{\omega}$ and take $\mathbf{C} = \{\Gamma^n\}_{n \geq 0}$, where $\{\Gamma^n\}_{n \geq 0}$ is the full subcategory of \mathcal{F}_{ω} whose objects are Γ^n , $n \geq 0$. By Proposition 1.4 one has an equivalence of categories $\text{mod-}\mathbf{C} \cong \mathcal{U}$. It follows from Theorem 1.3 that the condition of Theorem 1.5 holds. Applying Theorem 1.5 we obtain an adjoint pair of functors (l, r) between the category \mathcal{U} and \mathcal{F}_{ω} . One checks readily that $r = \mathbf{m}$ and we get all statements of Theorem 1.2 except the statement about the tensor product and the characterization of nilpotent modules. We refer to [20] and [34] for the proof of these facts.

REMARK 1.6. — By Proposition 1.4 one can reconstruct the category \mathcal{U} and the Steenrod algebra (without the Bokstein operations) from the endomorphism ring $\text{End}_{\mathcal{F}}(S^*, S^*)$, which has a meaning in a far more general setting. Based on this observation an appropriate notion of “Steenrod algebra over a finite field K ” was developed in [20] in such a way that the whole material in 1.3 works for any finite field.

REMARK 1.7. — Among other applications of Theorem 1.2 we would like to mention the important article of Henn-Lannes-Schwartz [17] and also the work of Kuhn [23] and Schwartz [35] on topological realizability of unstable modules over the Steenrod algebra.

1.4. Artinian conjecture

The following conjecture was posed by L. Schwartz.

If K is a finite field then the category \mathcal{F} is locally noetherian.

Here is another formulation:

If K is a finite field and $V \in \mathcal{V}$ then the injective object J_V is artinian.

It is known that the conjecture is true when $\dim V = 1$. As was proved by Powell [31] the conjecture is also true when $K = \mathbf{F}_2$ and $\dim V = 2$.

The conjecture implies the following assertions

Any injective object in \mathcal{F}_ω is also injective in \mathcal{F} and the functor $\mathbf{f} : \mathcal{U} \rightarrow \mathcal{F}$ respects injective objects.

2. THE CATEGORY \mathcal{P}

In this section K is a field. We give several equivalent interpretations of the category of *strict polynomial functors*. We start with recalling the definition of Schur algebras.

2.1. Polynomial representations of GL_n and Schur algebras.

Let $GL_n = GL_{n,K}$ be the general linear group over K considered as an algebraic group. We recall that the coordinate ring $K[GL_n]$ is the quotient of the polynomial algebra $K[x_{ij}, y]_{1 \leq i, j \leq n}$ by the ideal generated by $(y \det(x_{ij}) - 1)$. It has a Hopf algebra structure, whose comultiplication is given by $\Delta(x_{ij}) = \sum_{k=1}^n x_{ik} \otimes x_{kj}$. Moreover we let M_n be the algebraic monoid of $n \times n$ -matrices. The coordinate ring $K[M_n]$ is the polynomial algebra $K[x_{ij}]_{1 \leq i, j \leq n}$. Clearly it is a subalgebra of $K[GL_n]$. Comodules over $K[GL_n]$ are called *rational representations* of GL_n . A rational representation of GL_n is called *polynomial* if it is also a comodule over $K[M_n]$. One observes that for each $d \geq 0$ the subspace $A(n, d) \subset k[M_n]$ of homogeneous polynomials of degree d is a subcoalgebra of $k[M_n]$. The linear dual $A(n, d)^\diamond$ of $A(n, d)$ is known as *Schur algebra* and it is denoted by $S(n, d)$ (see [15]). A polynomial representation of GL_n is said to be *homogeneous of degree d* if it is a comodule over $A(n, d)$. Clearly the category of finite dimensional homogeneous polynomial representations of degree d is equivalent to the category of finite dimensional modules over the Schur algebra $S(n, d)$. We refer the reader to [25] and references there for extensive information on Schur algebras.

2.2. Polynomial laws and the category \mathcal{P}

For a vector space V we let \underline{V} be the functor from the category $-K/alg$ of commutative K -algebras to the category of sets given by $R \mapsto R \otimes V$. A *polynomial law* from a vector space V to a vector space W is by definition a natural transformation of functors $\underline{V} \rightarrow \underline{W}$

(see [32]). Let Pol be the category of vector spaces and polynomial laws. A polynomial law $f \in \text{Hom}_{Pol}(V, W)$ is called *homogeneous of degree d* if for any algebra R , $r \in R$ and $x \in R \otimes V$ one has $f_R(rx) = r^d f_R(x)$.

Following Friedlander and Suslin [14] a *strict polynomial functor* $F : \mathcal{V} \rightarrow \mathcal{V}$ is a rule which associates to each finite dimensional vector space V a vector space $F(V)$ and to each pair (V, W) of finite dimensional vector spaces a polynomial law

$$F(V, W) \in \text{Hom}_{Pol}(\text{Hom}(V, W), \text{Hom}(F(V), F(W)))$$

in such a way that for any $V \in \mathcal{V}$ one has $F(V, V)(1_V) = 1_{F(V, V)}$ and for any $U, V, W \in \mathcal{V}$ the following diagram of polynomial laws commutes:

$$\begin{array}{ccc} \text{Hom}(V, W) \times \text{Hom}(U, V) & \rightarrow & \text{Hom}(U, W) \\ \downarrow & & \downarrow \\ \text{Hom}(FV, FW) \times \text{Hom}(FU, FV) & \rightarrow & \text{Hom}(FU, FW), \end{array}$$

where the vertical arrows are induced by F , while the horizontal arrows come from the composition.

One says that a strict polynomial functor F is of finite degree if the degrees of the polynomial laws $F(X, Y)$ are bounded above. A strict polynomial functor F is called *homogeneous of degree n* if for any $V, W \in \mathcal{V}$ the polynomial $F(V, W)$ is homogeneous of degree n . We let \mathcal{P} (resp. \mathcal{P}_d) be the category of strict polynomial functors of finite degree (resp. strict homogeneous polynomial functors of degree d). Since any polynomial law is a sum of homogeneous ones, it follows that $\mathcal{P} \cong \bigoplus_d \mathcal{P}_d$.

2.3. The category $\Gamma^d \mathcal{V}$

The vector space $S^d V$ is the module of coinvariants of $V^{\otimes d}$ under the action of the symmetric group Σ_d on d symbols. Similarly $\Gamma^d V$ is the module of invariants $\Gamma^d(V) = (V^{\otimes d})^{\Sigma_d}$. For any vector space V one has a map $\gamma^d : V \rightarrow \Gamma^d(V)$, which is given by $x \mapsto \gamma^d(x) := x^{\otimes n} \in (V^{\otimes n})^{\Sigma_n}$. Moreover for any vector spaces V and W there exists a unique linear map

$$\mu : \Gamma^d V \otimes \Gamma^d W \rightarrow \Gamma^d(V \otimes W)$$

with the property $\mu(\gamma^d(x) \otimes \gamma^d(y)) = \gamma^d(x \otimes y)$. One can use the transformation μ to define the composition in the category $\Gamma^d \mathcal{V}$, whose objects are those of \mathcal{V} while morphisms are $\text{Hom}_{\Gamma^d \mathcal{V}}(V, W) = \Gamma^d(\text{Hom}(V, W))$. The identity arrows in $\Gamma^d \mathcal{V}$ are $\gamma^d(1_V)$. Clearly $\Gamma^d \mathcal{V}$ is a K -linear category, that is a category whose set of morphisms between two objects has a vector space structure and the composition is bilinear.

PROPOSITION 2.1. — *The following categories are equivalent:*

- the category of K -linear functors $\Gamma^d \mathcal{V} \rightarrow \mathcal{V}$,
- the category of homogeneous strict polynomial functors of degree d ,

- the category of homogeneous polynomial representations of degree d of GL_n , provided $n \geq d$,
- the category of finite dimensional modules over $S(n, d)$, provided $n \geq d$.

Proof. — Clearly \underline{V} is isomorphic to the functor $\text{Hom}_{K/\text{alg}}(S^*(V^\diamond), -)$ as soon as $V \in \mathcal{V}$. Hence by the Yoneda lemma one has $\text{Hom}_{\mathcal{P}_{ol}}(V, W) \cong S^*(V^\diamond) \otimes W \cong \bigoplus_d \text{Hom}(\Gamma^d V, W)$. Elements of $\text{Hom}(\Gamma^d V, W)$ correspond to homogeneous polynomial laws of degree d . This shows that a homogeneous strict polynomial functor of degree d is the same as a collection of $F(V) \in \mathcal{V}, V \in \mathcal{V}$ together with linear maps $\Gamma^d \text{Hom}(V, W) \otimes F(V) \rightarrow F(W)$ satisfying associativity and unity conditions. Therefore the first two categories are equivalent. The last two categories are also equivalent by the discussion in 2.1. It remains to show the equivalence between the first and the third category. For each $m \geq 0$ we let $\Gamma^{d,m} \in \mathcal{P}_d$ be the functor given by $V \mapsto \Gamma^d(\text{Hom}(K^m, V))$. By the Yoneda lemma one has an isomorphism $\text{Hom}_{\mathcal{P}}(\Gamma^{d,m}, F) \cong F(K^m)$ for any $F \in \mathcal{P}_d$. So $\Gamma^{d,m}, m \geq 0$ are small projective generators in the abelian category \mathcal{P}_d . By looking at cross-effects of $\Gamma^{d,m}$ it is clear that each $\Gamma^{d,m}$ is a direct summand of $\Gamma^{d,n}$ provided $m \leq n$ and each $\Gamma^{d,m}$ is a direct summand of an object of the form $\bigoplus_{k=1}^d \Gamma^{d,n}$ provided $n \geq d$. Therefore $\Gamma^{d,n}$ is a projective generator. Thus the category of K -linear functors $\Gamma^d \mathcal{V} \rightarrow \mathcal{V}$ is equivalent to the category of finite dimensional modules over $\text{End}_{\mathcal{P}}(\Gamma^{d,n}) = \Gamma^d(\text{End}(K^n))$. But the last algebra is isomorphic to the Schur algebra $S(n, d)$ and hence the result. \square

2.4. Elementary properties of \mathcal{P}

Clearly $f \mapsto \gamma^d(f)$ defines the (nonlinear) functor $\gamma^d : \mathcal{V} \rightarrow \Gamma^d \mathcal{V}$. The precomposition with γ^d yields the functor $\mathcal{P}_d \rightarrow \mathcal{F}_d$, which is a full embedding provided the field K contains at least d elements. By abuse of notation we denote the image of F under this functor by the same letter F .

The natural transformation $\Gamma^{d+l} \rightarrow \Gamma^d \otimes \Gamma^l$ coming from the Hopf algebra structure on the divided power algebra can be used to define the tensor product of strict homogeneous polynomial functors. It yields the functor $\mathcal{P}_d \times \mathcal{P}_l \rightarrow \mathcal{P}_{k+l}$, which corresponds to the obvious tensor functor in \mathcal{F} .

The dual of the natural transformation $S^d(S^l V) \rightarrow S^{dl}(V)$ is a transformation $\Gamma^{dl} \rightarrow \Gamma^d \circ \Gamma^l$, which can be used to define the composition of strict homogeneous polynomial functors. One observes also that the dual of a strict homogeneous polynomial functor is a strict homogeneous polynomial functor. Since Γ^d carries obvious structure of a strict homogeneous polynomial functor of degree d , we see that $T^d, S^d, \Lambda^d \in \mathcal{P}_d$. As was observed in the proof of Proposition 2.1 the functors $\Gamma^{d,m}$ are projective generators of \mathcal{P}_d . Therefore $S^{d,m} := D\Gamma^{d,m}$ are injective cogenerators of $\mathcal{P}_d, m \geq 0$. Another system of projective generators is $\Gamma^{d_1} \otimes \cdots \otimes \Gamma^{d_k}, d_1 + \cdots + d_k = d$.

2.5. Frobenius twist

In representation theory of algebraic groups over fields of characteristic $p > 0$ the Frobenius twist plays an important rôle (see [18]). By Proposition 2.1 to the Frobenius twist on polynomial representations corresponds a similar operation on strict polynomial functors, which can be described as follows. Let K be a field of characteristic $p > 0$. For a vector space V we let $V^{(1)}$ be the vector space obtained by extending scalars via the Frobenius homomorphism $f : K \rightarrow K$ given by $f(\lambda) = \lambda^p$. In other words $V^{(1)}$ is a vector space generated by $v^{(1)}$, $v \in V$ modulo the following relations

$$(v + w)^{(1)} = v^{(1)} + w^{(1)}, \quad (\lambda v)^{(1)} = \lambda^p v^{(1)}, \quad \lambda \in K.$$

One defines $V^{(r)}$, $r \geq 1$ by induction: $V^{(r+1)} = (V^{(r)})^{(1)}$.

The map $V^{(1)} \rightarrow S^p V$ given by $v^{(1)} \mapsto v^p$ is linear. By duality one obtains the natural transformation $\Gamma^p \rightarrow I^{(1)}$, where $I^{(1)}(V) = V^{(1)}$. Using this transformation one readily checks that $I^{(1)} \in \mathcal{P}_p$. For any $F \in \mathcal{P}_d$ we put $F^{(1)} = F \circ I^{(1)}$ and then by induction one defines $F^{(r+1)} = (F^{(r)})^{(1)}$, $r \geq 0$. Clearly $F^{(r)} \in \mathcal{P}_{dpr}$ if $F \in \mathcal{P}_d$.

Let us note that for finite prime fields the functors $I^{(1)}$ and I are isomorphic in \mathcal{F} . Here $I(V) = V$. However in \mathcal{P} these functors are different, they even have different degrees: $I \in \mathcal{P}_1$, while $I^{(1)} \in \mathcal{P}_p$. For finite fields the Frobenius twist induces an equivalence of categories $(-)^{(1)} : \mathcal{F}_d \rightarrow \mathcal{F}_d$.

2.6. $\text{Ext}_{\mathcal{P}}$ and cohomology of algebraic groups

The following theorem was proved by Donkin (see Section 3 of [14]).

THEOREM 2.2 ([6, 7]). — *Let A and B be finite dimensional polynomial representations. Then*

$$\text{Ext}_{K[M_n]}^*(A, B) \rightarrow \text{Ext}_{K[GL_n]}^*(A, B)$$

is an isomorphism. Here Ext is taken in the category of comodules.

By Proposition 2.1 we also have

COROLLARY 2.3. — *Let $F, T \in \mathcal{P}_d$. Then*

$$\text{Ext}_{\mathcal{P}}^*(F, T) \rightarrow \text{Ext}_{K[GL_n]}^*(F(K^n), T(K^n))$$

is an isomorphism provided $n \geq d$.

3. FUNCTOR COHOMOLOGY

In this section we deal with Ext -groups in the categories \mathcal{F} and \mathcal{P} . Calculations in \mathcal{P} are easier, because each piece \mathcal{P}_d of \mathcal{P} is of finite global dimension thanks to a result of Donkin [6], who proved this result in terms of Schur algebras (see also Totaro [36] for an

explicit formula for the global dimension of \mathcal{P}_d). Calculation in \mathcal{P} is easier also because the functors $S^{n_1} \otimes \cdots \otimes S^{n_k}$ (resp. $\Gamma^{n_1} \otimes \cdots \otimes \Gamma^{n_k}$) are injective (resp. projective) in \mathcal{P} . Of course S^n is no longer injective in \mathcal{F} , but the following result shows that if one inverts the Frobenius $S^n \rightarrow S^{np}$ then one gets an injective object in \mathcal{F}_ω .

LEMMA 3.1 ([22]). — Assume $F \in \mathcal{F}$ has a projective resolution of finite type. Then, for any $k > 0$, one has

$$\operatorname{colim} \operatorname{Ext}_{\mathcal{F}}^k(T, S^{p^n}) = 0,$$

where S^n denotes the n -th symmetric power and the limit is considered with respect to the Frobenius maps $\Phi : S^{p^n} \rightarrow S^{p^{n+1}}$.

Proof. — (see [12]). Let P_* be a projective resolution of F which is of finite type. Then DP_* is an injective resolution of DF . Hence

$$\operatorname{Ext}_{\mathcal{F}}^k(F, S^{2^n}) = H^k(\operatorname{Hom}_{\mathcal{F}}(P_*, S^{2^n})) \cong H^k((\mathbf{m}(DP_*))^{2^n}) \cong (H^k(\mathbf{m}(DP_*)))^{2^n}$$

and it is enough to show that $H^k(\mathbf{m}(DP_*)) \in \mathcal{N}$. Since

$$\mathbf{f}(H^k(\mathbf{m}(DP_*))) = H^k(\mathbf{f}(\mathbf{m}(DP_*))) = H^k(DP_*) = 0, \quad k > 0$$

Theorem 1.2 implies the expected inclusion. □

Lemma 3.1 together with Lemma 3.3 plays a crucial rôle in this section.

Let \mathbf{C} and \mathbf{D} be categories and let $l : \mathbf{C} \rightarrow \mathbf{D}$ and $r : \mathbf{D} \rightarrow \mathbf{C}$ be functors. If l is a left adjoint to r we will say that (l, r) is an adjoint pair from \mathbf{C} to \mathbf{D} . Moreover for a small category \mathbf{C} we let $\mathbf{C}\text{-mod}$ be the category of all functors $\mathbf{C} \rightarrow \mathbf{Vect}$.

LEMMA 3.2. — Let (l, r) be an adjoint pair from \mathbf{C} to \mathbf{D} . Assume \mathbf{C} and \mathbf{D} are small categories. Then for any $F : \mathbf{C} \rightarrow \mathbf{Vect}$ and $G : \mathbf{D} \rightarrow \mathbf{Vect}$ one has an isomorphism

$$\operatorname{Ext}_{\mathbf{D}\text{-mod}}^*(F \circ r, G) \cong \operatorname{Ext}_{\mathbf{C}\text{-mod}}^*(F, G \circ l).$$

LEMMA 3.3 ([27]). — Let $A \in \mathcal{F}$ be an additive functor and let $B : \mathcal{V} \times \mathcal{V} \rightarrow \mathbf{Vect}$ be a bifunctor with the property $B(X, 0) = 0 = B(0, X)$ for all $X \in \mathcal{V}$. Then $\operatorname{Ext}_{\mathcal{F}}^*(A, B^d) = 0 = \operatorname{Ext}_{\mathcal{F}}^*(B^d, A)$, where $B^d(V) = B(V, V)$.

Proof. — Apply the previous lemma to the adjoint pairs $(i_k, p_k) : \mathcal{V} \rightarrow \mathcal{V} \times \mathcal{V}$, $k = 1, 2, 3$, where $i_1(V) = (V, 0)$, $i_2(V) = (0, V)$, $i_3(V) = (V, V)$, $p_k(V_1, V_2) = V_k$, $k = 1, 2$, $p_3(V_1, V_2) = V_1 \oplus V_2$ and use the fact that $A \circ p_3 = A \circ p_1 \oplus A \circ p_2$. □

A strict polynomial functor T is called additive if T is additive as an object in \mathcal{F} . The previous argument shows that Lemma 3.3 is still true in the framework of \mathcal{P} . Actually the proof of Lemma 3.3 gives a bit more:

$$\operatorname{Ext}_{\mathcal{E}}^*(K, T_1 \otimes \cdots \otimes T_n) = 0 = \operatorname{Ext}_{\mathcal{E}}^*(T_1 \otimes \cdots \otimes T_n, K),$$

where $\operatorname{deg} K < n$, $T_1(0) = \cdots = T_n(0) = 0$ and \mathcal{E} is \mathcal{P} or \mathcal{F} .

Let us see how to use these facts for explicit calculations. First we consider calculations in the category \mathcal{P} . Since $I \in \mathcal{P}$ is projective we have $\text{Ext}_{\mathcal{P}}^i(I, -) = 0$ if $i > 0$. More interesting is $\text{Ext}_{\mathcal{P}}^*(I^{(r)}, -)$ for $r \geq 1$.

For simplicity we restrict ourselves to the case $p = 2$ and refer the reader to [12], [14] and [13] for the case $p > 2$. We follow ideas of [12].

It is not too hard to show that one has an exact sequence

$$(3.1) \quad 0 \rightarrow S^{2^h(1)} \rightarrow S^{2^h} \rightarrow \dots \rightarrow S^{2^h-i} \otimes S^i \rightarrow \dots \rightarrow S^{2^h} \rightarrow 0.$$

Here the first nontrivial map is given by the Frobenius, and the rest are the compositions $S^k \otimes S^l \rightarrow S^{k-1} \otimes S^1 \otimes S^l \rightarrow S^{k-1} \otimes S^{l+1}$, where first map (resp. second) is induced by the comultiplication (resp. multiplication) in the Hopf algebra $S^*(V)$. By the \mathcal{P} -version of Lemma 3.3 $\text{Ext}_{\mathcal{P}}^*(I^{(r)}, S^k \otimes S^l) = 0$ if $k, l \geq 1$ and a fortiori $\text{Hom}_{\mathcal{P}}^*(I^{(r)}, S^k \otimes S^l) = 0$. Let us note that $\text{Hom}_{\mathcal{P}}(I^{(h)}, S^{2^h}) = \text{Hom}_{\mathcal{F}}(I^{(h)}, S^{2^h})$ is one dimensional spanned by the iterated Frobenius map. Hence it follows from the injective resolution (3.1) that $\text{Ext}_{\mathcal{P}}^i(I^{(h+1)}, S^{2^h(1)}) = K$, if $i = 0$ and $i = 2^{h+1}$ and $\text{Ext}_{\mathcal{P}}^i(I^{(h+1)}, S^{2^h(1)}) = 0$ otherwise. Let us apply the degreewise action by the Frobenius twist on (3.1). Then one obtains a non-injective resolution of $S^{2^h(2)}$, which gives rise to a hypercohomology spectral sequence

$$E_{st}^1 = \text{Ext}_{\mathcal{P}}^t(I^{(h+2)}, S^{2^{h+1}-s(1)} \otimes S^{s(1)}) \implies \text{Ext}_{\mathcal{P}}^{s+t}(I^{(h+2)}, S^{2^h(2)}).$$

By the \mathcal{P} -version of Lemma 3.3 the spectral sequence has only two nontrivial columns corresponding to $s = 0$ and $s = 2^{h+1}$. Moreover the previous calculation shows that in both columns there are only two nontrivial terms corresponding to $t = 0$ and $t = 2^{h+2}$, so there is no space for differentials and we get $\text{Ext}_{\mathcal{P}}^i(I^{(h+2)}, S^{2^h(2)}) = K$, if $i = 0, i = 2^{h+1}, 2^{h+2}, 2^{h+2} + 2^{h+1}$ and $\text{Ext}_{\mathcal{P}}^i(I^{(h+2)}, S^{2^h(2)}) = 0$ otherwise. By iteration one obtains:

PROPOSITION 3.4 ([14]). — *For any $0 \leq j \leq r$ one has $\text{Ext}_{\mathcal{P}}^q(I^{(r)}, S^{2^{r-j}(j)}) = K$ if $q \equiv 0 \pmod{2^{r-j+1}}$ and $q < 2^{r+1}$ and $\text{Ext}_{\mathcal{P}}^q(I^{(r)}, S^{2^{r-j}(j)}) = 0$ otherwise.*

Such type of results was used in [14] to prove the following theorem:

The rational cohomology of any finite group scheme is a finitely generated algebra.

Now we do similar calculations in \mathcal{F} .

THEOREM 3.5 ([12]). — $\text{Ext}_{\mathcal{F}}^i(I^{(h)}, S^{2^h}) = K$ if $i \equiv \pmod{2^{h+1}}$ and $= 0$ otherwise.

Proof. — The exact sequence (3.1) together with Lemma 3.3 yields the exact sequence

$$\begin{aligned} & \dots \rightarrow \text{Ext}_{\mathcal{F}}^{k-2^h}(I^{(h)}, S^{2^h}) \rightarrow \\ & \rightarrow \text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^{h-1}}) \rightarrow \text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^h}) \rightarrow \text{Ext}_{\mathcal{F}}^{k-2^h+1}(I^{(h)}, S^{2^h}) \rightarrow \dots \end{aligned}$$

From this sequence it follows that $\text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^{h-1}}) \rightarrow \text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^h})$ is an isomorphism for $k < 2^h$. It is known that any finite functor admits a projective resolution of finite type (see [12]). Thus by Lemma 3.1 one has $\text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^h}) = 0$ as soon as $0 < k < 2^{h+1}$.

Now one can use this information in the same exact sequence to get $\text{Ext}_{\mathcal{F}}^k(I^{(h)}, S^{2^h}) = 0$ for $2^{h+1} < k < 2^{h+2}$. By iterating this process one gets the result. \square

It is worth to mention that Franjou, Friedlander, Scorichenko and Suslin give in [13] explicit calculations of $\text{Ext}^*(\Gamma^{d(h)}, S^n)$, $\text{Ext}^*(\Gamma^{d(h)}, \Lambda^n)$, $\text{Ext}^*(\Gamma^{d(h)}, \Gamma^n)$ and $\text{Ext}^*(\Lambda^{d(h)}, \Lambda^n)$ in the \mathcal{F} and \mathcal{P} framework. Let us also note that the groups $\text{Ext}_{\mathcal{F}}^*(I, -)$ are known as Mac Lane cohomology (see [19]) and they are dual to the topological Hochschild homology (see [29]). The previous theorem for $h = 0$ was first obtained by Breen [4] in the framework of “Extensions du groupe additif” and Bökstedt [3] in the framework of topological Hochschild homology.

Returning to Proposition 3.4 one sees that the Frobenius twist induces monomorphisms $\text{Ext}_{\mathcal{P}}^i(I^{(s+j)}, S^{2^s(j)}) \rightarrow \text{Ext}_{\mathcal{P}}^i(I^{(s+j+1)}, S^{2^s(j+1)})$ which stabilize for each i . This phenomenon is a particular case of the well known fact on cohomology of algebraic groups (see [18], page 347), which by virtue of Corollary 2.3 yields the following

THEOREM 3.6. — *Let $F, T \in \mathcal{P}_d$. Then the Frobenius twist gives a monomorphism $\text{Ext}_{\mathcal{P}}^i(F, T) \rightarrow \text{Ext}_{\mathcal{P}}^i(F^{(1)}, T^{(1)})$. Moreover, for all sufficiently large m the map $\text{Ext}_{\mathcal{P}}^i(F^{(m)}, T^{(m)}) \rightarrow \text{Ext}_{\mathcal{P}}^i(F^{(m+1)}, T^{(m+1)})$ is an isomorphism.*

By [13] the last assertion holds as soon as $i \leq 2^{m+1} - 1$.

We let $\text{Ext}_{\text{Gen}}^i(F, T)$ be the common value of $\text{Ext}_{\mathcal{P}}^i(F^{(m)}, T^{(m)})$, $m \gg 0$. Since the Frobenius twist in \mathcal{F}_d is an equivalence of categories we see that the canonical map $\text{Ext}_{\mathcal{P}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, T)$ gives rise to the homomorphism $\text{Ext}_{\text{Gen}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, T)$. Now comparing the Proposition 3.4 and Theorem 3.5 one sees that the homomorphism $\text{Ext}_{\text{Gen}}^*(I^{(h)}, S^{2^h}) \rightarrow \text{Ext}_{\mathcal{F}}^*(I^{(h)}, S^{2^h})$ is an isomorphism. However in general $\text{Ext}_{\text{Gen}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, T)$ is not an isomorphism even when $*$ = 0. The following rather surprising result of Franjou, Friedlander, Scorichenko and Suslin gives a condition when it is an isomorphism.

THEOREM 3.7 ([13]). — *Let $T, F \in \mathcal{P}_d$ be strict polynomial functors, then*

$$\text{Ext}_{\mathcal{P}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, T)$$

is a monomorphism. Moreover if the field K contains at least d elements then

$$\text{Ext}_{\text{Gen}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, T)$$

is an isomorphism.

The proof of the theorem is quite long and can be divided essentially into two parts. In the first part they prove a weaker version (Proposition 3.8) and then they use a base change argument in a very clever way to finish the proof. Here we prove only Proposition 3.8 and refer the interested reader to the original paper for the second part of the proof of Theorem 3.7.

PROPOSITION 3.8. — Let $F, T \in \mathcal{P}_d$ and let $m \geq 0$. Assume K has at least $2^m d$ elements. Then

$$\mathrm{Ext}_{\mathcal{P}}^i(F^{(m)}, T^{(m)}) \rightarrow \mathrm{Ext}_{\mathcal{F}}^i(F, T)$$

is an isomorphism provided $i \leq 2^{m-d+1} - 2$.

The rest of this section is devoted to the proof of Proposition 3.8.

DEFINITION 3.9 ([11]). — A graded object $E^* = \bigoplus_n E^n \in \mathcal{E}$, where $\mathcal{E} = \mathcal{F}$ or \mathcal{P} is called exponential if there exist natural isomorphisms $E^0(V) = K$, $E^*(V \oplus W) \cong E^*(V) \otimes E^*(W)$ and the values of E^n lie in \mathcal{V} .

If E^* is an exponential functor, then $E^n(0) = 0$, for $n > 0$ and $\mathrm{deg}(E^n) \leq n$. Typical examples of exponential functors are given by the symmetric algebra, the exterior algebra and the divided power algebra. If E^* and \bar{E}^* are exponential functors then $E^* \otimes \bar{E}^*$ is also an exponential functor.

The same argument as in the proof of Lemma 3.3 gives also the following

LEMMA 3.10 ([11, 13]). — Let E^* be an exponential functor in \mathcal{E} . Then for any functors $B, C \in \mathcal{E}$ and any $i \geq 0$ one has isomorphisms of graded vector spaces

$$\begin{aligned} \mathrm{Ext}_{\mathcal{E}}^*(E^n, B \otimes C) &\cong \bigoplus_{i+j=n} \mathrm{Ext}_{\mathcal{E}}^*(E^i, B) \otimes \mathrm{Ext}_{\mathcal{E}}^*(E^j, C) \\ \mathrm{Ext}_{\mathcal{E}}^*(B \otimes C, E^n) &\cong \bigoplus_{i+j=n} \mathrm{Ext}_{\mathcal{E}}^*(B, E^i) \otimes \mathrm{Ext}_{\mathcal{E}}^*(C, E^j). \end{aligned}$$

LEMMA 3.11 ([11, 13]). — For any exponential functor E^* and any injective $J \in \mathcal{P}_n$, $n \equiv 0 \pmod{2^h}$ one has $\mathrm{Ext}_{\mathcal{F}}^i(E^d, J) = 0$ provided $0 < i \leq 2^{n-d+2} - 2$.

Proof. — First we exploit the fact that $S^n \in \mathcal{F}$ behaves like an injective with respect to functors of degree $\ll n$. Assume $n = \epsilon_0 + 2\epsilon_1 + \dots + 2^k \epsilon_k$, $\epsilon_i = 0, 1$ is the 2-adic expansion of a natural number n and let $\alpha(n)$ be the number of nonzero elements of the set $\{\epsilon_0, \dots, \epsilon_k\}$. Then S^n is a retract of $(S^1)^{\otimes \epsilon_0} \otimes \dots \otimes (S^{2^k})^{\otimes \epsilon_k}$. Hence a discussion after Lemma 3.3 shows that $\mathrm{Ext}_{\mathcal{F}}^*(K, S^n \otimes S^m) = 0$ as soon as $\mathrm{deg}K < \alpha(n) + \alpha(m)$. Based on this fact and using the exact sequence (3.1) it is not too difficult to prove that the Frobenius transformation yields an isomorphism $\mathrm{Ext}_{\mathcal{F}}^i(K, S^{2^n}) \rightarrow \mathrm{Ext}_{\mathcal{F}}^i(K, S^{2^{n+1}})$ provided $\mathrm{deg}K \leq d$ and $i \leq 2^{n-d+2} - 2$. Since any finite functor has a projective resolution of finite type (see [12]), it follows from Lemma 3.1 that for such K one has $\mathrm{Ext}_{\mathcal{F}}^i(K, S^{2^n}) = 0$ provided $\mathrm{deg}K \leq d$ and $0 < i \leq 2^{n-d+2} - 2$. Lemma 3.10 together with the fact that the functors $S^{n_1} \otimes \dots \otimes S^{n_k}$, $n_1 + \dots + n_k = n$ are injective generators in \mathcal{P}_n can be used to finish the proof. \square

Proof of Proposition 3.8. — It consists of several reductions. Using a hypercohomology spectral sequence one can prove that it suffices to restrict ourselves to the case when F is a projective object. So we assume that $F = \Gamma^{d_1} \otimes \dots \otimes \Gamma^{d_k}$, $d_1 + \dots + d_k = d$. Let

$T^{(m)} \rightarrow J^*$ be an injective resolution in \mathcal{P}_{d2^m} . Then one has a hypercohomology spectral sequence

$$E_1^{ij} = \text{Ext}_{\mathcal{F}}^i(F^{(m)}, J^j) \implies \text{Ext}_{\mathcal{F}}^{i+j}(F^{(m)}, T^{(m)}).$$

Since $F^{(m)}$ is a direct summand of the exponential functor $(\Gamma^* \otimes \cdots \otimes \Gamma^*)^{(m)}$ one can use Lemma 3.11 to show that $E_1^{ij} = 0$ for all $0 < i \leq 2^{m-d+1} - 2$. So for all $i \leq 2^{m-d+1} - 2$ one has

$$\begin{aligned} \text{Ext}_{\mathcal{F}}^i(F^{(m)}, T^{(m)}) &= H^i(\text{Hom}_{\mathcal{F}}(F^{(m)}, J^*)) = \\ &= H^i(\text{Hom}_{\mathcal{P}}(F^{(m)}, J^*)) = \text{Ext}_{\mathcal{P}}^i(F^{(m)}, T^{(m)}) \end{aligned}$$

and hence the proposition. □

4. STABLE K -THEORY FOR FINITE FIELDS

4.1. Stable K -theory

Stable K -theory gives the possibility of reducing calculation of homology of the general linear groups with twisted coefficients to the homology of the general linear group with constant coefficients. This trick goes back to Waldhausen. Let R be a ring and let \mathcal{V}_R be the category of finitely generated free R -modules. Moreover we let \mathcal{B} be the category of all functors $\mathcal{V}_R^{op} \times \mathcal{V}_R \rightarrow Ab$. For any such functor D the abelian group $D(R^n, R^n)$ has a natural $GL_n R$ -action, so one can consider the corresponding homology groups $H_*(GL_n(R), D(R^n, R^n))$. These groups form a direct system and we let $H_*(GL(R), D)$ be the corresponding limit when $n \rightarrow \infty$. It is a classical theorem of Dwyer that this system always stabilizes as soon as R satisfies some finiteness condition and D is of finite degree with respect to both variables (see [9]). Let us denote $H_0(GL(R), D)$ by $K_0^{st}(R, D)$. The functor

$$K_0^{st}(R, -) : \mathcal{B} \rightarrow Ab$$

is a right exact functor, whose left derived functors are denoted by $K_*^{st}(R, -)$ and called the *stable K -theory* of R . Similarly one can introduce the *cohomological version of stable K -theory* $K_{st}^*(R, -) : \mathcal{B} \rightarrow Ab$ as a left derived functor of the functor $K_{st}^0(R, -)$. Here $K_{st}^0(R, D) = H^0(GL(R), D)$ for any $D \in \mathcal{B}$. It was proved in [2] that in this way one recovers the original definition of Waldhausen for some class of rings including all fields and for such rings one has the following direct sum decomposition (see [2])

$$H_n(GL(R), D) \cong \bigoplus_{i+j=n} H_i(GL(R), K_j^{st}(R, D)),$$

where **the action of $GL(R)$ on K^{st} is trivial**. It was conjectured in [2] that the groups $K_*^{st}(R, D)$ are isomorphic to the homology $H_*(\mathcal{V}_R, D)$ of the category \mathcal{V}_R with the coefficients in the bifunctor D provided D is of finite degree with respect to both variables. This conjecture is much stronger compared to a previous conjecture from [28] (see also [19], page 293), which was proved by Dundas and McCarthy [8]. We refer the

reader to [29] for an explicit definition of $H_*(\mathcal{V}_R, D)$. For purposes of these notes it suffices to note that $H_*(\mathcal{V}_R, D) \cong (\text{Ext}_{\mathcal{F}}^*(T, F))^{\diamond}$ if $D(X, Y) = \text{Hom}_R(FX, TY)$. Here $T, F \in \mathcal{F}(R) := (R - \text{mod})^{\mathcal{V}_R}$ and the values of F lie in \mathcal{V}_R . By [2] the conjecture is true for a given ring R if and only if it is true for such bifunctors. Quillen proved in [33] that for a finite field K the (co)homology of $GL(K)$ with coefficients in K is trivial. Thus for a finite field K one has

$$K_*^{st}(K, D) \cong H_*(GL(K), D), \quad K_*^{st}(K, D) \cong H^*(GL(K), D)$$

and hence the following theorem proved by Betley and Suslin independently solves our conjecture for finite fields.

THEOREM 4.1. — *Let K be a finite field and let F and T be strict homogeneous functors of degree d . Then for each $i \geq 0$ the natural map*

$$\text{Ext}_{\mathcal{F}}^i(F, T) \rightarrow \text{Ext}_{GL_n(K)}^i(F(K^n), T(K^n))$$

is an isomorphism provided n is big enough.

We give here a sketch of the proof following Suslin (Appendix of [13]), see also [1] for a different argument.

First let us consider the case when $F = T = I$. In this case the theorem follows from the result of Dundas and McCarthy [8]. Following Friedlander and Suslin it can be proved also using the famous result of [5]. In our situation it says that for a fixed $m \geq 0$ there exist numbers $t(m)$ and $d(m)$ such that if $r \geq t(m)$ and $[K : \mathbf{F}_p] \geq d(m)$ then

$$\text{Ext}_{\mathcal{P}}^i(I^{(r)}, I^{(r)}) \rightarrow H^i(GL_n(K), M_n(K))$$

is an isomorphism for all $i \leq m$. Here $M_n(K)$ is the adjoint representation of $GL_n(K)$. Now comparing this result to Theorem 3.7 we see that for a fixed m Theorem 4.1 is true for $F = T = I$, $i \leq m$ if the field is big enough. In order to handle small fields one uses the following trick. Since the $GL_n(K)$ -module $M_n(K)$ is selfdual, one can pass to homology because $H_i(GL_n(K), M_n(K)) \cong H^i(GL_n(K), M_n(K))^{\diamond}$. Now in homology we have a possibility to change a field as follows.

Let $K \rightarrow L$ be an extension of finite fields of degree e . Then it yields a canonical homomorphism

$$\text{can} : H_*(GL(K), M(K)) \rightarrow H_*(GL(L), M(L)),$$

where $M(K) = \text{colim} M_n(K)$. Choosing a basis for L over K one gets an embedding $L \subset M_e(L)$ and hence a homomorphism $H_*(GL_n(L), M_n(L)) \rightarrow H_*(GL_{en}(K), M_{en}(K))$ and therefore

$$\text{Tr} : H_*(GL(L), M(L)) \rightarrow H_*(GL(K), M(K)).$$

PROPOSITION 4.2. — *The composition*

$$\text{Tr} \circ \text{can} : H_*(GL(K), M(K)) \rightarrow H_*(GL(K), M(K))$$

coincides with the multiplication by $e = [L : K]$, while the composition

$$\text{can} \circ \text{Tr} : H_*(GL(L), M(L)) \rightarrow H_*(GL(L), M(L))$$

coincides with $\sum_{\sigma \in \text{Gal}(L/K)} \sigma_*$.

In order to finish the case $F = T = I$ we choose d sufficiently big with respect to m and prime to p and consider an extension $K \rightarrow L$ of degree d . Then for L the theorem is true in dimensions $i \leq m$. By Theorem 3.5 (for $h = 0$) we have $H_i(GL_n(L), M_n(L)) = 0$ for odd $i \leq m$ and $H_i(GL_n(L), M_n(L)) = L$ for even $i \leq m$. Based on Proposition 4.2 it is not too difficult to see that the same relations hold for $H_i(GL_n(K), M_n(K))$. Hence Theorem 4.1 is true for $F = T = I$.

Following Suslin, for a functor $F \in \mathcal{F}$ we let $aF \in \mathcal{F}$ be the functor which is given on objects by

$$(aF)(V) = \bigoplus_{W \subset V} F(V/W).$$

If $f : V \rightarrow V'$ is a linear map, then the W' -component of the induced map

$$(aF)(f) : \bigoplus_{W \subset V} F(V/W) \rightarrow \bigoplus_{W' \subset V'} F(V'/W')$$

is zero on all W except $W = f^{-1}(W')$. In this case the corresponding component of $(aF)(f)$ is given by $f_* : F(V/f^{-1}(W')) \rightarrow F(V'/W')$.

PROPOSITION 4.3. — Assume $F \in \mathcal{F}$ has a projective resolution of finite type and let $T \in \mathcal{F}$ be a finite functor. Then one has a natural isomorphism

$$\text{Ext}_{GL(K)}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, aT).$$

Proof. — By a hypercohomology spectral sequence argument it suffices to consider only the case when $F = P_V$ is a standard projective and $T = J_W$ is a standard injective. One can prove the claim in dimension zero by direct considerations, while in dimensions > 0 it follows from the fact that the corresponding bifunctor $D(X, Y) = \text{Hom}(FX, TY)$ is injective in \mathcal{B} and hence $\text{Ext}_{GL(K)}^i(F, T) \cong K_{st}^i(K, D) = 0, i > 0$. \square

The canonical homomorphism $T \rightarrow aT$ yields a homomorphism $\sigma(F, T) : \text{Ext}_{\mathcal{F}}^*(F, T) \rightarrow \text{Ext}_{\mathcal{F}}^*(F, aT)$. Thanks to Proposition 4.3 this map is an isomorphism if and only if Theorem 4.1 is true for the pair (F, T) . We will also need the following

LEMMA 4.4. — Let E^* be an exponential functor and let $B_1, \dots, B_n \in \mathcal{F}$ be functors with the property $B_i(0) = 0, i = 1, \dots, n$. Assume further that the natural homomorphisms $\sigma(E^s, B_i)$ are isomorphisms for all $s < m$. Then $\sigma(A^m, B_1 \otimes \dots \otimes B_n)$ is an isomorphism as well.

Proof of Theorem 4.1. — By the first part of the proof we know that $\sigma(I, I)$ is an isomorphism and we have to show that $\sigma(F, T)$ is an isomorphism provided F and T satisfy the conditions of the Theorem. This can be done by induction on the degree of

functors based on Lemma 4.4 as follows. First one uses the exact sequence (3.1) (here we start to assume that $p = 2$, for odd p one needs to use De Rham and Koszul complexes). This sequence remains exact after applying the exact functor $a : \mathcal{F} \rightarrow \mathcal{F}$. By induction one readily proves that $\sigma(I, S^n)$ is an isomorphism for all n . Indeed it suffices to note that the induction assumption and Lemma 4.4 imply $\text{Ext}_{\mathcal{F}}^*(I, a(S^i \otimes S^j)) = 0$, for $i + j < n$ and one can use the same method as in Section 3. By duality one obtains that $\sigma(\Gamma^n, I)$ is an isomorphism as well. Having these facts in mind and repeating the previous argument one sees that $\sigma(\Gamma^n, S^m)$ is an isomorphism for all n, m too. Now Lemma 4.4 yields that $\sigma(F, T)$ is an isomorphism when F is projective in \mathcal{P} and T is an injective, and the hypercohomology spectral sequence gives the result. \square

REFERENCES

- [1] S. BETLEY. *Stable K-theory of finite fields*. *K-Theory* 17 (1999), no. 2, 103–111.
- [2] S. BETLEY & T. PIRASHVILI. *Stable K-theory as a derived functor*. *J. Pure Appl. Algebra* 96 (1994), no. 3, 245–258.
- [3] M. BÖKSTEDT. *On the topological Hochschild homology of \mathbf{Z} and $\mathbf{Z}/p\mathbf{Z}$* . Bielefeld (1987).
- [4] L. BREEN. *Extensions du groupe additif*. *Publ. IHES* 48 (1978), 39–125.
- [5] E. CLINE, B. PARSHALL, L. SCOTT & W. VAN DER KALLEN. *Rational and generic cohomology*. *Invent. Math.* 39 (1977), 143–163.
- [6] S. DONKIN. *On Schur algebras and related algebras*. I. *J. Algebra* 104 (1986), 310–328.
- [7] S. DONKIN. *On Schur algebras and related algebras*. II. *J. Algebra* 111 (1987), 354–364.
- [8] B. DUNDAS & R. MCCARTHY. *Stable K-theory and topological Hochschild homology*, *Ann. of Math.* 140 (1994), 685–701; erratum: *Ann. of Math.* 142 (1995), 425–426.
- [9] W. DWYER. *Twisted homological stability for general linear groups*. *Ann. of Math.* 111 (1980), 239–251.
- [10] S. EILENBERG. & S. MAC LANE. *On the groups $H(\pi, n)$* II. *Ann. of Math.* 60 (1954), 49–139.
- [11] V. FRANJOU. *Extensions entre puissances extérieures et entre puissances symétriques*. *J. Algebra* 179 (1996), 501–522.
- [12] V. FRANJOU, J. LANNES & L. SCHWARTZ. *Autour de la cohomologie de Mac Lane des corps finis*. *Invent. Math.* 115 (1994), 513–538.

- [13] V. FRANJOU, E. FRIEDLANDER, A. SCORICHENKO & A. SUSLIN. *General linear and functor cohomology over finite fields*. Ann. of Math. (2) 150 (1999), 663–728.
- [14] E. FRIEDLANDER & A. SUSLIN. *Cohomology of finite group schemes over a field*. Invent. Math. 127(1997), 209–270.
- [15] J.A. GREEN. *Polynomial representations of GL_n* . Lecture Notes in Mathematics 830. Springer-Verlag, (1980).
- [16] H-W. HENN, J. LANNES & L. SCHWARTZ. *The categories of unstable modules and unstable algebras over the Steenrod algebra modulo nilpotent objects*. Amer. J. Math. 115 (1993), 1053–1106.
- [17] H-W. HENN, J. LANNES & L. SCHWARTZ. *Localizations of unstable A -modules and equivariant mod p cohomology*. Math. Ann. 301 (1995), 23–68.
- [18] J.C. JANTZEN. *Representations of algebraic groups*. Pure and Applied Mathematics 131. Academic Press, Inc., Boston, MA (1987).
- [19] M. JIBLADZE & T. PIRASHVILI. *Cohomology of algebraic theories*. J. Algebra 137 (1991), 253–296.
- [20] N.J. KUHN. *Generic representations of the finite general linear groups and the Steenrod algebra. I*. Amer. J. Math. 116 (1994), 327–360.
- [21] N.J. KUHN. *Generic representations of the finite general linear groups and the Steenrod algebra. II. K -Theory 8* (1994), 395–428.
- [22] N.J. KUHN. *Generic representations of the finite general linear groups and the Steenrod algebra. III. K -Theory 9* (1995), 273–303.
- [23] N.J. KUHN. *On topologically realizing modules over the Steenrod algebra*. Ann. of Math. (2) 141 (1995), 321–347.
- [24] I.G. MACDONALD. *Symmetric functions and Hall polynomials*. Second edition. Oxford University Press (1995).
- [25] S. MARTIN. *Schur algebras and representation theory*. Cambridge University Press (1993).
- [26] T. PIRASHVILI. *Polynomial functors*. (Russian) Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR 91 (1988), 55–66.
- [27] T. PIRASHVILI. *Higher additivizations*. (Russian) Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR 91 (1988), 44–54.
- [28] T. PIRASHVILI. *New homology and cohomology of rings*. (Russian) Soobshch. Akad. Nauk Gruzin. SSR 133 (1989), no. 3, 477–480.
- [29] T. PIRASHVILI & F. WALDHAUSEN. *Mac Lane homology and topological Hochschild homology*. J. Pure Appl. Algebra 82 (1992), no. 1, 81–98.
- [30] L. PIRIOU & L. SCHWARTZ. *Extensions de foncteurs simples*. K -Theory 15 (1998), 269–291.

- [31] G. POWELL. *The Artinian conjecture for $I^{\otimes 2}$* . J. Pure Appl. Algebra 128 (1998), 291–310.
- [32] N. ROBY. *Lois polynômes et lois formelles en théorie des modules*. Ann. Sci. Éc. Norm. Sup. (3) 80 (1963) 213–348.
- [33] D. QUILLEN. *On the cohomology and K-theory of the general groups over a finite field*, Ann. of Math. 96 (1972), 552–586.
- [34] L. SCHWARTZ. *Unstable modules over the Steenrod algebra and Sullivan's fixed point set conjecture*. University of Chicago Press (1994).
- [35] L. SCHWARTZ. *À propos de la conjecture de non-réalisation due à N. Kuhn*. Invent. Math. 134 (1998), no. 1, 211–227.
- [36] B. TOTARO. *Projective resolutions of representations of $GL(n)$* . J. Reine Angew. Math. 482 (1997), 1–13.

Teimuraz PIRASHVILI

A.M. Razmadze Math. Institute

Aleksidze str. 1

Tbilisi, 380093, Georgia

E-mail : pira@rmi.acnet.ge

FAITHFUL LINEAR REPRESENTATIONS OF THE BRAID GROUPS

by Vladimir TURAEV

The braid group on n strings B_n can be defined as the group generated by $n - 1$ generators $\sigma_1, \dots, \sigma_{n-1}$ with defining relations

$$(0.1) \quad \sigma_i \sigma_j = \sigma_j \sigma_i$$

if $|i - j| \geq 2$ and

$$(0.2) \quad \sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$$

for $i = 1, \dots, n - 2$. This group was introduced by Emil Artin in 1926. It has various interpretations, specifically, as the group of geometric braids in \mathbf{R}^3 , as the mapping class group of an n -punctured disc, as the fundamental group of the configuration space of n points on the plane, etc. The algebraic properties of B_n have been studied by many authors. To mention a few, note the solution of the conjugacy problem in B_n given by F. Garside, the papers of N. Ivanov and J. McCarthy who proved that the mapping class groups and in particular B_n satisfy the “Tits alternative”, and the work of P. Dehornoy establishing the existence of a left-invariant total order on B_n (see [Ga], [Iv2], [Ka]).

One of the most intriguing problems in the theory of braids is the question of whether B_n is linear, i.e., whether it admits a faithful representation into a group of matrices over a commutative ring. This question has its origins in a number of interrelated facts and first of all in the discovery by W. Burau [Bu] of an n -dimensional linear representation of B_n which for a long time had been considered as a candidate for a faithful representation. However, as it was established by J. Moody in 1991 this representation is not faithful for $n \geq 9$. Later it was shown to be unfaithful for $n \geq 5$. Thus, the question of the linearity of B_n remained open.

In 1999/2000 there appeared a series of papers of D. Krammer and S. Bigelow who proved that B_n is linear for all n . First there appeared a paper of Krammer

[Kr1] in which he constructed a homomorphism from B_n to $GL(n(n-1)/2, R)$ where $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ is the ring of Laurent polynomials on two variables. He proved that this homomorphism is injective for $n = 4$ and conjectured that the same is true for all n . Soon after that, Bigelow [Bi2] gave a beautiful topological proof of this conjecture. Another proof based on different ideas was obtained by Krammer [Kr2] independently.

The ring $\mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ can be embedded in the field of real numbers by assigning to q, t any algebraically independent non-zero real values. Therefore B_n embeds in $GL(n(n-1)/2, \mathbf{R})$. As an application, note that the linearity of B_4 implies the linearity of the group $\text{Aut}(F_2)$, where F_n is a free group of rank n (see [DFG]). It is known that $\text{Aut}(F_n)$ is not linear for $n \geq 3$, see [FP].

The representation of B_n considered by Krammer and Bigelow is one of a family of representations introduced earlier by R. Lawrence [La]. Her work was inspired by a study of the Jones polynomial of links and was concerned with representations of Hecke algebras arising from the actions of braids on homology of configuration spaces.

The same representation of B_n arises from a study of the so-called Birman-Murakami-Wenzl algebra C_n . This algebra is a quotient of the group ring $\mathbf{C}[B_n]$ by certain relations inspired by the theory of link polynomials. The irreducible finite dimensional representations of C_n were described by H. Wenzl in terms of Young diagrams. These representations yield irreducible finite dimensional representations of B_n . One of them was shown by M. Zinno [Zi] and independently by V. Jones to be equivalent to the Krammer representation which is henceforth irreducible.

The aim of this paper is to present these results. In Sect. 1 we consider the Burau representation and explain why it is not faithful. In Sect. 2 we outline Bigelow's approach following [Bi2]. In Sect. 3 we discuss the work of Krammer [Kr2]. Finally in Sect. 4 we discuss the Birman-Murakami-Wenzl algebras and the work of Zinno.

The author is indebted to S. Bigelow for useful comments on his work and to Ch. Kassel for careful reading of a preliminary version of this paper.

1. THE BURAU REPRESENTATION OF THE BRAID GROUP

1.1. Mapping class groups. It will be convenient for us to view the braid group as the mapping class group of a punctured disc. We recall here the definition and a few simple properties of the mapping class groups.

Let Σ be a connected oriented surface. By a *self-homeomorphism* of Σ we mean an orientation preserving homeomorphism $\Sigma \rightarrow \Sigma$ which fixes $\partial\Sigma$ pointwise. Two such

homeomorphisms are *isotopic* if they can be included in a continuous one-parameter family of self-homeomorphisms of Σ . The mapping class group $\text{Homeo}(\Sigma)$ of Σ is the group of isotopy classes of self-homeomorphisms of Σ with the group operation determined by composition.

Each self-homeomorphism of Σ induces an automorphism of the abelian group $H = H_1(\Sigma; \mathbf{Z})$. This is a “homological” representation $\text{Homeo}(\Sigma) \rightarrow \text{Aut}(H)$. The action of homeomorphisms preserves the skew-symmetric bilinear form $H \times H \rightarrow \mathbf{Z}$ determined by the algebraic intersection number. The value $[\alpha] \cdot [\beta] \in \mathbf{Z}$ of this form on the homology classes $[\alpha], [\beta] \in H$ represented by oriented loops α, β on Σ is computed as follows. Assume that α and β lie in a generic position so that they meet each other transversely in a finite set of points which are not self-crossings of α or β . Then

$$[\alpha] \cdot [\beta] = \sum_{p \in \alpha \cap \beta} \varepsilon_p$$

where $\varepsilon_p = +1$ if the tangent vectors of α, β at p form a positively oriented basis and $\varepsilon_p = -1$ otherwise.

An example of a self-homeomorphism of Σ is provided by the Dehn twist τ_α about a simple closed curve $\alpha \subset \Sigma$. It is defined as follows. Identify a regular neighborhood of α in Σ with the cylinder $S^1 \times [0, 1]$ so that $\alpha = S^1 \times (1/2)$. We choose this identification so that the product of the counterclockwise orientation on $S^1 = \{z \in \mathbf{C} \mid |z| = 1\}$ and the right-handed orientation on $[0, 1]$ corresponds to the given orientation on Σ . The Dehn twist $\tau_\alpha : \Sigma \rightarrow \Sigma$ is the identity map outside $S^1 \times [0, 1]$ and sends any $(x, s) \in S^1 \times [0, 1]$ to $(e^{2\pi i s} x, s) \in S^1 \times [0, 1]$. To compute the action of τ_α in homology, we orient α in an arbitrary way. The effect of τ_α on an oriented curve transversal to α is to insert $\alpha^{\pm 1}$ at each crossing of α with this curve, where ± 1 is the sign of the crossing. Therefore for any $g \in H$,

$$(1.1) \quad (\tau_\alpha)_*(g) = g + ([\alpha] \cdot g)[\alpha]$$

Note that τ_α and its action on H do not depend on the choice of orientation on α .

A similar construction applies to arcs in Σ whose endpoints are punctures. Assume that Σ is obtained from another surface Σ' by puncturing, i.e., by removing a finite set of points lying in $\text{Int}(\Sigma')$. These points will be called *punctures*. Let α be an embedded arc in Σ' whose endpoints are punctures x_1, x_2 and whose interior lies in Σ . One can define the Dehn “half-twist” $\tau_\alpha : \Sigma \rightarrow \Sigma$ which is the identity map outside a regular neighborhood of α in Σ' and which exchanges x_1 and x_2 . This homeomorphism is obtained by the isotopy of the identity map of Σ' rotating α about

its midpoint to the angle of π in the direction provided by the orientation of Σ . Restricting the resulting homeomorphism of $\Sigma' \rightarrow \Sigma'$ to Σ we obtain τ_α . To compute the action of τ_α on $H = H_1(\Sigma)$ we orient α from x_1 to x_2 and associate to α a loop α' in Σ as follows. Choose a point $z \in \alpha$ and for $i = 1, 2$ denote by μ_i the loop in Σ beginning at z and moving along α until coming very closely to x_i , then encircling x_i in the direction determined by the orientation of Σ and finally moving back to z along α . Set $\alpha' = \mu_1^{-1}\mu_2$. The homotopy class of the loop α' on Σ does not depend on the choice of z . The effect of τ_α on an oriented curve transversal to α is to insert $(\alpha')^{\pm 1}$ at each crossing of α with this curve. Thus for any $g \in H$, we have

$$(1.2) \quad (\tau_\alpha)_*(g) = g + ([\alpha] \cdot g)[\alpha']$$

where $[\alpha] \cdot g = -g \cdot [\alpha] \in \mathbf{Z}$ is the algebraic intersection number of g with the 1-dimensional homology class $[\alpha]$ of Σ “modulo infinity” represented by α .

In general, the action of $\text{Homeo}(\Sigma)$ on $H = H_1(\Sigma)$ is not faithful. We point out one source of non-faithfulness. If $\alpha, \beta \subset \Sigma$ are simple closed curves with $[\alpha] \cdot [\beta] = 0$ then formula (1.1) implies that $(\tau_\alpha)_*$ and $(\tau_\beta)_*$ commute in $\text{Aut}(H)$. The Dehn twists τ_α, τ_β themselves commute if and only if α is isotopic to a simple closed curve disjoint from β , see for instance [Iv1]. It is easy to give examples of simple closed curves $\alpha, \beta \subset \Sigma$ which are not disjoint up to isotopy but have zero algebraic intersection number. Then the commutator $[\tau_\alpha, \tau_\beta]$ lies in the kernel of the homological representation. Using (1.2), one can similarly derive elements of the kernel from embedded arcs with endpoints in punctures.

1.2. Braid groups. Let $D = \{z \in \mathbf{C} \mid |z| \leq 1\}$ be the unit disc with counterclockwise orientation. Fix a set of $n \geq 1$ distinct punctures $X = \{x_1, \dots, x_n\} \subset \text{Int}(D)$. We shall assume that $x_1, \dots, x_n \in (-1, +1) = \mathbf{R} \cap \text{Int}(D)$ and $x_1 < x_2 < \dots < x_n$. Set $D_n = D \setminus X$. The group $\text{Homeo}(D_n)$ is denoted B_n and called the n -th *braid group*. An element of B_n is an isotopy class of a homeomorphism $D_n \rightarrow D_n$ which fixes $\partial D_n = \partial D = S^1$ pointwise. Such a homeomorphism uniquely extends to a homeomorphism $D \rightarrow D$ permuting x_1, \dots, x_n . This defines a group homomorphism from B_n onto the symmetric group S_n . We can equivalently define B_n as the group of isotopy classes of homeomorphisms $D \rightarrow D$ which fix ∂D pointwise and preserve X as a set.

For $i = 1, \dots, n-1$, the linear interval $[x_i, x_{i+1}] \subset (-1, +1) \subset \mathbf{R}$ is an embedded arc in D with endpoints in the punctures x_i, x_{i+1} . The corresponding Dehn half-twist $D_n \rightarrow D_n$ is denoted by σ_i . It is a classical fact that B_n is generated by $\sigma_1, \dots, \sigma_{n-1}$ with defining relations (0.1), (0.2). The image of σ_i in S_n is the permutation $(i, i+1)$.

Another definition of B_n can be given in terms of braids. A (geometric) *braid on n strings* is an n -component one-dimensional manifold $E \subset D \times [0, 1]$ such that E meets $D \times \{0, 1\}$ orthogonally along the set $(X \times 0) \cup (X \times 1)$ and the projection on $[0, 1]$ maps each component of E homeomorphically onto $[0, 1]$. The braids are considered up to isotopy in $D \times [0, 1]$ constant on the endpoints. The group operation in the set of braids is defined by glueing one braid on the top of the other one and compressing the result into $D \times [0, 1]$.

The equivalence between these two definitions of B_n is established as follows. Any homeomorphism $h : D \rightarrow D$ which fixes ∂D pointwise is related to the identity map id_D by an isotopy $\{h_s : D \rightarrow D\}_{s \in [0, 1]}$ such that $h_0 = \text{id}_D$ and $h_1 = h$. If $h(X) = X$ then the set $\cup_{s \in [0, 1]} (h_s(X) \times s) \subset D \times [0, 1]$ is a braid. Its isotopy class depends only on the element of B_n represented by h . This establishes an isomorphism between B_n and the group of braids on n strings. The generator $\sigma_i \in B_n$ corresponds to the i -th “elementary” braid represented by a plane diagram consisting of n linear intervals which are disjoint except at one intersection point where the i -th interval goes over the $(i + 1)$ -th interval.

1.3. The Burau representation. Let Λ denote the ring $\mathbf{Z}[t, t^{-1}]$. The Burau representation $B_n \rightarrow GL_n(\Lambda)$ sends the i -th generator $\sigma_i \in B_n$ into the matrix

$$I_{i-1} \oplus \begin{pmatrix} 1-t & t \\ 1 & 0 \end{pmatrix} \oplus I_{n-i-1}$$

where I_k denotes the identity $(k \times k)$ -matrix and the non-trivial (2×2) -block appears in the i -th and $(i + 1)$ -th rows and columns (see [Bu]). Substituting $t = 1$, we obtain the standard representation of the symmetric group S_n by permutation matrices or equivalently the homological action of B_n on $H_1(D_n) = \mathbf{Z}^n$. The Burau representation is reducible: it splits as a direct sum of an $(n - 1)$ -dimensional representation and the trivial one-dimensional representation.

The Burau representation is known to be faithful for $n \leq 3$, see [Bir]. J. Moody [Mo] proved in 1991 that it is not faithful for $n \geq 9$. D. Long and M. Paton [LP] extended Moody’s argument to $n \geq 6$. Recently, S. Bigelow [Bi1] proved that this representation is not faithful for $n = 5$. The case $n = 4$ remains open.

The geometric idea allowing to detect non-trivial elements in the kernel of the Burau representation is parallel to the one at the end of Sect. 1.1. We first give a homological description of the Burau representation. Observe that $H_1(D \setminus x_i) = \mathbf{Z}$ is generated by the class of a small loop encircling x_i in the counterclockwise direction. Each loop in $D \setminus x_i$ represents k times the generator where k is the *winding number* of

the loop around x_i . Consider the homomorphism $H_1(D_n) \rightarrow \mathbf{Z}$ sending the homology class of a loop to its *total winding number* defined as the sum of its winding numbers around x_1, \dots, x_n . Let $\tilde{D}_n \rightarrow D_n$ be the corresponding regular covering. The group of covering transformations of \tilde{D}_n is \mathbf{Z} which we write as a multiplicative group with generator t . The group $H_1(\tilde{D}_n)$ acquires thus the structure of a Λ -module. It is easy to check that this is a free Λ -module of rank $n - 1$.

Fix a basepoint $d \in \partial D$. Any homeomorphism $h : D_n \rightarrow D_n$ representing an element of B_n lifts uniquely to a homeomorphism $\tilde{h} : \tilde{D}_n \rightarrow \tilde{D}_n$ which fixes the fiber over d pointwise. This induces a Λ -linear automorphism \tilde{h}_* of $H_1(\tilde{D}_n)$. The map $h \mapsto \tilde{h}_*$ defines a representation $B_n \rightarrow \text{Aut}(H_1(\tilde{D}_n))$ equivalent to the $(n - 1)$ -dimensional Burau representation.

Now we extend the algebraic intersection to arcs and refine it so that it takes values in Λ . Let α, β be two embedded oriented arcs in D_n with endpoints in the punctures. (We assume that all four endpoints of α, β are distinct so that $n \geq 4$). Let $\tilde{\alpha}, \tilde{\beta}$ be lifts of α, β to \tilde{D}_n , respectively. Set

$$\langle \alpha, \beta \rangle = \sum_{k \in \mathbf{Z}} (t^k \tilde{\alpha} \cdot \tilde{\beta}) t^k \in \Lambda$$

where $t^k \tilde{\alpha} \cdot \tilde{\beta} \in \mathbf{Z}$ is the algebraic intersection number of the arcs $t^k \tilde{\alpha}, \tilde{\beta}$ in \tilde{D}_n . This finite sum is only defined up to multiplication by a power of t depending on the choice of $\tilde{\alpha}, \tilde{\beta}$. This will not be important for us since we are only interested in whether or not $\langle \alpha, \beta \rangle = 0$. To compute $\langle \alpha, \beta \rangle$ explicitly one deforms α in general position with respect to β . Then $\langle \alpha, \beta \rangle = \sum_{p \in \alpha \cap \beta} \varepsilon_p t^{k_p}$ where $\varepsilon_p = \pm 1$ is the intersection sign at p and $k_p \in \mathbf{Z}$. The exponents $\{k_p\}_p$ are determined by the following condition: if $p, q \in \alpha \cap \beta$, then $k_p - k_q$ is the total winding number of the loop going from p to q along α and then from q to p along β . Note that

$$\langle \beta, \alpha \rangle = \sum_{k \in \mathbf{Z}} (t^k \tilde{\beta} \cdot \tilde{\alpha}) t^k = \sum_{k \in \mathbf{Z}} (\tilde{\beta} \cdot t^{-k} \tilde{\alpha}) t^k = - \sum_{k \in \mathbf{Z}} (t^{-k} \tilde{\alpha} \cdot \tilde{\beta}) t^k = -\overline{\langle \alpha, \beta \rangle}$$

where the overline denotes the involution in Λ sending any t^k to t^{-k} . Hence $\langle \alpha, \beta \rangle = 0$ if and only if $\langle \beta, \alpha \rangle = 0$.

We claim that if $\langle \alpha, \beta \rangle = 0$, then the automorphisms $(\tilde{\tau}_\alpha)_*, (\tilde{\tau}_\beta)_*$ of $H_1(\tilde{D}_n)$ commute. Observe that the loops α', β' on D_n associated to α, β as in Sect. 1.1 have zero total winding numbers and therefore lift to certain loops $\tilde{\alpha}', \tilde{\beta}'$ in \tilde{D}_n . The effect of $\tilde{\tau}_\alpha$ on any oriented loop γ in \tilde{D}_n is to insert a lift of $(\alpha')^{\pm 1}$ at each crossing of γ with the preimage of α in \tilde{D}_n . Thus

$$(\tilde{\tau}_\alpha)_*([\gamma]) = [\gamma] + \lambda_\gamma [\tilde{\alpha}']$$

for a certain Laurent polynomial $\lambda_\gamma \in \Lambda$. The coefficients of λ_γ are the algebraic intersection numbers of γ with lifts of α to \tilde{D}_n . By $\langle \alpha, \beta \rangle = 0$, any lift of α has algebraic intersection number zero with any lift of β and hence with any lift of β' . Therefore, $\lambda_{\tilde{\beta}'} = 0$ and $(\tilde{\tau}_\alpha)_*([\tilde{\beta}']) = [\tilde{\beta}']$. Similarly, $(\tilde{\tau}_\beta)_*([\gamma]) = [\gamma] + \mu_\gamma [\tilde{\beta}']$ with $\mu_\gamma \in \Lambda$ and $(\tilde{\tau}_\beta)_*([\tilde{\alpha}']) = [\tilde{\alpha}']$. We conclude that

$$(\tilde{\tau}_\alpha \tilde{\tau}_\beta)_*([\gamma]) = [\gamma] + \lambda_\gamma [\tilde{\alpha}'] + \mu_\gamma [\tilde{\beta}'] = (\tilde{\tau}_\beta \tilde{\tau}_\alpha)_*([\gamma]).$$

To show that the Burau representation is not faithful it remains to provide an example of oriented embedded arcs α, β in D_n with endpoints in distinct punctures such that $\langle \alpha, \beta \rangle = 0$ and $\tau_\alpha \tau_\beta \neq \tau_\beta \tau_\alpha$ in B_n . For $n \geq 6$, the simplest known example (see [Bi1]) is provided by the pair $\alpha = \varphi_1([x_3, x_4])$, $\beta = \varphi_2([x_3, x_4])$ where

$$\varphi_1 = \sigma_1^2 \sigma_2^{-1} \sigma_5^{-2} \sigma_4, \quad \varphi_2 = \sigma_1^{-1} \sigma_2 \sigma_5 \sigma_4^{-1}.$$

To compute $\langle \alpha, \beta \rangle$ one can draw α, β and use the recipe above. To prove that the braids $\tau_\alpha = \varphi_1 \sigma_3 \varphi_1^{-1}$ and $\tau_\beta = \varphi_2 \sigma_3 \varphi_2^{-1}$ do not commute, one can use the solution of the word problem in B_n or the methods of the Thurston theory of surfaces (cf. Sect. 2.2). Thus the commutator $[\tau_\alpha, \tau_\beta]$ lies in the kernel of the Burau representation. This commutator can be represented by a word of length 44 in the generators $\sigma_1, \dots, \sigma_5$.

Similar ideas apply in the case $n = 5$, although one has to extend them to arcs relating the punctures to the base point $d \in \partial D$. The shortest known word in the generators $\sigma_1, \dots, \sigma_4$ representing an element of the kernel has length 120.

2. THE WORK OF BIGELOW

2.1. A representation of B_n . We use the notation $D, D_n, X = \{x_1, \dots, x_n\}$ introduced in Sect. 1. Let C be the space of all unordered pairs of distinct points in D_n . This space is obtained from $(D_n \times D_n) \setminus \text{diagonal}$ by the identification $\{x, y\} = \{y, x\}$ for any distinct $x, y \in D_n$. It is clear that C is a connected non-compact 4-manifold with boundary. It has a natural orientation induced by the counterclockwise orientation of D_n . Set $d = -i \in \partial D$ where $i = \sqrt{-1}$ and $d' = -i e^{\frac{\varepsilon \pi i}{2}} \in \partial D$ with small positive ε . We take $c_0 = \{d, d'\}$ as the basepoint for C .

A closed curve $\alpha : [0, 1] \rightarrow C$ can be written in the form $\alpha(s) = \{\alpha_1(s), \alpha_2(s)\}$ where $s \in [0, 1]$ and α_1, α_2 are arcs in D_n such that $\{\alpha_1(0), \alpha_2(0)\} = \{\alpha_1(1), \alpha_2(1)\}$. The arcs α_1, α_2 are either both loops or can be composed with each other. They form thus a closed oriented one-manifold mapped to D_n . Let $a(\alpha) \in \mathbf{Z}$ be the total winding

number of this one-manifold around the punctures $\{x_1, \dots, x_n\}$. Composing the map $s \mapsto (\alpha_1(s) - \alpha_2(s))/|\alpha_1(s) - \alpha_2(s)| : [0, 1] \rightarrow S^1$ with the projection $S^1 \rightarrow \mathbf{RP}^1$ we obtain a loop in RP^1 . The corresponding element of $H_1(RP^1) = \mathbf{Z}$ is denoted by $b(\alpha)$. The formula $\alpha \mapsto q^{a(\alpha)}t^{b(\alpha)}$ defines a homomorphism, ϕ , from $H_1(C)$ to the (multiplicatively written) free abelian group with basis q, t . Let $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ be the group ring of this group.

Let $\tilde{C} \rightarrow C$ be a regular covering corresponding to the kernel of ϕ . The generators q, t act on \tilde{C} as commuting covering transformations. The homology group $H_2(\tilde{C}) = H_2(\tilde{C}; \mathbf{Z})$ becomes thus an R -module.

Any self-homeomorphism h of D_n induces by $h(\{x, y\}) = \{h(x), h(y)\}$ a homeomorphism $C \rightarrow C$ also denoted h . It is easy to check that $h(c_0) = c_0$ and the action of h on $H_1(C)$ commutes with ϕ . Therefore this homeomorphism $C \rightarrow C$ lifts uniquely to a map $\tilde{h} : \tilde{C} \rightarrow \tilde{C}$ which fixes the fiber over c_0 pointwise and commutes with the covering transformations. Consider the representation $B_n \rightarrow \text{Aut}(H_2(\tilde{C}))$ sending the isotopy class of h to the R -linear automorphism \tilde{h}_* of $H_2(\tilde{C})$.

2.2. Theorem (S. Bigelow [Bi2]). – *The representation $B_n \rightarrow \text{Aut}(H_2(\tilde{C}))$ is faithful for all $n \geq 1$.*

We outline below the main ideas of Bigelow's proof. The proof uses almost no information about the structure of the R -module $H_2(\tilde{C})$. The only thing needed is the absence of R -torsion or more precisely the fact that multiplication by a non-zero polynomial of type $q^a t^b - 1$ has zero kernel in $H_2(\tilde{C})$. In fact, $H_2(\tilde{C})$ is a free R -module of rank $n(n-1)/2$, as it was essentially shown in [La].

We shall use one well-known fact concerning isotopies of arcs on surfaces. Let N, T be embedded arcs in D_n with distinct endpoints lying either in the punctures or on ∂D_n . Assume that the interiors of N, T do not meet ∂D_n , and that N intersects T transversely (in a finite number of points). A *bigon* for the pair (N, T) is an embedded disc in $\text{Int}(D_n)$ whose boundary is formed by one subarc of N and one subarc of T and whose interior is disjoint from N and T . It is clear that in the presence of a bigon there is an isotopy of T constant on the endpoints and decreasing $\#(N \cap T)$ by two. Thurston's theory of surfaces implies the converse: if there is an isotopy of T (rel endpoints) decreasing $\#(N \cap T)$ then the pair (N, T) has at least one bigon, cf. [FLP, Prop. 3.10].

2.3. Noodles and forks. We need the following notation. For arcs $\alpha, \beta : [0, 1] \rightarrow D_n$ such that $\alpha(s) \neq \beta(s)$ for all $s \in [0, 1]$, we denote by $\{\alpha, \beta\}$ the arc in C given by

$\{\alpha, \beta\}(s) = \{\alpha(s), \beta(s)\}$. We fix once and forever a point $\tilde{c}_0 \in \tilde{C}$ lying over $c_0 \in C$.

A *noodle* in D_n is an embedded arc $N \subset D_n$ with endpoints d and d' . For a noodle N , the set $\Sigma_N = \{\{x, y\} \in C \mid x, y \in N, x \neq y\}$ is a surface in C containing c_0 . This surface is homeomorphic to a triangle with one edge removed. We orient N from d to d' and orient Σ_N as follows: at a point $\{x, y\} = \{y, x\} \in \Sigma_N$ such that x is closer to d along N than y , the orientation of Σ_N is the product of the orientations of N at x and y in this order. Let $\tilde{\Sigma}_N$ be the lift of Σ_N to \tilde{C} containing \tilde{c}_0 . The orientation of Σ_N lifts to $\tilde{\Sigma}_N$ in the obvious way. Clearly, $\tilde{\Sigma}_N$ is a proper surface in \tilde{C} in the sense that $\tilde{\Sigma}_N \cap \partial\tilde{C} = \partial\tilde{\Sigma}_N$.

A *fork* in D_n is an embedded tree $F \subset D$ formed by three edges and four vertices d, x_i, x_j, z such that $F \cap \partial D = d, F \cap X = \{x_i, x_j\}$ and z is a common vertex of all 3 edges. The edge, H , relating d to z is called the *handle* of F . The union, T , of the other two edges is an embedded arc with endpoints $\{x_i, x_j\}$. This arc is called the *tines* of F . Note that in a small neighborhood of z , the handle H lies on one side of T which distinguishes a side of T . We orient T so that its distinguished side lies on its right. The handle H also has a distinguished side determined by d' . Pushing slightly the graph $F = T \cup H$ to the distinguished side (fixing the vertices x_i, x_j and pushing d to d') we obtain a “parallel copy” $F' = T' \cup H'$. The graph F' is a fork with handle H' , tines T' , and vertices d', x_i, x_j, z' where $z' = T' \cap H'$ lies on the distinguished side of both T and H . We can assume that F' meets F only in common vertices $\{x_i, x_j\} = T \cap T'$ and in one point lying on $H \cap T'$ close to z, z' .

For a fork F , the set $\Sigma_F = \{\{y, y'\} \in C \mid y \in T \setminus \{x_i, x_j\}, y' \in T' \setminus \{x_i, x_j\}\}$ is a surface in C homeomorphic to $(0, 1)^2$. Let α_0 be an arc from d to z along H and let α'_0 be an arc from d' to z' along H' . Consider the arc $\{\alpha_0, \alpha'_0\}$ in C and denote by $\tilde{\alpha}$ its lift to \tilde{C} which starts in \tilde{c}_0 . Let $\tilde{\Sigma}_F$ be the lift of Σ_F to \tilde{C} which contains the lift $\tilde{\alpha}(1)$ of the point $\{z, z'\} \in \Sigma_F$. The surfaces Σ_F and $\tilde{\Sigma}_F$ have a natural orientation determined by the orientation of T and the induced orientation of T' .

We shall use the surfaces $\tilde{\Sigma}_N, \tilde{\Sigma}_F$ to establish a duality between noodles and forks. More precisely, for any noodle N and any fork F we define an element $\langle N, F \rangle$ of R as follows. By applying a preliminary isotopy we can assume that N intersects T transversely in $m \geq 0$ points z_1, \dots, z_m (the numeration is arbitrary; the intersection of N with H may be not transversal). We choose the parallel fork $F' = T' \cup H'$ as above so that T' meets N transversely in m points z'_1, \dots, z'_m where each pair z_i, z'_i is joined by a short arc in N which lies in the narrow strip bounded by $T \cup T'$ and meets no other z_j, z'_j . Then the surfaces Σ_F and Σ_N intersect transversely in m^2 points $\{z_i, z'_j\}$ where $i, j = 1, \dots, m$. Therefore for any $a, b \in \mathbf{Z}$, the image $q^a t^b \tilde{\Sigma}_N$

of $\tilde{\Sigma}_N$ under the covering transformation $q^a t^b$ meets $\tilde{\Sigma}_F$ transversely. Consider the algebraic intersection number $q^a t^b \tilde{\Sigma}_N \cdot \tilde{\Sigma}_F \in \mathbf{Z}$ and set

$$\langle N, F \rangle = \sum_{a,b \in \mathbf{Z}} (q^a t^b \tilde{\Sigma}_N \cdot \tilde{\Sigma}_F) q^a t^b.$$

The sum on the right-hand side is finite (it has $\leq m^2$ terms) and thus defines an element of R .

2.4. Lemma. – $\langle N, F \rangle$ is invariant under isotopies of N and F in D_n constant on the endpoints.

Proof. – We first compute $\langle N, F \rangle$ explicitly. Let $N \cap T = \{z_1, \dots, z_m\}$ and $N \cap T' = \{z'_1, \dots, z'_m\}$ as above. For every pair $i, j \in \{1, \dots, m\}$, there exist unique integers $a_{i,j}, b_{i,j} \in \mathbf{Z}$ such that $q^{a_{i,j}} t^{b_{i,j}} \tilde{\Sigma}_N$ intersects $\tilde{\Sigma}_F$ at a point lying over $\{z_i, z'_j\} \in C$. Let $\varepsilon_{i,j} = \pm 1$ be the sign of that intersection. Then

$$(2.1) \quad \langle N, F \rangle = \sum_{i=1}^m \sum_{j=1}^m \varepsilon_{i,j} q^{a_{i,j}} t^{b_{i,j}}.$$

The numbers $a_{i,j}, b_{i,j}$ can be computed as follows. Let α_0 be an arc from d to z along H and let α'_0 be an arc from d' to z' along H' . Let β_i be an arc from z to z_i along T and let β'_j be an arc from z' to z'_j along T' . Finally, let $\gamma_{i,j}$ and $\gamma'_{i,j}$ be disjoint arcs in N connecting the points z_i, z'_j to the endpoints of N . Note that $\delta_{i,j} = \{\alpha_0, \alpha'_0\} \{\beta_i, \beta'_j\} \{\gamma_{i,j}, \gamma'_{i,j}\}$ is a loop in C . Then

$$(2.2) \quad q^{a_{i,j}} t^{b_{i,j}} = \phi(\delta_{i,j}).$$

Indeed, we can lift $\delta_{i,j}$ to a path $\tilde{\alpha}\tilde{\beta}\tilde{\gamma}$ in \tilde{C} beginning at \tilde{c}_0 where $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ are lifts of $\{\alpha_0, \alpha'_0\}, \{\beta_i, \beta'_j\}, \{\gamma_{i,j}, \gamma'_{i,j}\}$, respectively. By definition of $\tilde{\Sigma}_F$, the point $\tilde{\alpha}(1) = \tilde{\beta}(0)$ lies on $\tilde{\Sigma}_F$. Hence the lift $\tilde{\beta}$ of $\{\beta_i, \beta'_j\}$ lies on $\tilde{\Sigma}_F$. The path $\tilde{\alpha}\tilde{\beta}\tilde{\gamma}$ ends at $\phi(\delta_{i,j})(\tilde{c}_0) = \tilde{\gamma}(1)$. Hence the lift $\tilde{\gamma}$ of $\{\gamma_{i,j}, \gamma'_{i,j}\}$ lies on $\phi(\delta_{i,j})\tilde{\Sigma}_N$. Therefore the point $\tilde{\beta}(1) = \tilde{\gamma}(0)$ lying over $\{z_i, z'_j\}$ belongs to both $\tilde{\Sigma}_F$ and $\phi(\delta_{i,j})\tilde{\Sigma}_N$. This yields (2.2).

Note that the residue $b_{i,j} \pmod{2}$ depends on whether the two points of D_n forming a point in C switch places moving along the loop $\delta_{i,j}$. This is determined by which of z_i, z'_j lies closer to d along N .

To compute $\varepsilon_{i,j}$ we observe that $\varepsilon_{i,j}$ is determined by the signs of the intersections of N and T at z_i, z_j and by which of z_i, z'_j lies closer to d on N . The sign of the

intersection of N and T at z_i is $+$ if N crosses T from the left to the right and is $-$ otherwise. By our choice of orientations on N and T , this sign is $+$ if z_i lies closer to d along N than z'_i and is $-$ otherwise. Hence, this sign is $(-1)^{b_{i,i}}$. Therefore $\varepsilon_{i,j}$ is determined by $b_{i,i} + b_{j,j} + b_{i,j} \pmod{2}$. A precise computation shows that

$$(2.3) \quad \varepsilon_{i,j} = -(-1)^{b_{i,i} + b_{j,j} + b_{i,j}}.$$

Now we can prove the lemma. It suffices to fix F and to prove that $\langle N, F \rangle$ is invariant under isotopies of N . A generic isotopy of N in D_n can be split into a finite sequence of local moves of two kinds: (i) isotopies keeping N transversal to T , (ii) a move pushing a small subarc of N across a subarc of $T \setminus z$. It is clear from the discussion above that the move (i) does not change $\langle N, F \rangle$. The move (ii) adds two new intersection points z_{m+1}, z_{m+2} to the set $N \cap T = \{z_1, \dots, z_m\}$. It follows from definitions and the discussion above that for any $i = 1, \dots, m + 2$,

$$a_{i,m+1} = a_{i,m+2}, \quad b_{i,m+1} = b_{i,m+2}, \quad \varepsilon_{i,m+1} = -\varepsilon_{i,m+2}.$$

Hence for any $i = 1, \dots, m + 2$, the terms $\varepsilon_{i,m+1} q^{a_{i,m+1}} t^{b_{i,m+1}}$ and $\varepsilon_{i,m+2} q^{a_{i,m+2}} t^{b_{i,m+2}}$ cancel each other. Similarly, for any $j = 1, \dots, m$, the terms $\varepsilon_{i,j} q^{a_{i,j}} t^{b_{i,j}}$ with $i = m + 1, m + 2$ cancel each other. Therefore $\langle N, F \rangle$ is the same before and after the move.

2.5. Lemma. *– $\langle N, F \rangle = 0$ if and only if there is an isotopy $\{T(s)\}_{s \in [0,1]}$ of the tines $T = T(0)$ of F in D_n (rel endpoints) such that $T(1)$ is disjoint from N .*

Proof. – Any isotopy $\{T(s)\}_{s \in [0,1]}$ of $T = T(0)$ extends to an ambient isotopy of D_n constant on ∂D_n and therefore extends to an isotopy $\{F(s)\}_{s \in [0,1]}$ of the fork $F = F(0)$. If $T(1)$ is disjoint from N then by Lemma 2.4, $\langle N, F \rangle = \langle N, F(1) \rangle = 0$.

The hard part of the lemma is the opposite implication. By applying a preliminary isotopy, we can assume that T intersects N transversely at a *minimal* number of points z_1, \dots, z_m with $m \geq 0$. We assume that $m \geq 1$ and show that $\langle N, F \rangle \neq 0$. To this end we use the lexicographic ordering on monomials $q^a t^b$. Namely we write $q^a t^b \geq q^{a'} t^{b'}$ with $a, b, a', b' \in \mathbf{Z}$ if either $a > a'$ or $a = a'$ and $b \geq b'$. We say that the ordered pair (i, j) with $i, j \in \{1, \dots, m\}$ is *maximal* if $q^{a_{i,j}} t^{b_{i,j}} \geq q^{a_{k,l}} t^{b_{k,l}}$ for any $k, l \in \{1, \dots, m\}$. We claim that

$$(*) \text{ if the pair } (i, j) \text{ is maximal, then } b_{i,i} = b_{j,j} = b_{i,j}.$$

This claim and (2.3) imply that all entries of the maximal monomial, say $q^a t^b$, in (2.1) occur with the same sign $-(-1)^b$. Hence $\langle N, F \rangle \neq 0$.

To prove (*) we first compute $a_{i,j}$ for any i, j (not necessarily maximal). Let ξ_i be the loop obtained by moving from d to z_i along F then back to d along N . Let a_i be the total winding number of ξ_i around all n punctures. Let ξ be the loop obtained by moving from d to d' along N , and then moving clockwise along ∂D back to d . Let a be the total winding number of ξ . We claim that

$$(2.4) \quad a_{i,j} = a_i + a_j + a.$$

Indeed, if $b_{i,j}$ is even then the paths $\alpha_0\beta_i\gamma_{i,j}$ and $\alpha'_0\beta'_j\gamma'_{i,j}$ (in the notation of Lemma 2.4) are loops and $a_{i,j}$ is the sum of their total winding numbers. These loops are homotopic in D_n to ξ_i and $\xi_j\xi$, respectively. This implies (2.4). If $b_{i,j}$ is odd then $a_{i,j}$ is the total winding number of the loop $\alpha_0\beta_i\gamma_{i,j}\alpha'_0\beta'_j\gamma'_{i,j}$. This loop is homotopic in D_n to $\xi_i\xi\xi_j$ which implies (2.4) in this case.

Suppose now that the pair (i, j) is maximal. Then $a_{i,j}$ is maximal among all the integers $a_{k,l}$. By (2.4), it follows that $a_i = a_j$ is maximal among all the integers a_k . (Although we shall not need it, observe that then $a_{i,i} = a_{j,j} = a_{i,j}$.)

We now show that $b_{i,i} = b_{i,j}$. By the maximality of (i, j) , we have that $b_{i,i} \leq b_{i,j}$. Suppose, seeking a contradiction, that $b_{i,i} < b_{i,j}$. Let α be an embedded arc from z'_i to z'_j along T' . Let β be an embedded arc from z'_j to z'_i along N .

If β does not pass through the point z_i , then we denote by w the winding number of the loop $\alpha\beta$ around z_i . Observe that $b_{i,j} - b_{i,i} = 2w$. To see this, consider the loop $\delta_{i,j} = \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_j\}\{\gamma_{i,j}, \gamma'_{i,j}\}$ appearing in (2.2). Clearly, $\beta'_j \sim \beta'_i\alpha$, where \sim denotes homotopy of paths in $D_n \setminus z_i$ constant on the endpoints. The assumption that β does not pass through z_i implies that $\gamma_{i,j} = \gamma_{i,i}$ and $\gamma'_{i,j} \sim \beta\gamma'_{i,i}$. Then

$$\delta_{i,j} = \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_j\}\{\gamma_{i,j}, \gamma'_{i,j}\} \sim \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_i\}\{z_i, \alpha\beta\}\{\gamma_{i,i}, \gamma'_{i,i}\}$$

where z_i denotes the constant path in z_i . This implies that $b_{i,j} - b_{i,i} = 2w$.

If β passes through z_i , we first modify β in a small neighborhood of z_i so that z_i lies to its left. Let w be the winding number of the loop $\alpha\beta$ around z_i . A little more difficult but similar argument shows that $b_{i,j} - b_{i,i} = 2w - 1$. In either case $w > 0$.

Let $D_0 = D \setminus z_i$ and $p : \hat{D}_0 \rightarrow D_0$ be the universal (infinite cyclic) covering. Let $\hat{\alpha}$ be a lift of α to \hat{D}_0 . Let $\hat{\beta}$ be the lift of β to \hat{D}_0 which starts at $\hat{\beta}(1)$. Consider a small neighborhood $V \subset D$ of the short arc in N connecting z_i to z'_i such that V meets $\alpha\beta$ only at z'_i . Let γ be a generic loop in $V \setminus z_i$ based at z'_i which winds w times around z_i in the clockwise direction. Let $\hat{\gamma}$ be the lift of γ to \hat{D}_0 beginning at $\hat{\beta}(1)$ and ending at $\hat{\alpha}(0)$. We can assume that $\hat{\gamma}$ is an embedded arc meeting $\hat{\alpha}\hat{\beta}$ only at

the endpoints. Let \hat{z}'_k be the first point of $\hat{\alpha}$ which lies also on $\hat{\beta}$ (possibly $\hat{z}'_k = \hat{\alpha}(1)$). Then $p(\hat{z}'_k) = z'_k$ for some $k = 1, \dots, n$. Let $\hat{\alpha}'$ be the initial segment of $\hat{\alpha}$ ending at \hat{z}'_k . Let $\hat{\beta}'$ be the final segment of $\hat{\beta}$ starting at \hat{z}'_k . Set $\hat{\delta} = \hat{\alpha}'\hat{\beta}'\hat{\gamma}$. It is clear that $\hat{\delta}$ is a simple closed curve in \hat{D}_0 . By the Jordan curve theorem it bounds a disc, $B \subset \hat{D}_0$, which lies either on the left or on the right of $\hat{\gamma}$. Since γ passes clockwise around z_i , the component of $D_0 \setminus \text{Im}(\gamma)$ adjacent to z_i lies on the right of γ . The set $p(B) \subset D_0$ being compact can not pass non-trivially over this component. Hence B lies on the left of $\hat{\gamma}$. Therefore $\hat{\delta}$ passes counterclockwise around B .

The number $a_k - a_i$ is equal to the total winding number of the loop $p(\hat{\alpha}')p(\hat{\beta}')$ in D_n around the punctures x_1, \dots, x_n . Since γ is contractible in D_n , the loop $p(\hat{\alpha}')p(\hat{\beta}')$ is homotopic in D_n to $p(\hat{\alpha}')p(\hat{\beta}')\gamma = p(\hat{\delta})$. Therefore $a_k - a_i$ is equal to the number of points in $B \cap p^{-1}(X)$. Since a_i is maximal, we must have $a_k = a_i$ and $B \cap p^{-1}(X) = \emptyset$, so that $p(B) \subset D_n \setminus z_i$. Then we can isotop T so as to have fewer points of intersection with N . To see this we shall construct a bigon for the pair (N, T) . If B meets $p^{-1}(N \cup T)$ only along $\hat{\alpha}'\hat{\beta}'$ then the projection $p|_{\text{Int}(B)} : \text{Int}(B) \rightarrow D_n \setminus z_i$ must be injective. It follows that $w = 1$ and the union of $p(B)$ with the small disc bounded by γ in V is a bigon for (N, T) . Assume that $\text{Int}(B) \cap p^{-1}(N \cup T) \neq \emptyset$. Note that $p^{-1}(N)$ (resp. $p^{-1}(T)$) is an embedded one-manifold in \hat{D}_0 whose components are non-trivially permuted by any covering transformation. If $p^{-1}(N)$ intersects $\text{Int}(B)$ then this intersection consists of a finite set of disjoint arcs with endpoints on $\hat{\alpha}'$. At least one of these arcs bounds together with a subarc of $\hat{\alpha}'$ a disc, $B_0 \subset B$, whose interior does not meet $p^{-1}(N)$. If $p^{-1}(N)$ does not meet $\text{Int}(B)$ then we set $B_0 = B$. Applying the same construction to the intersection of B_0 with $p^{-1}(T)$ we obtain a bigon $B_{00} \subset B_0$ for the pair $(p^{-1}(N), p^{-1}(T))$. The restriction of p to B_{00} is injective and yields a bigon for (N, T) . Hence the intersection $N \cap T$ is not minimal. This contradicts our choice of N, T . Therefore, the assumption $b_{i,i} < b_{i,j}$ must have been false. So, $b_{i,i} = b_{i,j}$. Similarly, $b_{j,j} = b_{i,j}$. This completes the proof of (*) and of Lemma 2.5.

2.6. Lemma. – *If a self-homeomorphism h of D_n represents an element of $\text{Ker}(B_n \rightarrow \text{Aut}(H_2(\tilde{C})))$ then for any noodle N and any fork F , we have $\langle N, h(F) \rangle = \langle N, F \rangle$.*

Proof. – Let $\{U_i \subset \text{Int}(D)\}_{i=1}^m$ be disjoint closed disc neighborhoods of the points $\{x_i\}_{i=1}^m$, respectively. Let U be the set of points $\{x, y\} \in C$ such that at least one of x, y lies in $\cup_{i=1}^m U_i$. Let $\tilde{U} \subset \tilde{C}$ be the preimage of U under the covering map $\tilde{C} \rightarrow C$. Observe that the surface $\tilde{\Sigma}_F$ is an open square such that for a sufficiently

big concentric closed subsquare $S \subset \tilde{\Sigma}_F$ we have $\tilde{\Sigma}_F \setminus S \subset \tilde{U}$. Hence $\tilde{\Sigma}_F$ represents a relative homology class $[\tilde{\Sigma}_F] \in H_2(\tilde{C}, \tilde{U})$. The boundary homomorphism $H_2(\tilde{C}, \tilde{U}) \rightarrow H_1(\tilde{U})$ maps $[\tilde{\Sigma}_F]$ into $[\partial S] \in H_1(\tilde{U})$. A direct computation in $\pi_1(U)$ (see [Bi2]) shows that $(q - 1)^2(qt + 1)[\partial S] = 0$. Therefore $(q - 1)^2(qt + 1)[\tilde{\Sigma}_F] = j(v_F)$ where j is the inclusion homomorphism $H_2(\tilde{C}) \rightarrow H_2(\tilde{C}, \tilde{U})$ and $v_F \in H_2(\tilde{C})$. Deforming if necessary N , we can assume that $N \cap (\cup_i U_i) = \emptyset$. Then $\tilde{\Sigma}_N \cap \tilde{U} = \emptyset$ and therefore

$$(q - 1)^2(qt + 1)\langle N, F \rangle = \sum_{a,b \in \mathbf{Z}} (q^a t^b \tilde{\Sigma}_N \cdot v_F) q^a t^b$$

where $q^a t^b \tilde{\Sigma}_N \cdot v_F$ is the (well-defined) algebraic intersection number between a properly embedded surface and a 2-dimensional homology class. (This number does not depend on the choice of v_F as above).

Any self-homeomorphism h of D_n is isotopic to a self-homeomorphism of D_n preserving the set U . Therefore $v_{h(F)} = \tilde{h}_*(v_F)$. If $\tilde{h}_* = \text{id}$, then

$$q^a t^b \tilde{\Sigma}_N \cdot v_F = q^a t^b \tilde{\Sigma}_N \cdot \tilde{h}_*(v_F) = q^a t^b \tilde{\Sigma}_N \cdot v_{h(F)}.$$

This implies that $(q - 1)^2(qt + 1)\langle N, F \rangle = (q - 1)^2(qt + 1)\langle N, h(F) \rangle$ and therefore $\langle N, F \rangle = \langle N, h(F) \rangle$.

2.7. Deduction of Theorem 2.2 from the lemmas. – We shall prove that a self-homeomorphism h of D_n representing an element of $\text{Ker}(B_n \rightarrow \text{Aut}(H_2(\tilde{C})))$ is isotopic to the identity map $\text{rel } \partial D_n$. We begin with the following assertion.

(**) *An embedded arc T in D_n with endpoints in (distinct) punctures can be isotopped off a noodle N if and only if $h(T)$ can be isotopped off N .*

Indeed, we can extend T to a fork F so that T is the tines of F . By Lemma 2.6, $\langle N, h(F) \rangle = 0$ if and only if $\langle N, F \rangle = 0$. Now Lemma 2.5 implies (**).

We shall apply (**) to the following arcs and noodles. For $i = 1, \dots, n - 1$, denote by T_i the embedded arc $[x_i, x_{i+1}] \subset (-1, +1) \subset D$ and denote by N_i the i -th “elementary” noodle obtained by rushing from d towards x_i , encircling x_i in the clockwise direction and then moving straight to d' . It is clear that T_i can be isotopped off N_j if and only if $j \neq i, i + 1$. This and (**) imply that h induces the identity permutation on the punctures of D_n .

Since T_1 is disjoint from N_3 , we can isotop $h \text{ rel } \partial D_n$ so that $h(T_1)$ is disjoint from N_3 . Similarly, $h(T_1)$ can be made disjoint from N_4 . As it was explained in

Sect. 2.2, this can be done by a sequence of isotopies eliminating bigons for the pair $(N_4, h(T_1))$. Since N_4 and $h(T_1)$ do not meet N_3 , neither do these bigons. Hence our isotopies do not create intersections of $h(T_1)$ with N_3 . Repeating this argument, we can assume that $h(T_1)$ is disjoint from all N_i with $i = 3, 4, \dots, n - 1$. By applying one final isotopy we can make $h(T_1) = T_1$. Applying the same procedure to T_2 we can ensure that $h(T_2) = T_2$ while keeping $h(T_1) = T_1$. Continuing in this way, we can assume that $h(T_i) = T_i$ for all $i = 1, \dots, n - 1$. Such a homeomorphism h is isotopic to a k -th power ($k \in \mathbf{Z}$) of the Dehn twist about a circle in $\text{Int}(D_n)$ going very closely to ∂D_n . This Dehn twist acts on $H_2(\tilde{C})$ by multiplication by $q^{2n}t^2$. Since by assumption h acts trivially on $H_2(\tilde{C})$, we must have $k = 0$ so that h is isotopic to the identity rel ∂D_n .

2.8. Remarks. – The proof of Lemma 2.6 shows that each fork F determines (a priori non-uniquely) a certain homology class $v_F \in H_2(\tilde{C})$. It follows from the computations in [Bi2] that this class is in fact well-determined by F . Thus, the forks yield a nice geometric way of representing elements of $H_2(\tilde{C})$ (this was implicit in [Kr1]). For instance, for any $1 \leq i < j \leq n$ we can consider the fork consisting of three linear segments connecting the point $-\sqrt{-1}/2$ to d, x_i, x_j . The corresponding classes $\{v_{i,j} \in H_2(\tilde{C})\}_{i,j}$ form a basis of the free R -module $H_2(\tilde{C})$. The action of the braid generators $\sigma_1, \dots, \sigma_{n-1}$ on this basis can be described by explicit formulas (see [Bi2], cf. Sect. 3.1).

3. THE WORK OF KRAMMER

3.1. A representation of B_n . Following [Kr2], we denote by $\text{Ref} = \text{Ref}_n$ the set of pairs of integers (i, j) such that $1 \leq i < j \leq n$. Clearly, $\text{card}(\text{Ref}) = n(n - 1)/2$.

Let R be a commutative ring with unit and $q, t \in R$ be two invertible elements. Let $V = \bigoplus_{s \in \text{Ref}} Rx_s$ be the free R -module of rank $n(n - 1)/2$ with basis $\{x_s\}_{s \in \text{Ref}}$. Krammer [Kr2] defines an R -linear action of B_n on V by

$$\sigma_k(x_{i,j}) = \begin{cases} x_{i,j} & \text{if } k < i - 1 \text{ or } j < k, \\ x_{i-1,j} + (1 - q)x_{i,j} & \text{if } k = i - 1, \\ tq(q - 1)x_{i,i+1} + qx_{i+1,j} & \text{if } k = i < j - 1, \\ tq^2x_{i,j} & \text{if } k = i = j - 1, \\ x_{i,j} + tq^{k-i}(q - 1)^2x_{k,k+1} & \text{if } i < k < j - 1, \\ x_{i,j-1} + tq^{j-i}(q - 1)x_{j-1,j} & \text{if } i < k = j - 1, \\ (1 - q)x_{i,j} + qx_{i,j+1} & \text{if } k = j, \end{cases}$$

where $1 \leq i < j \leq n$ and $k = 1, \dots, n - 1$. That the action of σ_k is invertible and

that relations (0.1), (0.2) are satisfied should be verified by a direct computation. For $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$, this representation is equivalent to the one considered in Sect. 2. In terms of the basis $\{v_{i,j} \in H_2(\tilde{C})\}_{i,j}$ mentioned in Sect. 2.8, the equivalence is given by

$$v_{i,j} = x_{i,j} + (1 - q) \sum_{i < k < j} x_{k,j}, \quad x_{i,j} = v_{i,j} + (q - 1) \sum_{i < k < j} q^{k-1-i} v_{k,j}.$$

3.2. Theorem (D. Krammer [Kr2]). – *Let $R = \mathbf{R}[t^{\pm 1}]$, $q \in \mathbf{R}$, and $0 < q < 1$. Then the representation $B_n \rightarrow \text{Aut}(V)$ defined in Sect. 3.1 is faithful for all $n \geq 1$.*

This Theorem implies Theorem 2.2: if a representation over $\mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ becomes faithful after assigning a real value to q , then it is faithful itself.

Below we outline the main ideas of Krammer’s proof.

3.3. Positive braids and the set $\Omega \subset B_n$. We recall a few facts about the braid group B_n , see [Ch], [Ga], [Mi]. For $i = 1, 2, \dots, n - 1$ denote by s_i the transposition $(i, i + 1) \in S_n$. The set $\{s_1, \dots, s_{n-1}\}$ generates the symmetric group S_n . Let $|| : S_n \rightarrow \mathbf{Z}$ be the length function with respect to this generating set: for $x \in S_n$, $|x|$ is the smallest natural number k such that x is a product of k elements of the set $\{s_1, \dots, s_{n-1}\}$. The canonical projection $B_n \rightarrow S_n$ has a unique set-theoretic section $r : S_n \rightarrow B_n$ such that $r(s_i) = \sigma_i$ for $i = 1, \dots, n - 1$ and $r(xy) = r(x)r(y)$ whenever $|xy| = |x| + |y|$. The group B_n admits a presentation by generators $\{r(x) \mid x \in S_n\}$ and relations $r(xy) = r(x)r(y)$ for all $x, y \in S_n$ such that $|xy| = |x| + |y|$. Set $\Omega = r(S_n) \subset B_n$. Note that $1 = r(1) \in \Omega$ and $\sigma_i = r(s_i) \in \Omega$ for all i .

The *positive braid monoid* B_n^+ is the submonoid of B_n generated by $\{\sigma_1, \dots, \sigma_{n-1}\}$. The elements of B_n^+ are called *positive braids*. Clearly, $\Omega \subset B_n^+$.

For any $x \in B_n^+$ there is a unique longest $x' \in S_n$ such that $x \in r(x')B_n^+$. We denote $r(x') \in \Omega$ by $LF(x)$ where LF stands for the leftmost factor. Observe that

$$(3.1) \quad LF(xy) = LF(x LF(y))$$

for any $x, y \in B_n^+$. This implies that the map $B_n^+ \times \Omega \rightarrow \Omega$ defined by $(x, y) \mapsto LF(xy)$ is an action of the monoid B_n^+ on Ω .

3.4. Half-permutations. A set $A \subset \text{Ref}$ is called a *half-permutation* if whenever $(i, j), (j, k) \in A$, one has $(i, k) \in A$. Each half-permutation A determines an ordering $<_A$ on the set $\{1, 2, \dots, n\}$ by $i <_A j \Leftrightarrow (i, j) \in A$ (and vice versa).

To state deeper properties of half-permutations we consider the set 2^{Ref} of all subsets of Ref and define a map $L : S_n \rightarrow 2^{\text{Ref}}$ by

$$L(x) = \{(i, j) \mid 1 \leq i < j \leq n, x^{-1}(i) > x^{-1}(j)\} \subset \text{Ref}.$$

Note that the set $L(x)$ is a half-permutation and $\text{card}(L(x)) = |x|$. It is obvious that the map $L : S_n \rightarrow 2^{\text{Ref}}$ is injective.

The key property of half-permutations is the following assertion ([Kr2, Lemma 4.3]): for every half-permutation $A \subset \text{Ref}$ there is a greatest set $A' \subset A$ (with respect to inclusion) such that $A' = L(x)$ for a certain $x \in S_n$. The corresponding braid $r(x) \in \Omega$ is denoted by $GB(A)$ where GB stands for the greatest braid. This defines a map GB from the set of half-permutations to Ω . In particular, for any $x \in S_n$ we have $GB(L(x)) = r(x)$.

3.5. Actions of B_n of 2^{Ref} and on half-permutations. Let $R = \mathbf{R}[t^{\pm 1}]$, $q \in \mathbf{R}$, and $0 < q < 1$. The action of B_n on V defined in Sect. 3.1 has the following property: for any positive braid $x \in B_n^+$ the entries of the matrix of the map $v \mapsto xv : V \rightarrow V$ belong to $\mathbf{R}_{\geq 0} + t\mathbf{R}[t]$. (This is obvious for the generators $\sigma_1, \dots, \sigma_{n-1}$ of B_n^+). Therefore the action of B_n^+ preserves the set

$$W = \bigoplus_{s \in \text{Ref}} (\mathbf{R}_{\geq 0} + t\mathbf{R}[t]) x_s \subset V.$$

For a set $A \subset \text{Ref}$, define W_A as the subset of W consisting of vectors $\sum_{s \in \text{Ref}} k_s x_s$ with $k_s \in \mathbf{R}_{\geq 0} + t\mathbf{R}[t]$ such that $k_s \in t\mathbf{R}[t] \Leftrightarrow s \in A$. Clearly, W is the disjoint union of the sets W_A corresponding to various $A \subset \text{Ref}$. For any $x \in B_n^+$ and $A \subset \text{Ref}$ there is a unique $B \subset \text{Ref}$ such that $xW_A \subset W_B$. We denote this set B by xA . This defines an action of B_n^+ on 2^{Ref} . By [Kr2, Lemma 4.2], this action maps half-permutations to half-permutations. This defines an action of B_n^+ on the set of half-permutations. Finally, Krammer observes that the map GB from this set to Ω is B_n^+ -equivariant. Thus, for any positive braid $x \in B_n^+$ and a half-permutation $A \subset \text{Ref}$, we have

$$(3.2) \quad GB(xA) = LF(xGB(A)).$$

The rest of the argument is contained in the following two lemmas.

3.6. Lemma. – *Let B_n act on a set U . Suppose we are given non-empty disjoint sets $\{C_x \subset U\}_{x \in \Omega}$ such that $xC_y \subset C_{LF(xy)}$ for all $x, y \in \Omega$. Then the action of B_n on U is faithful.*

Proof. – Denote by ℓ the group homomorphism $B_n \rightarrow \mathbf{Z}$ mapping $\sigma_1, \dots, \sigma_{n-1}$ to 1. We check first that the inclusion $xC_y \subset C_{LF(xy)}$ holds for all $x \in B_n^+$ and $y \in \Omega$. Clearly, $\ell(B_n^+) \geq 0$ so that we can use induction on $\ell(x)$. If $\ell(x) = 0$ then $x = 1$ and the inclusion follows from the equality $LF(y) = y$. Let $\ell(x) \geq 1$. Then $x = \sigma_i v$ where $i = 1, \dots, n-1$ and $v \in B_n^+$. Clearly, $\ell(v) = \ell(x) - 1$. We have

$$xC_y = \sigma_i v C_y \subset \sigma_i (C_{LF(vy)}) \subset C_{LF(\sigma_i LF(vy))} = C_{LF(\sigma_i v y)} = C_{LF(xy)}.$$

Here the first inclusion follows from the induction hypothesis, the second inclusion follows from the assumptions of the lemma, the middle equality follows from (3.1).

It is known that for any $b \in B_n$ there are $x, y \in B_n^+$ such that $b = xy^{-1}$. Therefore to prove the lemma it suffices to show that, if two elements $x, y \in B_n^+$ act in the same way on U , then $x = y$. We will show this by induction on $\ell(x) + \ell(y)$. If $\ell(x) + \ell(y) = 0$, then $x = y = 1$. Assume that $\ell(x) + \ell(y) > 0$. By assumption, C_1 is non-empty; choose any $u \in C_1$. We have $xu \in xC_1 \subset C_{LF(x)}$ and similarly $yu \in C_{LF(y)}$. Hence $xu = yu \in C_{LF(x)} \cap C_{LF(y)}$. By the disjointness assumption, this is possible only if $LF(x) = LF(y)$. Write $z = LF(x) \in \Omega$ and consider $x', y' \in B_n^+$ such that $x = zx'$ and $y = zy'$. We have $z \neq 1$ since otherwise $x = y = 1$. Then $\ell(z) > 0$ and $\ell(x') + \ell(y') < \ell(x) + \ell(y)$. The induction assumption yields that $x' = y'$. Therefore $x = y$. This proves the inductive step and the lemma.

3.7. Lemma. – For $x \in \Omega$, set

$$C_x = \bigcup_{A \in GB^{-1}(x)} W_A \subset V$$

where A runs over all half-permutations such that $GB(A) = x$. Then the sets $\{C_x\}_{x \in \Omega}$ satisfy all the conditions of Lemma 3.6. Therefore the action of B_n on V is faithful.

Proof. – It is obvious that the sets $\{C_x\}_{x \in \Omega}$ are disjoint. The set C_x is non-empty because $\emptyset \neq W_{L(r^{-1}(x))} \subset C_x$.

To prove the inclusion $xC_y \subset C_{LF(xy)}$ for $x, y \in \Omega$, it suffices to prove that $xW_A \subset C_{LF(xy)}$ whenever A is a half-permutation such that $GB(A) = y$. This follows from the inclusions

$$xW_A \subset W_{xA} \subset C_{GB(xA)} = C_{LF(xy)}.$$

Here the first inclusion follows from the definition of the B_n^+ -action of Ref. The second inclusion follows from the definition of $C_{GB(xA)}$. The equality follows from (3.2).

3.8. More about the representation. – Let $\ell_\Omega : B_n \rightarrow \mathbf{Z}$ be the length function with respect to the generating set Ω : for $x \in B_n$, $\ell_\Omega(x)$ is the minimal natural number k such that $x = x_1 \dots x_k$ where $x_i \in \Omega$ or $x_i^{-1} \in \Omega$ for each $i = 1, \dots, k$. Among other related results, Krammer gives an explicit computation of ℓ_Ω in terms of his representation. Namely, take the Laurent polynomial ring $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ as the ground ring and denote by ρ Krammer's representation $B_n \rightarrow \text{Aut}(V)$ defined in Sect. 3.1. For $x \in B_n$, consider the Laurent expansion $\rho(x) = A_k t^k + A_{k+1} t^{k+1} + \dots + A_l t^l$ where $\{A_i\}_{i=k}^l$ are $(m \times m)$ -matrices over $\mathbf{Z}[q^{\pm 1}]$ and $A_k \neq 0, A_l \neq 0$. Then

$$\ell_\Omega(x) = \max(l - k, l, -k).$$

This formula yields another proof of the faithfulness of ρ : If $x \in \text{Ker} \rho$, then $k = l = 0$ so that $\ell_\Omega(x) = 0$ and $x = 1$.

The length function ℓ_Ω was first considered by Charney [Ch] who proved that the formal power series $\sum_{x \in B_n} z^{\ell_\Omega(x)} \in \mathbf{Z}[[z]]$ is a rational function. It is unknown whether the similar formal power series determined by the length function with respect to the generators $\sigma_1, \dots, \sigma_{n-1}$ is rational.

4. BMW-ALGEBRAS AND REPRESENTATIONS OF B_n

4.1. Hecke algebras and representations of B_n . A vast family of representations of B_n including the Burau representation arise from a study of Hecke algebras. The Hecke algebra $H_n(\alpha)$ corresponding to $\alpha \in \mathbf{C}$ can be defined as the quotient of the complex group ring $\mathbf{C}[B_n]$ by the relations $\sigma_i^2 = (1 - \alpha)\sigma_i + \alpha$ for $i = 1, \dots, n - 1$. This family of finite dimensional \mathbf{C} -algebras is a one-parameter deformation of $\mathbf{C}[S_n] = H_n(1)$. For α sufficiently close to 1, the algebra $H_n(\alpha)$ is isomorphic to $\mathbf{C}[S_n]$. For such α , the algebra $H_n(\alpha)$ is semisimple and its irreducible representations are indexed by the Young diagrams with n boxes. Their decomposition rules and dimensions are the same as for the irreducible representations of S_n .

Each representation of $H_n(\alpha)$ yields a representation of B_n via the natural projection $B_n \subset \mathbf{C}[B_n] \rightarrow H_n(\alpha)$. This gives a family of irreducible finite dimensional representations of B_n indexed by the Young diagrams with n boxes. These representations were extensively studied by V. Jones [Jo]. In particular, he observed that the $(n - 1)$ -dimensional Burau representation of B_n appears as the irreducible representation associated with the two-column Young diagram whose columns have $n - 1$ boxes and one box, respectively.

4.2. Birman-Murakami-Wenzl algebras and their representations. Jun Murakami [Mu] and independently J. Birman and H. Wenzl [BW] introduced a two-parameter family of finite dimensional \mathbf{C} -algebras $C_n(\alpha, l)$ where α and l are non-zero complex numbers such that $\alpha^4 \neq 1$ and $l^4 \neq 1$. For $i = 1, \dots, n-1$, set

$$e_i = (\alpha + \alpha^{-1})^{-1}(\sigma_i + \sigma_i^{-1}) - 1 \in \mathbf{C}[B_n].$$

The algebra $C_n(\alpha, l)$ is the quotient of $\mathbf{C}[B_n]$ by the relations

$$e_i \sigma_i = l^{-1} e_i, \quad e_i \sigma_{i-1}^{\pm 1} e_i = l^{\pm 1} e_i$$

for all i . (The original definition in [BW] involves more relations; for the shorter list given above, see [We].) The algebra $C_n(\alpha, l)$ admits a geometric interpretation in terms of so-called Kauffman skein classes of tangles in Euclidean 3-space. This family of algebras is a deformation of an algebra introduced by R. Brauer [Br] in 1937.

The algebraic structure and representations of $C_n(\alpha, l)$ were studied by Wenzl [We], who established (among other results) the following three facts.

(i) *For generic α, l , the algebra $C_n(\alpha, l)$ is semisimple.*

Here “generic” means that α is not a root of unity and $\sqrt{-1}l$ is not an integer power of $-\sqrt{-1}\alpha$. (The latter two numbers correspond to r and q in Wenzl’s notation). In the sequel we assume that α, l are generic in this sense. We denote the number of boxes in a Young diagram λ by $|\lambda|$.

(ii) *The irreducible finite dimensional $C_n(\alpha, l)$ -modules are indexed by the Young diagrams λ such that $|\lambda| \leq n$ and $|\lambda| \equiv n \pmod{2}$.*

The irreducible $C_n(\alpha, l)$ -module corresponding to λ will be denoted by $V_{n,\lambda}$. Composing the natural projection $B_n \subset \mathbf{C}[B_n] \rightarrow C_n(\alpha, l)$ with the action of $C_n(\alpha, l)$ on $V_{n,\lambda}$ we obtain an irreducible representation of B_n .

Observe that the inclusion $B_{n-1} \hookrightarrow B_n$ sending each $\sigma_i \in B_{n-1}$ with $i = 1, \dots, n-2$ to $\sigma_i \in B_n$ induces an inclusion $C_{n-1}(\alpha, l) \hookrightarrow C_n(\alpha, l)$ for all $n \geq 2$.

(iii) *The $C_n(\alpha, l)$ -module $V_{n,\lambda}$ decomposes as a $C_{n-1}(\alpha, l)$ -module into a direct sum $\bigoplus_{\mu} V_{n-1,\mu}$ where μ ranges over all Young diagrams obtained by removing or (if $|\lambda| < n$) adding one box to λ . Each such μ appears in this decomposition with multiplicity 1.*

4.3. The Bratelli diagram for the BMW-algebras. The assertions (ii) and (iii) in Sect. 4.2 allow us to draw the Bratelli diagram for the sequence $C_1(\alpha, l) \subset C_2(\alpha, l) \subset \dots$. On the level $n = 1, 2, \dots$ of the Bratelli diagram one puts all Young diagrams λ such that $|\lambda| \leq n$ and $|\lambda| \equiv n \pmod{2}$. Then one connects by an edge each λ on the n -th level to all Young diagrams on the $(n - 1)$ -th level obtained by removing or (if $|\lambda| < n$) adding one box to λ . For instance, the $n = 1$ level consists of a single Young diagram with one box corresponding to the tautological one-dimensional representation of $C_1(\alpha, l) = \mathbf{C}$. The $n = 2$ level contains the empty Young diagram and two Young diagrams with two boxes. All three are connected by an edge to the diagram on the level 1. Note that every Young diagram λ appears on the levels $|\lambda|, |\lambda| + 2, |\lambda| + 4, \dots$

The Bratelli diagram yields a useful method of computing the dimension of $V_{n, \lambda}$ where λ is a Young diagram on the n -th level. It is clear from (iii) that $\dim(V_{n, \lambda})$ is the number of paths on the Bratelli diagram leading from λ to the only diagram on the level 1. Here by a path we mean a path with vertices lying on consecutively decreasing levels. We give three examples of computations based on (iii).

(a) Let λ_n be the Young diagram represented by a column of n boxes. There is only one path from λ_n , positioned on the level n , to the top of the Bratelli diagram. Hence, $\dim(V_{n, \lambda_n}) = 1$ for all $n \geq 1$.

For $n \geq 2$, the algebra $C_n(\alpha, l)$ has two one-dimensional representations. In both of them all e_i act as 0 and all σ_i act as multiplication by one and the same number equal either to α or to α^{-1} . We choose the correspondence between the irreducible $C_n(\alpha, l)$ -modules and the Young diagrams so that all σ_i act on V_{n, λ_n} as multiplication by α . If λ_n^T is the Young diagram represented by a row of n boxes, then similarly to (a) we have that $\dim(V_{n, \lambda_n^T}) = 1$ and all σ_i act on V_{n, λ_n^T} as multiplication by α^{-1} .

(b) For $n \geq 2$, let λ'_n be the two-column Young diagram whose columns have $n - 1$ boxes and one box, respectively. For $n \geq 3$, the diagram λ'_n , positioned on the level n , is connected to only two Young diagrams on the previous level, namely, to λ'_{n-1} and λ_{n-1} . Hence

$$\dim(V_{n, \lambda'_n}) = \dim(V_{n-1, \lambda'_{n-1}}) + \dim(V_{n-1, \lambda_{n-1}}) = \dim(V_{n-1, \lambda'_{n-1}}) + 1.$$

We have $\lambda'_2 = \lambda_2^T$ so that $\dim(V_{2, \lambda'_2}) = 1$. Hence $\dim(V_{n, \lambda'_n}) = n - 1$ for all $n \geq 2$.

(c) For $n \geq 2$, consider the module $V_{n, \lambda_{n-2}}$ corresponding to the Young diagram λ_{n-2} positioned on the level n . If $n \geq 3$, then this diagram is connected to three Young diagrams on the previous level, namely, to $\lambda_{n-1}, \lambda_{n-3}, \lambda'_{n-1}$. Hence

$$\dim(V_{n, \lambda_{n-2}}) = \dim(V_{n-1, \lambda_{n-1}}) + \dim(V_{n-1, \lambda_{n-3}}) + \dim(V_{n-1, \lambda'_{n-1}})$$

$$= \dim(V_{n-1, \lambda_{n-3}}) + n - 1.$$

We gave $\lambda_0 = \emptyset$ and by (iii) above, $\dim(V_{2, \lambda_0}) = \dim(V_{1, \lambda_1}) = 1$. Thus for all $n \geq 2$,

$$\dim(V_{n, \lambda_{n-2}}) = n(n-1)/2,$$

i.e., $V_{n, \lambda_{n-2}}$ has the same dimension as the Krammer representation of B_n . We now rescale the representation $B_n \rightarrow \text{Aut}(V_{n, \lambda_{n-2}})$ by dividing the action of each σ_i by α .

4.4. Theorem (M. Zinno [Zi]). – *The Krammer representation corresponding to $q = -\alpha^{-2}$ and $t = \alpha^3 l^{-1}$ is isomorphic to the rescaled representation $B_n \rightarrow \text{Aut}(V_{n, \lambda_{n-2}})$.*

The proof given in [Zi] goes by a direct comparison of both actions of B_n on certain bases. Theorem 4.4 implies that the Krammer-Bigelow representation considered in Sect. 2 and 3 is irreducible.

BIBLIOGRAPHY

- [Bi1] S. BIGELOW – *The Burau representation is not faithful for $n \geq 5$* , Preprint math. GT/9904100.
- [Bi2] S. BIGELOW – *Braid groups are linear*, Preprint math. GR/0005038.
- [Bir] J. S. BIRMAN – *Braids, links, and mapping class groups*, Ann. of Math. Stud., vol. 82, Princeton Univ. Press, Princeton, N.J., 1974.
- [BW] J. S. BIRMAN, H. WENZL – *Braids, link polynomials and a new algebra*, Trans. Amer. Math. Soc. **313** (1989), 249–273.
- [Br] R. BRAUER – *On algebras which are connected with the semisimple continuous groups*, Ann. of Math. **38** (1937), 857–872.
- [Bu] W. BURAU – *Über Zopfgruppen und gleichsinnig verdrillte Verkettungen*, Abh. Math. Semin. Hamburg. Univ. **11** (1935), 179–186.
- [Ch] R. CHARNEY – *Geodesic automation and growth functions for Artin groups of finite type*, Math. Ann. **301** (1995), 307–324.
- [DFG] J. L. DYER, E. FORMANEK, E. K. GROSSMAN – *On the linearity of automorphism groups of free groups*, Arch. Math. (Basel) **38** (1982), 404–409.
- [FLP] A. FATHI, F. LAUDENBACH, V. POENARU – *Travaux de Thurston sur les surfaces*, Astérisque **66–67** (1991), Soc. Math. France, Paris.
- [FP] E. FORMANEK, C. PROCESI – *The automorphism group of a free group is not linear*, J. Algebra **149** (1992), 494–499.

- [Ga] F. A. GARSIDE – *The braid group and other groups*, Quart. J. Math. Oxford **20** (1969), 235–254.
- [Iv1] N. V. IVANOV – *Automorphisms of Teichmüller modular groups*, Topology and geometry – Rohlin Seminar, Lecture Notes in Math., vol. 1346, Springer, Berlin-New York (1988), 199–270.
- [Iv2] N. V. IVANOV – *Subgroups of Teichmüller modular groups*, Transl. of Math. Monographs, vol. 115. Amer. Math. Soc., Providence, RI, 1992.
- [Jo] V. F. R. JONES – *Hecke algebra representations of braid groups and link polynomials*, Ann. of Math. **126** (1987), 335–388.
- [Ka] C. KASSEL – *L'ordre de Dehornoy sur les tresses*, Séminaire Bourbaki, exposé 865 (novembre 1999), to appear in Astérisque, Soc. Math. France, Paris.
- [Kr1] D. KRAMMER – *The braid group B_4 is linear*, Preprint, Basel (1999).
- [Kr2] D. KRAMMER – *Braid Groups are linear*, Preprint, Basel (2000).
- [La] R. J. LAWRENCE – *Homological representations of the Hecke algebra*, Comm. Math. Phys. **135** (1990), 141–191.
- [LP] D. D. LONG, M. PATON – *The Burau representation is not faithful for $n \geq 6$* , Topology **32** (1993), 439–447.
- [Mi] J. MICHEL – *A note on words in braid monoids*, J. Algebra **215** (1999), 366–377.
- [Mo] J. A. MOODY – *The Burau representation of the braid group B_n is unfaithful for large n* , Bull. Amer. Math. Soc. (N.S.) **25** (1991), 379–384.
- [Mi] J. MURAKAMI – *The Kauffman polynomial of links and representation theory*, Osaka J. Math. **24** (1987), 745–758.
- [We] H. WENZL – *Quantum groups and subfactors of type B, C, and D*, Comm. Math. Phys. **133** (1990), 383–432.
- [Zi] M. G. ZINNO – *On Krammer's Representation of the Braid Group*, Preprint math. RT/0002136.

Vladimir TURAEV

Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.

7 rue Descartes

F-67084 STRASBOURG Cedex

E-mail : turaev@math.u-strasbg.fr

RANDOM MATRICES AND PERMUTATIONS, MATRIX INTEGRALS AND INTEGRABLE SYSTEMS

by **Pierre VAN MOERBEKE**

TABLE OF CONTENTS

0. Introduction	1
1. Largest increasing sequences in Random Permutations	3
1.1. Robinson-Schensted-Knuth correspondence and symmetric functions	3
1.2. Plancherel measure, Integrals over $U(n)$ and Toeplitz matrices	5
1.3. Virasoro constraints, Integrable systems and Painlevé V equation	7
1.4. Random permutations π_N for large N	9
2. The spectrum of Random Matrices	11
2.1. Virasoro constraints, Toda and Pfaff lattices and KP equations	12
2.2. The Gaussian ensemble: PDE's for the statistics of the spectrum	13
2.3. Infinite Hermitian matrix ensembles	13
3. Large random matrices and Permutations: a direct connection via enumerative geometry	16
4. Integrals, moment matrices and integrable systems	17
REFERENCES	18

0. INTRODUCTION

The purpose of this lecture is to give a survey of recent interactions between the theory of random matrices, the theory of random permutations and so-called integrable models. The latter are described either by integrals over tangent spaces to symmetric spaces, or by integrals over classical groups.

Going back in time, t'Hooft was led, in the 70's, to so-called matrix models in order to understand the behavior of quantum gauge theories with gauge group $SU(N)$ when N

The support of a National Science Foundation grant # DMS-98-4-50790, a Nato, a FNRS and a Francqui Foundation grant is gratefully acknowledged.

gets large. In their pioneering work, Bessis, Itzykson and Zuber [8] considered the integral $Z_N^{(\varepsilon)}$ taken over the space \mathcal{H}_N of Hermitian matrices of size N , whose log has the following expansion:

$$\log Z_N^{(\varepsilon)} = \log \frac{\int_{\mathcal{H}_N} dM e^{-\frac{1}{2}trM^2 - \frac{\varepsilon}{N}trM^4}}{\int_{\mathcal{H}_N} dM e^{-\frac{1}{2}trM^2}} = \sum_{g=0}^{\infty} N^{2-2g} \sum_{k=0}^{\infty} (-\varepsilon)^k W(g, k)$$

where $W(g, n)$ is the number of graphs with k vertices drawn on a surface of genus g , or in a dual language, the number of ways to cover a surface of genus g with k squares. Replacing $\frac{\varepsilon}{N}trM^4$ with $\sum_{i \geq 1} t_i trM^i$ leads to coverings of surfaces of genus g by n -gons of various n . The sequence of such integrals over N is a solution of the standard Toda lattice equations.

Two-matrix integrals (Chadha-Mahoux-Mehta [11], Itzykson-Zuber [20])

$$\int \int_{\mathcal{H}_N \times \mathcal{H}_N} dM_1 dM_2 e^{-tr(M_1^2 + M_2^2 - cM_1 M_2 + \alpha M_1^4 + \beta M_2^4)}$$

provide a solution to the 2d-Toda lattice. They describe an Ising model on a random lattice, with spin $\sigma = \pm 1$ at each site with nearest neighbor interaction; the interaction between two sites depends on whether the spins are equal or not.

Kontsevich [25] proves that $Z_N(t)$ is a τ -function for the KdV equation for N large, with expansion (set $t_i = -c_i trZ^{-i}$ for appropriate integer c_i 's)

$$\begin{aligned} \log Z_N(t) &= \log \frac{\int_{\mathcal{H}_N} dX e^{-\frac{1}{2}tr X^2 Z + \frac{1}{6}tr X^3}}{\int_{\mathcal{H}_N} dX e^{-\frac{1}{2}tr X^2 Z}} = \sum_{\Gamma \in G_N} \frac{(\frac{1}{2})^{\# \text{ vertices in } \Gamma}}{\# \text{ Aut } \Gamma} \prod_{\substack{\text{ribbons} \\ \alpha, \beta}} \frac{2}{z_\alpha + z_\beta} \\ &= \sum_{n \geq 1, d_1, \dots, d_n \geq 0} \frac{t_{2d_1+1} \dots t_{2d_n+1}}{n!} \int_{\mathcal{M}_{g,n}} \prod_{i=1}^n c_1(\mathcal{L}_i)^{d_i}, \end{aligned}$$

where G_N are all non-equivalent ribbon graphs Γ with N distinct loops obtained by picking vertices from which emerge 3 ribbons, interconnecting them and associating a z_α with each edge ($1 \leq \alpha \leq N$). Each ribbon has two edges, one going with some z_α and another going with some z_β ($1 \leq \alpha, \beta \leq N$). The product in the formula above is taken over all such ribbons of the graph. The second formula involves Chern classes $c_1(\mathcal{L}_i)$ of line bundles on the Deligne-Mumford smooth compactification $\overline{\mathcal{M}}_{g,n}$ of the moduli space $\mathcal{M}_{g,n}$ of smooth Riemann surfaces of genus g with n distinct marked points x_1, \dots, x_n ; the fibers of the line bundle are given by $T_{x_i}^* \mathcal{C}$ at each point $(\mathcal{C}, x_1, \dots, x_n) \in \overline{\mathcal{M}}_{g,n}$.

Recently, other integrals have arisen, like

$$\int_{O(n)} e^{xtrM} dM, \int_{U(n)} e^{\sqrt{x}tr(M+\bar{M})} dM, \int_{U(n)} \det(I+M)^k e^{-x tr\bar{M}} dM,$$

whose expansion in x relate to problems in combinatorics. The integrals over the space of Hermitian, symmetric and symplectic matrices, namely

$$\int_{\mathcal{H}_n, S_n \text{ or } \mathcal{T}_{2n}} e^{-trV(M)} dM,$$

play a prominent role in theory of random matrices. These two sets of integrals provide the main thrust of this lecture. Sequences of such integrals (in n) provide solutions to integrable lattices. For a more detailed overview on these questions, see the MSRI lectures [39]. The last section contains a table listing these connections between matrix integrals, moment matrices and integrable lattices.

1. LARGEST INCREASING SEQUENCES IN RANDOM PERMUTATIONS

Let S_N be the group of permutations π_N , equipped with the uniform probability distribution:

$$(1.0.1) \quad P(\pi_N) = 1/N!.$$

An *increasing subsequence* of $\pi_N \in S_N$ is a sequence $1 \leq j_1 < \dots < j_k \leq N$, such that $\pi(j_1) < \dots < \pi(j_k)$. Define

$$(1.0.2) \quad L_N(\pi_N) = \text{length of the longest increasing subsequence of } \pi_N .$$

Problem: Find the probability $P(L_N \leq n)$?

$$\text{Examples: } \begin{cases} \text{for } \pi_7 = (\underline{3}, 1, \underline{4}, 2, \underline{6}, \underline{7}, 5), \text{ we have } L(\pi_7) = 4. \\ \text{for } \pi_5 = (5, \underline{1}, \underline{4}, 3, 2), \text{ we have } L(\pi_5) = 2. \end{cases}$$

1.1. Robinson-Schensted-Knuth correspondence and symmetric functions

I give here a very brief summary of well-known, but necessary facts to approach the problem:

- A *Young diagram* λ is a finite sequence of non-increasing, non-negative integers $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\ell \geq 0$; also called a *partition* of $n = |\lambda| := \lambda_1 + \dots + \lambda_\ell$, with $|\lambda|$ being the weight. It can be represented by a diagram, having λ_1 boxes in the first row, λ_2 boxes in the second row, etc..., all aligned to the left. A *dual Young diagram* $\hat{\lambda} = (\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots)$ is the diagram obtained by flipping the diagram λ about its diagonal; thus $\hat{\lambda}_1 = \ell = \text{length of first column of } \lambda$.
- A *Young tableau* of shape λ is an array of positive integers a_{ij} (at place (i, j) in the Young diagram) placed in the Young diagram λ , which are non-decreasing from left to right *and* strictly increasing from top to bottom.
- A *standard Young tableau* of shape λ is an array of integers $1, \dots, n$ placed in the Young diagram, which are strictly increasing from left to right *and* from top to bottom. The number of Young tableaux of a given shape $\lambda = (\lambda_1 \geq \dots \geq \lambda_m)$ is

given by a number of formulae (for the Schur polynomial s_λ , see below)¹

$$\begin{aligned}
 f^\lambda &= \#\{\text{standard tableaux of shape } \lambda\} \\
 &= \text{coefficient of } x_1 x_2 \dots x_n \text{ in } s_\lambda(x) \\
 &= \frac{|\lambda|!}{\prod_{\text{all } i,j} h_{ij}^\lambda} = |\lambda|! \det \left(\frac{1}{(\lambda_i - i + j)!} \right) \\
 (1.1.1) \quad &= |\lambda|! \prod_{1 \leq i < j \leq m} (h_i - h_j) \prod_1^m \frac{1}{h_i!}, \text{ with } h_i := \lambda_i - i + m, \quad m := \hat{\lambda}_1.
 \end{aligned}$$

- The *Schur polynomial* s_λ associated with a Young diagram λ is a symmetric function in the variables x_1, x_2, \dots (finite or infinite), defined by

$$(1.1.2) \quad s_\lambda(x_1, x_2, \dots) := \sum_{\{a_{ij}\} \text{ tableaux of } \lambda} \prod_{ij} x_{a_{ij}}.$$

- The linear space Λ_n of *symmetric polynomials* in x_1, \dots, x_n with rational coefficients comes equipped with an inner product, which can also be expressed as an integral over the unitary group $U(n)$ for Haar measure dM :

$$(1.1.3) \quad \langle f, g \rangle = \frac{1}{n!} \int_{(S_1)^n} f(z_1, \dots, z_n) g(\bar{z}_1, \dots, \bar{z}_n) \prod_{1 \leq k < \ell \leq n} |z_k - z_\ell|^2 \prod_1^n \frac{dz_k}{2\pi i z_k} = \int_{U(n)} f(M) g(\bar{M}) dM.$$

- An *orthonormal basis of the space* Λ_n is given by the Schur polynomials above $s_\lambda(x_1, \dots, x_n)$, in which the numbers a_{ij} are restricted to $1, \dots, n$; therefore we have the following “*Fourier series*”:

$$(1.1.4) \quad f(x_1, \dots, x_n) = \sum_{\substack{\lambda \text{ with} \\ \hat{\lambda}_1 \leq n}} \langle f, s_\lambda \rangle s_\lambda(x_1, \dots, x_n), \text{ with } \langle s_\lambda, s_{\lambda'} \rangle = \delta_{\lambda\lambda'}.$$

In particular, one computes the following Fourier series:

$$(1.1.5) \quad (x_1 + \dots + x_n)^k = \sum_{\substack{|\lambda|=k \\ \hat{\lambda}_1 \leq n}} f^\lambda s_\lambda.$$

If $\lambda = (\lambda_1 \geq \dots \geq \lambda_\ell)$, with² $\hat{\lambda}_1 = \ell > n$, then obviously $s_\lambda = 0$.

- *Robinson-Schensted-Knuth correspondence*: There is a 1-1 correspondence

$$(1.1.6) \quad S_n \longrightarrow \left\{ \begin{array}{l} (P, Q), \text{ two standard Young} \\ \text{tableaux from } 1, \dots, n, \text{ where} \\ P \text{ and } Q \text{ have the same shape} \end{array} \right\}.$$

Given a permutation i_1, \dots, i_n , the correspondence constructs two standard Young tableaux P, Q having the same shape λ . This construction is inductive. Namely, having obtained two equally shaped Young diagrams P_k, Q_k from i_1, \dots, i_k , with the

¹ $h_{ij}^\lambda := \lambda_i + \hat{\lambda}_j - i - j + 1$ is the *hook length* of the i, j th box in the Young diagram.

²Remember, from the definition of the dual Young diagram, that $\hat{\lambda}_1 =$ the length of the first column of λ

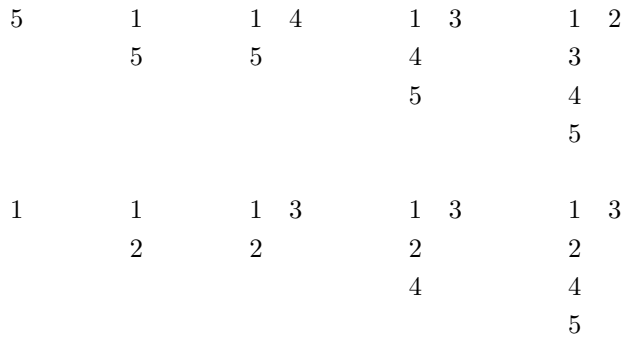
numbers (i_1, \dots, i_k) in the boxes of P_k and the numbers $(1, \dots, k)$ in the boxes of Q_k , one creates a new diagram Q_{k+1} , by putting the *next number* i_{k+1} in the first row of P , according to the following rule:

- (i): if $i_{k+1} \geq$ all numbers appearing in the first row of P_k , then one creates a new box with i_{k+1} in that box to the right of the first column,
- (ii): if not, place i_{k+1} in the box (of the first row) with the smallest higher number. That number then gets pushed down to the second row of P_k according to the rule (i) or (ii), as if the first row had been removed.

The diagram Q is a bookkeeping device; namely, add a box (with the number $k + 1$ in it) to Q_k exactly at the place, where the new box has been added to P_k . This produces a new diagram Q_{k+1} of same shape as P_{k+1} .

The inverse of this map is constructed essentially by reversing the steps above.

Example: $\pi = (5, 1, 4, 3, 2) \in S_5$,



Hence $\pi \longrightarrow (P(\pi), Q(\pi)) = \left(\left(\begin{pmatrix} 1 & 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 3 \\ 2 \\ 4 \\ 5 \end{pmatrix} \right) \right)$ and $L_5(\pi) = 2 = \# \text{columns of } \begin{cases} P \\ \text{or} \\ Q \end{cases}$

The Robinson-Schensted-Knuth correspondence has the following properties

- $\pi \mapsto (P, Q)$, then $\pi^{-1} \mapsto (Q, P)$
- length (longest increasing subsequence of π) = # (columns in P)
- length (longest decreasing subsequence of π) = # (rows in P)
- $\pi^2 = I$, then $\pi \mapsto (P, P)$
- $\pi^2 = I$, with k fixed points, then P has exactly k columns of odd length.

1.2. Plancherel measure, Integrals over $U(n)$ and Toeplitz matrices

A next set of ideas is due to Vershik & Kerov [40], Diaconis & Shashahani [13], Biane [9], Rains [33, 34], Baik & Rains [7]. For a nice state-of-the-art account, see Aldous & Diaconis [5]. The uniform probability measure (see (1.1.1) for f^λ) (1.0.1) on S_N induces

on Young diagrams, via the RSK bijection (1.1.6), the “Plancherel” probability measure

$$(1.2.1) \quad P_N(\lambda) = \frac{(f^\lambda)^2}{N!}, \quad |\lambda| = N,$$

and so from (1.2.1) and the RSK bijection, one deduces a number of formulae below; notice, equality (ii) follows from (1.1.1) and (iii) follows from (i) and (1.1.5):

$$\begin{aligned}
 P_N(L(\pi_N) \leq n) &= \frac{1}{N!} \# \left\{ (P, Q), \text{ standard Young tableaux each of } \right. \\
 &\quad \left. \text{arbitrary shape } \lambda, \text{ with } |\lambda| = N, \lambda_1 \leq n \right\} \\
 &\stackrel{(i)}{=} \frac{1}{N!} \sum_{\substack{|\lambda|=N \\ \lambda_1 \leq n}} (f^\lambda)^2 = \frac{1}{N!} \sum_{\substack{|\lambda|=N \\ \hat{\lambda}_1 \leq n}} (f^\lambda)^2, \text{ by duality of Young diagrams} \\
 &\stackrel{(ii)}{=} N! \sum_{m=1}^n \frac{1}{m!} \sum_{\substack{h \in \mathbb{Z}^m, h_i \geq 1 \\ \sum h_j = N + \frac{m(m-1)}{2}}} \prod_{1 \leq i < j \leq m} (h_i - h_j)^2 \prod_1^m \frac{1}{h_j!^2} \\
 &\stackrel{(iii)}{=} \frac{1}{N!} \sum_{\substack{|\lambda|=|\mu|=N \\ \hat{\lambda}_1, \hat{\mu}_1 \leq n}} f^\lambda f^\mu \langle s_\lambda, s_\mu \rangle = \frac{1}{N!} \langle (x_1 + \dots + x_n)^N, (x_1 + \dots + x_n)^N \rangle \\
 (1.2.2) \quad &\stackrel{(iv)}{=} \frac{1}{N!} \int_{U(n)} |\text{tr} M|^{2N} dM = \frac{1}{N!} \binom{2N}{N}^{-1} \int_{U(n)} (\text{tr}(M + \bar{M}))^{2N} dM,
 \end{aligned}$$

using in the last equality the fact that $\int_{U(n)} ((\text{tr} M)^i (\overline{\text{tr} M})^j) dM = 0$ for $i \neq j$. In 1990, Gessel [17] considered the generating function below and showed that it equals a Toeplitz determinant (determinant of a matrix, whose (i, j) th entry depends on $i - j$ only).

$$\begin{aligned}
 (1.2.3) \quad \tau_n(x) &:= \sum_{N=0}^{\infty} \frac{x^N}{N!} P(L(\pi_N) \leq n) \\
 &\stackrel{(i)}{=} \int_{U(n)} e^{\sqrt{x} \text{tr}(M + \bar{M})} dM \\
 &\stackrel{(ii)}{=} \frac{1}{n!} \int_{(S^1)^n} \Delta_n(z) \Delta_n(\bar{z}) \prod_{k=1}^n \left(e^{\sqrt{x}(z_k + \bar{z}_k)} \frac{dz_k}{2\pi i z_k} \right) \\
 &\stackrel{(iii)}{=} \det \left(\int_0^{2\pi} e^{2\sqrt{x} \cos \theta} e^{i(k-\ell)\theta} d\theta \right)_{1 \leq k, \ell \leq n} \quad (\text{Toeplitz determinant}) \\
 &\stackrel{(iv)}{=} \prod_{k=0}^{n-1} h_k(x)^{-1}, \text{ where } h_k(x) := \frac{\tau_{k+1}}{\tau_k} = \int_{S^1} |p_k(z)|^2 e^{\sqrt{x}(z + \bar{z})} dz, \\
 &\stackrel{(v)}{=} e^x \prod_{k=n}^{\infty} h_k(x)^{-1},
 \end{aligned}$$

where $p_k(z)$ are monic orthogonal polynomials on the circle S^1 for the weight $e^{\sqrt{x}(z+\bar{z})} dz$:

$$(1.2.4) \quad \int_{S^1} p_n(z) \overline{p_m(z)} e^{\sqrt{x}(z+\bar{z})} \frac{dz}{2\pi iz} = h_n(x) \delta_{mn}.$$

Equality (1.2.3)(i) follows from (1.2.2)(iv); moreover (1.2.3)(ii) uses Haar measure on $U(n)$. Identity (1.2.3) (iii) follows from the fact that the product of the two Vandermonde appearing in the integral (ii) can be expressed as sum of determinants:

$$\Delta_n(u) \Delta_n(v) = \sum_{\sigma \in S_n} \det \left(u_{\sigma(k)}^{\ell-1} v_{\sigma(k)}^{k-1} \right)_{1 \leq \ell, k \leq n} = \det \left(p_{\ell-1}^{(1)}(u_k) \right)_{1 \leq \ell, k \leq n} \det \left(p_{\ell-1}^{(2)}(v_k) \right)_{1 \leq \ell, k \leq n}.$$

The fact above that $\Delta_n(u) \Delta_n(v)$ can also be expressed as a product of two determinants, involving monic polynomials, implies (iv). Finally, by Szegő's strong limit theorem for Toeplitz determinants, we have $\lim_{n \rightarrow \infty} \tau_n = e^x$, thus leading to (v).

1.3. Virasoro constraints, Integrable systems and Painlevé V equation

THEOREM 1.1. — For every $\ell \geq 0$, the Gessel generating function

$$(1.3.1) \quad \sum_{N=0}^{\infty} \frac{x^N}{N!} P(L(\pi_N) \leq \ell) = \int_{U(\ell)} e^{\sqrt{x} \operatorname{tr}(M+\bar{M})} dM = \exp \int_0^x \log \left(\frac{x}{u} \right) g_{\ell}(u) du,$$

where g_{ℓ} is the unique solution to the initial value problem:

$$(1.3.2) \quad \begin{cases} g'' - \frac{g'^2}{2} \left(\frac{1}{g-1} + \frac{1}{g} \right) + \frac{g'}{u} + \frac{2}{u} g(g-1) - \frac{\ell^2}{2u^2} \frac{g-1}{g} = 0 \\ g_{\ell}(u) = 1 - \frac{u^{\ell}}{(\ell!)^2} + O(u^{\ell+1}), \text{ near } u = 0. \text{ (**Painlevé V**)} . \quad \square \end{cases}$$

Theorem 1.1 is due to Hisakado [19], Tracy-Widom [36], by methods of functional analysis and Adler-van Moerbeke [3], by integrable methods. A similar statement holds for the set S_{2n}^0 of fixed-point free involutions π^0 (i.e., $(\pi^0)^2 = I$ and $\pi^0(k) \neq k$ for $1 \leq k \leq 2n$). Put the uniform distribution on S_{2n}^0 :

$$(1.3.3) \quad P(\pi_{2n}^0) = \frac{1}{(2n-1)!!} = \frac{2^n n!}{(2n)!}.$$

Then, in the following statement, due to Adler-van Moerbeke [3] and Baik-Rains [7], $O(n)_{\pm}$ refers to orthogonal matrices, with determinant $= \pm 1$.

THEOREM 1.2. — The generating function

$$(1.3.4) \quad \begin{aligned} 2 \sum_{k=0}^{\infty} \frac{(x^2/2)^k}{k!} P(L(\pi_{2k}^0) \leq \ell + 1) &= E_{O(\ell+1)_-} e^{x \operatorname{tr} M} + E_{O(\ell+1)_+} e^{x \operatorname{tr} M} \\ &= \exp \left(\int_0^x \frac{f_{\ell}^-(u)}{u} du \right) + \exp \left(\int_0^x \frac{f_{\ell}^+(u)}{u} du \right), \end{aligned}$$

where $f = f_\ell^\pm$ is the unique solution to the initial value problem:

$$(1.3.5) \quad \begin{cases} f''' + \frac{1}{u}f'' + \frac{6}{u}f'^2 - \frac{4}{u^2}ff' - \frac{16u^2 + \ell^2}{u^2}f' + \frac{16}{u}f + \frac{2(\ell^2 - 1)}{u} = 0 \\ f_\ell^\pm(u) = u^2 \pm \frac{u^{\ell+1}}{\ell!} + O(u^{\ell+2}), \text{ near } u = 0. \text{ (Painlevé V)} \end{cases}$$

□

Proof of Theorem 1.1: A brief outline will be given here, because of its interesting connection with an integrable system, called the Toeplitz lattice (see [3]). Inserting times t_i , with $i = \dots, -2, -1, 0, 1, 2, \dots$, and $t_0 = 0$, in the integrals (1.2.3)(ii) over $(S^1)^n$, one obtains, setting $t := (\dots, t_{-1}, t_0, t_1, \dots)$:

$$(1.3.6) \quad I_n(t) = \int_{(S^1)^n} |\Delta_n(z)|^2 \prod_{k=1}^n e^{\sum_{i \neq 0} t_i z_k^i} \frac{dz_k}{2\pi i z_k}, \quad n \geq 0.$$

Virasoro constraints. [3] The I_n 's satisfy a set of three linear partial differential equations, forming an $\mathfrak{sl}(2, \mathbb{Z})$ -algebra (Virasoro constraints), explicitly given by the three equations:

$$(1.3.7) \quad \left(\sum_{i \neq \mp 1, 0} (i \pm 1) t_{i \pm 1} \frac{\partial}{\partial t_i} + n \left(t_{\pm 1} + \frac{\partial}{\partial t_{\mp 1}} \right) \right) I_n(t) = 0 \text{ and } \left(\sum_{i \neq 0} i t_i \frac{\partial}{\partial t_i} \right) I_n(t) = 0$$

Toeplitz lattice. The t -dependent semi-infinite moment matrix,

$$(1.3.8) \quad m_n(t) := (\langle z^k, z^\ell \rangle_t)_{0 \leq k, \ell \leq n-1} = \left(\oint_{S^1} \frac{\rho(z) dz}{2\pi i z} z^{k-\ell} e^{\sum_{i \neq 0} (t_i z^i)} \right)_{0 \leq k, \ell \leq n-1}$$

is a Toeplitz matrix and satisfies the simple differential equations:

$$(1.3.9) \quad \frac{\partial \mu_{ij}}{\partial t_k} = \mu_{i+k, j}, \quad \frac{\partial \mu_{ij}}{\partial t_{-k}} = \mu_{i, j+k}, \quad \text{for } k \geq 1.$$

In analogy with (1.2.3)(iii), define $\tau_n(t)$, instead of $\tau_n(x)$:

$$\tau_n(t) := \det m_n(t).$$

The Borel factorization $m_\infty(t) = S_1^{-1}(t)S_2(t)$ in a lower-triangular matrix S_1 (with 1's along the diagonal) and an upper-triangular matrix S_2 , enables one to define³ $L_1 := S_1 \Lambda S_1^{-1}$ and $L_2 := S_2 \Lambda^\top S_2^{-1}$, which combined with the Toeplitz nature of m_∞ has a peculiar “rank 2”-structure, where x_i and y_i are certain integrals over $U(n)$ and where

³where Λ is the shift matrix $(\Lambda v)_n = v_{n+1}$, $n \geq 0$, i.e., Λ is the semi-infinite matrix with all 0 entries, except for 1's just above the diagonal.

$h_i = \tau_{i+1}/\tau_i$, $h_i/h_{i-1} = 1 - x_i y_i$, $h := \text{diag}(h_0, h_1, \dots)$ and $x_0 = y_0 = 1$:

$$L_1 = h \begin{pmatrix} -x_1 y_0 & 1 - x_1 y_1 & 0 & 0 & & \\ -x_2 y_0 & -x_2 y_1 & 1 - x_2 y_2 & 0 & & \\ -x_3 y_0 & -x_3 y_1 & -x_3 y_2 & 1 - x_3 y_3 & & \\ -x_4 y_0 & -x_4 y_1 & -x_4 y_2 & -x_4 y_3 & & \\ & & & & \ddots & \end{pmatrix} h^{-1}$$

and

$$(1.3.10) \quad L_2 = \begin{pmatrix} -x_0 y_1 & -x_0 y_2 & -x_0 y_3 & -x_0 y_4 & & \\ 1 - x_1 y_1 & -x_1 y_2 & -x_1 y_3 & -x_1 y_4 & & \\ 0 & 1 - x_2 y_2 & -x_2 y_3 & -x_2 y_4 & & \\ 0 & 0 & 1 - x_3 y_3 & -x_3 y_4 & & \\ & & & & \ddots & \end{pmatrix}.$$

The **2d-Toda Lattice** hierarchy⁴

$$(1.3.11) \quad \frac{\partial L_i}{\partial t_n} = [(L_1^n)_+, L_i] \quad \text{and} \quad \frac{\partial L_i}{\partial t_{-n}} = [(L_2^n)_-, L_i], \quad i = 1, 2, n \geq 1,$$

maintains the “rank 2” nature of the semi-infinite matrices L_1 and L_2 . The first equation in this hierarchy is equivalent to the **discrete sinh-Gordon equation**

$$\frac{\partial^2 q_n}{\partial t_1 \partial t_{-1}} = -e^{q_n - q_{n-1}} + e^{q_{n+1} - q_n}, \quad \text{where } q_n = \log \frac{\tau_{n+1}}{\tau_n}.$$

The equations (1.3.11) induce on the x_i and y_i the **Toeplitz lattice** equations; see [3].

Using the three equations (1.3.7) and $\partial/\partial t_1$ of the first equation of (1.3.7), and setting all $t_i = 0$, enable one to extract various t -partials, like $\partial^2 \log \tau(t)/\partial t_{-1} \partial t_1$ and $\partial^2 \log \tau(t)/\partial t_{-2} \partial t_1$, in terms of pure partials in $\partial^k/\partial x^k$, where $x = t_1 t_{-1}$. Substituting these expressions in the integrable Toeplitz lattice equations leads to Painlevé V (1.3.2). The integrable system associated with the integral (1.3.4) in Theorem 1.2 (with t_i 's inserted) is the standard Toda lattice, instead of the Toeplitz lattice; the integral satisfies the Virasoro constraints associated with the Toda lattice; see [3].

1.4. Random permutations π_N for large N

Around 1960 and based on Monte-Carlo methods, Ulam [38] conjectured that

$$\lim_{n \rightarrow \infty} \frac{E(L_n)}{\sqrt{n}} = c \text{ exists.}$$

An argument of Erdős & Szekeres [14], dating back from 1935 showed that $E(L_n) \geq \frac{1}{2} \sqrt{n-1}$, and thus $c \geq 1/2$. In '72, Hammersley [18] showed rigorously that the limit exists. Logan and Shepp [23] showed the limit $c \geq 2$, and finally Vershik and Kerov

⁴ $()_+$ denotes the upper-triangular part of $()$, including the diagonal, whereas $()_-$ denotes the strictly lower-triangular part of $()$

[40] that $c = 2$. The next major contribution was due to Johansson [22] and Baik-Deift-Johansson [6]:

THEOREM 1.3. — (“Law of large numbers” and “Central limit theorem”) One has:

$$(1.4.1) \quad \lim_{n \rightarrow \infty} \frac{L_n}{2\sqrt{n}} = 1, \text{ and } \lim_{n \rightarrow \infty} P\left(\frac{L_n - 2\sqrt{n}}{n^{1/6}} \leq x\right) = \exp\left(-\int_x^\infty (u-x)g^2(u)du\right),$$

where $g(x)$ is a solution of (a version of) Painlevé II,

$$(1.4.2) \quad \begin{cases} g'' = xg + 2g^3 \text{ (Painlevé II)} \\ g(x) \cong A(x) := \int_{-\infty}^\infty e^{ixy-y^3/3} dy \cong \frac{e^{-\frac{2}{3}x^{\frac{3}{2}}}}{2\sqrt{\pi x^{1/4}}} \text{ for } x \nearrow \infty. \end{cases}$$

□

The point $\gamma = \frac{2\sqrt{x}}{n} = 1$ corresponds to a “phase transition” for the partition function

$$e^{-x}\tau_n(x) = e^{-x} \int_{U(n)} e^{\sqrt{x} \operatorname{tr}(M+\bar{M})} dM = \prod_{k=n}^\infty h_k(x)^{-1} = e^{-x} \sum_{n=0}^\infty \frac{x^N}{N!} P(L(\pi_N) \leq n)$$

On the one hand, Johansson shows, for all $\varepsilon > 0$, there exist C and $\delta > 0$ such that

$$(1.4.3) \quad e^{-x}\tau_n(x) \leq Ce^{-\delta x}, \text{ for } \frac{2\sqrt{x}}{n} > 1 + \varepsilon \text{ and } e^{-x}\tau_n(x) \geq 1 - \frac{C}{n}, \text{ for } \frac{2\sqrt{x}}{n} < 1 - \varepsilon$$

On the other hand, setting $P_{n,N} := P(L(\pi_N) \leq n)$, Johansson’s de-Poissonization Lemma goes as follows: given any $\alpha > 3$, there are constants $C = C(\alpha)$ and $N_0 = N_0(\alpha)$ such that the following holds for $N \geq N_0$ and all $0 \leq n \leq N$, with $m(\alpha) = (\alpha^2 - 2\alpha - 1)/2$:

$$(1.4.4) \quad e^{-x} \sum_{n=0}^\infty \frac{x^N}{N!} P_{n,N} \Big|_{x=N+\alpha\sqrt{N \log N}} - \frac{C}{N^{m(\alpha)}} \leq P_{n,N} \leq e^{-x} \sum_{n=0}^\infty \frac{x^N}{N!} P_{n,N} \Big|_{x=N-\alpha\sqrt{N \log N}} + \frac{C}{N^{m(\alpha)}}.$$

The two relations (1.4.3) and (1.4.4) combined lead to the law of large numbers in (1.4.1). To prove the “central limit theorem” (1.4.1), one uses the Riemann-Hilbert approach to obtain asymptotics. Indeed, as a first observation, setting $\rho_x(z) := e^{\sqrt{x}(z+z^{-1})}$, the only solution Y_{n+1} (2×2 matrix) to the following Riemann-Hilbert problem *à la* Fokas, Its and Kitaev [15], but adapted to S^1 :

1. $Y(z)$ holomorphic in $\mathbb{C} \setminus S^1$

$$2. Y_-(z) = Y_+(z) \begin{pmatrix} 1 & \frac{\rho_x(z)}{z^{n+1}} \\ 0 & 1 \end{pmatrix}.$$

$$3. Y(z) \begin{pmatrix} z^{-(n+1)} & 0 \\ 0 & z^{n+1} \end{pmatrix} = (I + O(z^{-1})), \text{ when } z \rightarrow \infty.$$

is given by the matrix

$$(1.4.5) \quad Y_{n+1}(z) = \begin{pmatrix} p_{n+1}(z) & \int_{S^1} p_{n+1}(u) \frac{\rho_x(u)}{u-z} \frac{du}{2\pi i u^{n+1}} \\ -h_n^{-1} z^n \overline{p_n(1/\bar{z})} & -h_n^{-1} \int_{\mathbb{R}} \overline{p_n(1/\bar{u})} \frac{\rho_x(u)}{u-z} \frac{du}{2\pi i u^{n+1}} \end{pmatrix},$$

where $p_k(z)$ are the monic orthogonal polynomials (1.2.4) on the circle S^1 for the weight $\rho_x(z)dz$. The Riemann-Hilbert formulation is an efficient tool to find the asymptotics of $(Y_{k+1}(0))_{21} = -h_k^{-1}(x)$. Then, summing up, $\sum_n \log h_k^{-1}(x) = \log(e^{-x}\tau_n(x))$, Baik-Deift-Johansson [6] show the following result: Define u such that $\frac{2\sqrt{x}}{n+1} = 1 - \frac{u}{2^{1/3}(n+1)^{2/3}}$, with $-M < u < M$ for a given $M > 0$. Then the estimate below holds, where g is the solution to (1.4.2):

$$\left| \log(e^{-x}\tau_n(x)) - \int_u^\infty 2if(y)dy \right| = \frac{C(M)}{n^{1/3}} + Ce^{-1/4M^{3/2}} \text{ with } 2if(y) = - \int_y^\infty g^2(\alpha)d\alpha.$$

This estimate combined with Johansson’s de-Poissonization Lemma leads to (1.4.1) and (1.4.2).

2. THE SPECTRUM OF RANDOM MATRICES

Random matrix theory deals with an ensemble of matrices M , having some symmetry condition to guarantee the reality of the spectrum, e.g., the Hermitian ensemble \mathcal{H}_n , the symmetric ensemble \mathcal{S}_n or the symplectic ensemble \mathcal{T}_{2n} . Define a probability measure, based on a Haar measure, which in “polar coordinates” is expressed in terms of the Vandermonde with a power β :

$$c_n e^{-trV(M)} dM = c_n |\Delta_n(z)|^\beta \prod_1^n e^{-V(z_k)} dz_k dU, \text{ with } V' = \frac{g}{f} \text{ rational, } dM = \text{Haar measure.}$$

The three ensembles $\mathcal{H}_n, \mathcal{S}_n, \mathcal{T}_{2n}$ appear very naturally as the tangent spaces (at the identity) to the simplest symmetric spaces:

non-compact symmetric space G/K	$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with \mathfrak{p} =tangent space to G/K at I	space $\mathfrak{p} = \{M \in \mathfrak{g}, \text{with...}\}$	Haar measure on \mathfrak{p}
$SL(n, \mathbb{C})/SU(n)$	$sl(n, \mathbb{C}) = su(n) \oplus \mathcal{H}_n$	$M = \bar{M}^\top$	$\beta = 2$
$SL(n, \mathbb{R})/SO(n)$	$sl(n, \mathbb{R}) = so(n) \oplus \mathcal{S}_n$	$M = M^\top$	$\beta = 1$
$SU^*(2n)/USp(n)$	$su^*(2n) = (sp(n, \mathbb{C}) \cap u(2n)) \oplus \mathcal{T}_{2n}$	$M = \bar{M}^\top, M = J\bar{M}J^{-1}$	$\beta = 4$

Question: What is the statistics of the spectrum of M ?

For \mathcal{H}_n and \mathcal{S}_n , if the probability $P(M \in dM)$ satisfies the following two requirements: (i) invariance under conjugation by unitary transformations $M \mapsto UMU^{-1}$, (ii) the random variables $M_{ii}, \Re M_{ij}, \Im M_{ij}, 1 \leq i < j \leq n$ are independent, then $V(z)$ is quadratic (Gaussian ensemble) ([28].)

2.1. Virasoro constraints, Toda and Pfaff lattices and KP equations

Consider weights of the form $\rho(z)dz := e^{-V(z)}dz$ on an interval $F = [A, B] \subseteq \mathbb{R}$, with rational logarithmic derivative and subjected to the following boundary conditions:

$$(2.1.1) \quad -\frac{\rho'}{\rho} = \frac{g}{f} = \frac{\sum_0^\infty b_i z^i}{\sum_0^\infty a_i z^i}, \quad \lim_{z \rightarrow A, B} f(z)\rho(z)z^k = 0 \text{ for all } k \geq 0,$$

and a disjoint union of intervals $E = \bigcup_1^{2r} [x_{2i-1}, x_{2i}] \subset F \subset \mathbb{R}$.

THEOREM 2.1. — (Adler-van Moerbeke [1]) The vector of integrals $I(t, x; \beta) = (I_0 = 1, I_1(t, x; \beta), \dots)$, with $t := (t_1, t_2, \dots)$ and $x := (x_1, \dots, x_{2r})$ and

$$(2.1.2) \quad I_n(t, x; \beta) := \int_{E^n} |\Delta_n(z)|^\beta \prod_{k=1}^n \left(e^{\sum_1^\infty t_i z_k^i} \rho(z_k) dz_k \right) \text{ for } n > 0$$

satisfies the following Virasoro constraints⁵ for all $k \geq -1$:

$$\left(-\sum_1^{2r} x_i^{k+1} f(x_i) \frac{\partial}{\partial x_i} + \sum_{i \geq 0} \left(a_i \beta \mathbb{J}_{k+i}^{(2)}(t, n) - b_i \beta \mathbb{J}_{k+i+1}^{(1)}(t, n) \right) \right) I(t, x; \beta) = 0,$$

in terms of the coefficients a_i, b_i of the rational function $(-\log \rho)'$ and the end points x_i of the subset E , as in (2.1.1). The $\beta \mathbb{J}_k^{(2)}$ and $\beta \mathbb{J}_k^{(1)}$ form a Virasoro and a Heisenberg algebra respectively, interacting as follows

$$(2.1.3) \quad \begin{aligned} \left[\beta \mathbb{J}_k^{(2)}, \beta \mathbb{J}_\ell^{(2)} \right] &= (k - \ell) \beta \mathbb{J}_{k+\ell}^{(2)} + c \left(\frac{k^3 - k}{12} \right) \delta_{k, -\ell} \\ \left[\beta \mathbb{J}_k^{(2)}, \beta \mathbb{J}_\ell^{(1)} \right] &= -\ell \beta \mathbb{J}_{k+\ell}^{(1)} + c' k(k + 1) \delta_{k, -\ell}. \\ \left[\beta \mathbb{J}_k^{(1)}, \beta \mathbb{J}_\ell^{(1)} \right] &= \frac{k}{\beta} \delta_{k, -\ell}, \end{aligned}$$

with central charge $c = 1 - 6 \left(\left(\frac{\beta}{2} \right)^{1/2} - \left(\frac{\beta}{2} \right)^{-1/2} \right)^2$ and $c' = \left(\frac{1}{\beta} - \frac{1}{2} \right)$.

Moreover:

(i) $\tau_n(t) := \frac{1}{n!} I_n(t, x, \beta)|_{\beta=2}$ form the τ -functions of the standard Toda lattice; in particular, each τ_n satisfies the KP equation:

$$\left(\left(\frac{\partial}{\partial t_1} \right)^4 + 3 \left(\frac{\partial}{\partial t_2} \right)^2 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau_n + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau_n \right)^2 = 0.$$

(ii) $\tau_{2n}(t) := \begin{cases} \frac{1}{(2n)!} I_{2n}(t, x; \beta)|_{\beta=1} \\ \frac{1}{n!} I_n(2t, x; \beta)|_{\beta=4} \end{cases}$ form the τ -functions of the Pfaff lattice (i.e., they are Pfaffians, rather than determinants); in particular, they satisfy the Pfaff-KP equation:

⁵When E equals the whole range F , then the $\partial/\partial x_i$'s are absent in the formulae (2.1.7).

$$\left(\left(\frac{\partial}{\partial t_1} \right)^4 + 3 \left(\frac{\partial}{\partial t_2} \right)^2 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau_{2n} + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau_{2n} \right)^2 = 12 \frac{\tau_{2n-2} \tau_{2n+2}}{\tau_{2n}^2}.$$

2.2. The Gaussian ensemble: PDE’s for the statistics of the spectrum

THEOREM 2.2. — For the Gaussian ensemble, the probabilities⁶: ($\beta = 2, 1, 4$)

$$\begin{aligned} P_n(E) := P_n(\text{all spectral points of } M \in E) &= \frac{\int_{\mathcal{H}_n(E), \mathcal{S}_n(E) \text{ or } \mathcal{T}_n(E)} e^{-tr M^2} dM}{\int_{\mathcal{H}_n(\mathbb{R}), \mathcal{S}_n(\mathbb{R}) \text{ or } \mathcal{T}_n(\mathbb{R})} e^{-tr M^2} dM} \\ (2.2.1) \qquad \qquad \qquad &= \frac{\int_{E^n} |\Delta_n(z)|^\beta \prod_{k=1}^n e^{-z_k^2} dz_k}{\int_{\mathbb{R}^n} |\Delta_n(z)|^\beta \prod_{k=1}^n e^{-z_k^2} dz_k} \end{aligned}$$

satisfy the following PDE’s in the x_i ’s ($\mathcal{B}_k = \sum_1^{2r} x_i^{k+1} \frac{\partial}{\partial x_i}$, $\delta_{1,4}^\beta = \begin{cases} 1 & \text{for } \beta = 1, 4 \\ 0 & \text{otherwise} \end{cases}$, $f_n(x) = \frac{d}{dx} \log P_n$):⁷

$$\begin{aligned} \delta_{1,4}^\beta Q \left(\frac{P_{n-1} P_{n+1}}{P_n^2} - 1 \right) & \begin{cases} 2 \text{ and } n \text{ even, for } \beta = 1 \\ 1 \text{ and } n \text{ arbitrary, for } \beta = 4 \end{cases} \\ = \begin{cases} \left(\mathcal{B}_{-1}^4 + (Q_2 + 6\mathcal{B}_{-1}^2 F) \mathcal{B}_{-1}^2 + (2 - \delta_{1,4}^\beta) \frac{4}{\beta} (3\mathcal{B}_0^2 - 4\mathcal{B}_{-1} \mathcal{B}_1 + 6\mathcal{B}_0) \right) \log P_n & \text{for } E = \bigcup_1^{2r} [x_{2i-1}, x_{2i}] \\ f_n''' + 6f_n'^2 + \left(\frac{4x^2}{\beta} (\delta_{1,4}^\beta - 2) + Q_2 \right) f_n' - \frac{4x}{\beta} (\delta_{1,4}^\beta - 2) f_n, & \text{for } E = [-\infty, x] \end{cases} \end{aligned}$$

Note that for $\beta = 2$ and for $E = [-\infty, x]$, the equation above takes on the simple form:

$$(2.2.2) \qquad f_n''' + 6f_n'^2 + 8nf_n' + 4xf_n = 0 \quad (\text{Painlevé IV}) \quad \square$$

Equation (2.2.2) was obtained by Tracy-Widom [35] and the rest of Theorem 2.2 is due to M. Adler, T. Shiota and P. van Moerbeke ([4] and [1]).

2.3. Infinite Hermitian matrix ensembles

Consider probability (2.2.1) for $\beta = 2$ and let the size n of the matrices go to ∞ . To perform this limit, one uses a different representation; namely,

$$(2.3.1) \qquad P_n(E) := \frac{\int_{\mathcal{H}_n(E)} e^{-tr M^2} dM}{\int_{\mathcal{H}_n} e^{-tr M^2} dM} = \frac{1}{n!} \int_{E^n} \det(K_n(z_k, z_\ell))_{1 \leq k, \ell \leq n} \prod_1^n \rho(dz_i),$$

⁶ $\mathcal{H}_n(E)$, $\mathcal{S}_n(E)$ or $\mathcal{T}_n(E)$ denotes the subset of matrices with spectrum in $E \subseteq \mathbb{R}$

⁷Also, define the *invariant* polynomials

$$Q = 12n \left(n + 1 - \frac{2}{\beta} \right) \quad \text{and} \quad Q_2 = 4(1 + \delta_{1,4}^\beta) \left(2n + \delta_{1,4}^\beta \left(1 - \frac{2}{\beta} \right) \right).$$

can be represented in terms of a reproducing kernel

$$(2.3.2) \quad K_n(y, z) := \sum_{j=1}^n \frac{p_{j-1}(y) p_{j-1}(z)}{\sqrt{h_{j-1}} \sqrt{h_{j-1}}}, \text{ where } \int_{\mathbb{R}} K_n(y, z) K_n(z, u) \rho(dz) = K_n(y, u).$$

The reproducing property follows from the orthogonality of the *monic orthogonal polynomials* $p_k = p_k(z)$ for the Gaussian weight $e^{-z^2} dz$ on \mathbb{R} , and the L^2 -norms $h_k = \int_{\mathbb{R}} p_k^2(z) e^{-z^2} dz$ of the p_k 's. We also have the following Fredholm determinant formula

$$P(\text{exactly } k \text{ eigenvalues } \in [x_1, x_2]) = \frac{(-1)^k}{k!} \left(\frac{\partial}{\partial \lambda} \right)^k \det(I - \lambda K_n(y, z) I_{[x_1, x_2]}(z)) \Big|_{\lambda=1}$$

When the size n of the Hermitian matrices tends to ∞ , the following holds:

- *Wigner's semi-circle law*: For this ensemble and for very large n , the density of eigenvalues tends to the semi-circle distribution on the interval $[-\sqrt{2n}, \sqrt{2n}]$.
- *Bulk scaling limit*: From the formula above, it follows that the average number of eigenvalues per unit length near $z = 0$ ("the bulk") is given by $\sqrt{2n}/\pi$ and thus the average distance between two consecutive eigenvalues is given by $\pi/\sqrt{2n}$. Upon using this rescaling, one shows ([24, 27, 29, 32, 21])

$$\lim_{n \nearrow \infty} \frac{\pi}{\sqrt{2n}} K_n \left(\frac{\pi y}{\sqrt{2n}}, \frac{\pi z}{\sqrt{2n}} \right) = \frac{\sin \pi(y - z)}{\pi(y - z)} =: K(y, z) \quad (\text{Sine kernel})$$

with

$$(2.3.3) \quad P(\text{no eigenvalues } \in [0, x]) = \det(I - K(y, z) I_{[0, x]}(z)) = \exp \int_0^{\pi x} \frac{f(x)}{x} dx,$$

where $f(x, \lambda)$ is a solution to the **Painlevé V** differential equation, (see Jimbo, Miwa, Mori, Sato [21]), ($' = \partial/\partial x$)

$$(xf'')^2 - 4(xf' - f)(-f'^2 - xf' + f) = 0 \text{ with } f(x; \lambda) \cong -\frac{x}{\pi}, \text{ for } x \simeq 0.$$

- *Edge scaling limit*: I will particularly concentrate on the "edge" $\sqrt{2n}$ of the Wigner semi-circle. There the scaling is $\sqrt{2n}^{1/6}$, which makes the problems considerably subtler; this will be used in (2.3.7) (see [10, 16, 28, 35]). The main result can be stated as follows:

THEOREM 2.3. — Given the spectrum $z_1 \geq z_2 \geq \dots$ of the large random Hermitian matrix M , define the "eigenvalues" in a new scale:

$$(2.3.4) \quad u_i = 2n^{\frac{2}{3}} \left(\frac{z_i}{\sqrt{2n}} - 1 \right) \text{ for } n \nearrow \infty$$

then the statistics of the largest "eigenvalue" u_1 (in the new scale) is given by the same probability distribution as the length of the longest increasing sequence:

$$(2.3.5)$$

$$P(u_1 \leq x) = \exp \left(- \int_x^\infty (\alpha - x) g^2(\alpha) d\alpha \right), \text{ with } \begin{cases} g'' = xg + 2g^3 \text{ (Painlevé II)} \\ g(x) \cong A(x) \cong -\frac{e^{-\frac{2}{3}x^{\frac{3}{2}}}}{2\sqrt{\pi}x^{1/4}} \text{ for } x \nearrow \infty. \end{cases}$$

More generally, $P(\text{no points } u_i \in [x_1, x_2])$ satisfies the partial differential equation

(2.3.6)

$$\left(\mathcal{B}_{-1}^3 - 4\left(\mathcal{B}_0 - \frac{1}{2}\right)\right) f + 6(\mathcal{B}_{-1}f)^2 = 0, \text{ with } f := \sum_1^2 \frac{\partial \log P}{\partial x_i}, \mathcal{B}_k = \sum_1^2 x^{k+1} \frac{\partial}{\partial x_i}.$$

□

Tracy-Widom [35] have first obtained result (2.3.5) by methods of functional analysis. In [4], equations (2.3.5) and (2.3.6) were derived using integrable systems and Virasoro constraints.

Proof: Setting $z = \sqrt{2n} + \frac{u}{\sqrt{2n^{1/6}}}$, in the kernel $K_n(x, y)$ of (2.3.2), one proves

$$(2.3.7) \quad \lim_{n \nearrow \infty} \frac{1}{\sqrt{2n^{1/6}}} K_n \left(\sqrt{2n} + \frac{u}{\sqrt{2n^{1/6}}}, \sqrt{2n} + \frac{v}{\sqrt{2n^{1/6}}} \right) = K(u, v),$$

where the Airy kernel K is defined in terms of the Airy function:

(2.3.8)

$$K(u, v) = \int_0^\infty A(x+u)A(x+v)dx, \quad A(u) = \int_{-\infty}^\infty e^{iux-x^3/3} dx \cong \frac{e^{-(2/3)u^{3/2}}}{2\sqrt{\pi}u^{1/4}}, \text{ for } u \nearrow \infty$$

To find the differential equations (2.3.6), one proceeds (sketchily) as follows: as a first step, notice that, setting $L^2 := (\frac{\partial}{\partial t_1})^2 + q(t)$, $t := (t_1, t_2, \dots)$, the solution $q(t)$ and $\Psi(t; z)$ (with asymptotics $\Psi(t; z) = e^{\sum_1^\infty t_i z^i} (1 + O(z^{-1}))$, $z \in \mathbb{C}$, $z \rightarrow \infty$) to the initial value problem:

$$(2.3.9) \quad \begin{cases} (L^2 - z^2)\Psi(t; z) = 0, \quad \frac{\partial L^2}{\partial t_n} = [(L^n)_+, L^2], \quad \frac{\partial \Psi}{\partial t_n} = (L^n)_+ \Psi, \\ q(t^{(0)}) = -x, \text{ at } t = t^{(0)} := (x, 0, \frac{2}{3}, 0, 0, \dots) \end{cases},$$

is given by Kontsevich's integral (Z diagonal and N arbitrary), which itself is intimately related to the Airy function $A(x)$,

(2.3.10)

$$\begin{cases} q(t) = 2 \frac{\partial^2}{\partial t_1^2} \log \tau(t), \text{ with } \tau(t) = \frac{\int_{\mathcal{H}_N} dX e^{-tr(X^3/3 + X^2 Z)}}{\int_{\mathcal{H}_N} dX e^{-tr(X^2 Z)}}, \quad t_n := -\frac{1}{n} tr Z^{-n} + \frac{2}{3} \delta_{n3} \\ \Psi(t; z) = e^{\sum_1^\infty t_i z^i} \tau(t - [z^{-1}]) / \tau(t), \text{ with } [z^{-1}] = (z^{-1}, z^{-2}/2, z^{-3}/3, \dots) \in \mathbb{C}^\infty \\ \Psi(t^{(0)}; z) = z^{1/2} A(x + z^2) = e^{xz + \frac{2}{3}z^3} (1 + O(z^{-1})), \quad z \rightarrow \infty. \end{cases}$$

These data define a kernel

$$(2.3.11) \quad K_t(z^2, z'^2) := \frac{1}{z^{\frac{1}{2}} z'^{\frac{1}{2}}} \int_0^\infty \Psi(t; z) \Psi(t; z') dt_1, \text{ where } t = (t_1, t_2, \dots)$$

which flows off the Airy kernel $K(z^2, z'^2)$, defined in (2.3.8), upon using (2.3.10).

Then, one shows that both, Kontsevich's integral $\tau(t)$ and the product $\tau(t, E) := \tau(t) \cdot \det(I - K_t(\lambda, \lambda')I_E(\lambda'))$, with $E = \bigcup_1^{2r} [x_{2i-1}, x_{2i}] \subset \mathbb{R}$, satisfy Virasoro constraints, of which the two first ones read: (after the time shift $t_3 \mapsto t_3 + 2/3$, in view of (2.3.9))

$$(2.3.12) \quad \begin{aligned} \sum_{i=1}^{2r} \frac{\partial}{\partial x_i} \log \tau(t, E) &= \left(\frac{\partial}{\partial t_1} + \frac{1}{2} \sum_{i \geq 3} it_i \frac{\partial}{\partial t_{i-2}} \right) \log \tau(t, E) + \frac{t_1^2}{4} \\ \sum_{i=1}^{2r} x_i \frac{\partial}{\partial x_i} \log \tau(t, E) &= \left(\frac{\partial}{\partial t_3} + \frac{1}{2} \sum_{i \geq 1} it_i \frac{\partial}{\partial t_i} \right) \log \tau(t, E) + \frac{1}{16}, \end{aligned}$$

Both, $\tau(t)$ and $\tau(t, E)$, also satisfy the Korteweg-de Vries equation (KP equation, depending on odd t_i 's only)

$$(2.3.13) \quad \left(\left(\frac{\partial}{\partial t_1} \right)^4 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau \right)^2 = 0. \quad (\mathbf{KdV})$$

Differentiating (2.3.12) in t_1 and t_3 , and setting all $t_i = 0$ enable one to express the t -partials, appearing in the KdV equation, $\frac{\partial^2}{\partial t_1^2} \log \tau \Big|_{t=0}$, $\frac{\partial^4}{\partial t_1^4} \log \tau \Big|_{t=0}$, $\frac{\partial^2}{\partial t_1 \partial t_3} \log \tau \Big|_{t=0}$ in terms of partials $\partial/\partial x_i$. Setting these expressions in the the KdV -equation at $t = 0$, implies that the statistics of the scaled eigenvalues u_i , (for the Airy kernel (2.3.8))

$$P(E^c) := P(\text{all "eigenvalues" } u_i \in E^c) = \det(I - KI_E) = \frac{\tau(t, E)}{\tau(t)} \Big|_{t=0},$$

satisfies the partial differential equation (2.3.6) with $f := \sum_1^{2r} \frac{\partial \log P}{\partial x_i}$, $\mathcal{B}_k = \sum_1^{2r} x^{k+1} \frac{\partial}{\partial x_i}$.

When $E = (x, \infty)$, the equation (2.3.6) for $f = \partial \log P((-\infty, x])/\partial x$ becomes a 3rd order ODE (Chazy-type equation)

$$(2.3.14) \quad f''' - 4xf' + 2f + 6f'^2 = 0.$$

According to [12], this equation has a first integral, which is a Painlevé II equation

$$(2.3.15) \quad f''^2 + 4f'(f'^2 - xf' + f) = 0.$$

and this equation has a solution given by $f := g'^2 - xg^2 - g^4$ and $f' = -g^2$, which establishes (2.3.5) and (2.3.6).

3. LARGE RANDOM MATRICES AND PERMUTATIONS: A DIRECT CONNECTION VIA ENUMERATIVE GEOMETRY

For large random Hermitian matrices, whose entries have Gaussian real and imaginary part of variance = 1/2, the Wigner semi-circle has support on $(-\sqrt{2n}, \sqrt{2n})$, instead of $[-\sqrt{2n}, \sqrt{2n}]$. Then the (slightly modified) scaling (2.3.4) is given by

$$(3.0.16) \quad u_i = n^{2/3} \left(\frac{z_i}{2n^{1/2}} - 1 \right) \text{ for } n \nearrow \infty.$$

Given the set of partitions $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots)$ of n , in view of (1.4.1), consider

$$(3.0.17) \quad v_i = n^{1/3} \left(\frac{\lambda_i}{2n^{1/2}} - 1 \right) \text{ for } n \nearrow \infty.$$

Theorems 1.3 and 2.3 show that the variables u_1 and v_1 have the same probability distribution. Okounkov [31] shows *in a direct way* that the joint distribution of all u_i and v_i coincide; more specifically:

THEOREM 3.1. — (Okounkov [31]) For any $m \geq 1$, any $\xi_1, \dots, \xi_m > 0$, the following holds:

$$E\left(\sum_{i=1}^{\infty} e^{\xi_1 u_i} \dots \sum_{i=1}^{\infty} e^{\xi_m u_i}\right) = E\left(\sum_{i=1}^{\infty} e^{\xi_1 v_i} \dots \sum_{i=1}^{\infty} e^{\xi_m v_i}\right), \text{ when } n \nearrow \infty. \quad \square$$

The left hand side, which relates to random permutations, hinges on the following expressions:

$$\left(\frac{1}{2\sqrt{n}}\right)^{k_1+\dots+k_m} \frac{n^{m/2}}{n!} \text{tr}(X_1^{k_1} \dots X_m^{k_m}) = \left(\frac{1}{2}\right)^{k_1+\dots+k_m} \sum_{\mathbf{S}} n^{\frac{1}{2}(\chi(\mathbf{S})-m)} |\text{Cov}_{\mathbf{S}}(k_1, \dots, k_m)|,$$

In this expression, the X_i are certain matrices, such that e.g. $\lambda_j - j$ is an eigenvalue of X_1 , if (j, λ_j) is a corner of the Young diagram (i.e., $\lambda_j > \lambda_{j+1}$). Also $\text{Cov}_{\mathbf{S}}(k_1, \dots, k_m)$ are coverings of S^2 ramified according to a rule, given by the numbers k_1, \dots, k_m .

The right hand side relates to random matrices and has an interpretation as a “map” on a surface. It hinges on the following formula for $M \in U(n)$:

$$\left(\frac{1}{2\sqrt{n}}\right)^{k_1+\dots+k_m} E(\text{tr}(M^{k_1}) \dots \text{tr}(M^{k_m})) = \left(\frac{1}{2}\right)^{k_1+\dots+k_m} \sum_{\mathbf{S}} n^{\chi(\mathbf{S})-m} |\text{Map}_{\mathbf{S}}(k_1, \dots, k_m)|,$$

where the sum is taken over all the homeomorphism classes of orientable, not necessarily connected surfaces \mathbf{S} ; $\chi(\mathbf{S})$ is its Euler number and $\text{Map}_{\mathbf{S}}(k_1, \dots, k_m)$ denotes the ways to cover the surface \mathbf{S} with m polygons (a k_1 -gon, a k_2 -gon, ..., a k_m -gon) by pairwise gluing the sides of different or the same k_i -gons. Each polygon has a marked vertex, to distinguish a k -gon from its $k - 1$ rotations.

The main proposition, established by Okounkov and leading to Theorem 3.1, is the following: for connected surfaces, as $k_i \rightarrow \infty$,

$$\text{Cov}_{\mathbf{S}}(k_1, \dots, k_m) \cong \text{Map}_{\mathbf{S}}(k_1, \dots, k_m).$$

4. INTEGRALS, MOMENT MATRICES AND INTEGRABLE SYSTEMS

The first column contains a list of matrix integrals and a Fredholm determinant. After perturbing the integral by inserting times t_i 's and possibly s_i 's, the new integral thus obtained satisfies linear PDE's (Virasoro constraints). The integral can be represented as a determinant of a “moment” matrix, (defined by an appropriate inner-product) or a

Pfaffian, if the moment matrix is skew. Performing an appropriate ‘‘Borel factorization’’ of this associated moment matrix, one shows that each of these integrals, as a function of the t_i and s_i 's, satisfies an integrable lattice or an integrable PDE. In all the cases, combining both systems of equations leads to ODE's or PDE's for the corresponding integral, in x or in the boundary of E . The last column lists the connection with probability. For more information on the 6th integral and the double matrix integral, see [37, 3, 2].

τ -functions satisfying Virasoro constraints, after inserting t_i 's	= determinant or Pfaffian of m_ℓ of the form:	underlying integrable lattice or PDE	connecting with
$\int_{\mathcal{H}_\ell(E)} e^{-trV(M)} dM$	Hankel	Toda lattice	spectrum of random Hermitian matrices
$\int_{\mathcal{S}_\ell(E)} e^{-trV(M)} dM$	skew-symmetric	Pfaff lattice	spectrum of random symmetric matrices
$\int_{\mathcal{T}_\ell(E)} e^{-trV(M)} dM$	skew-symmetric	Pfaff lattice	spectrum of random symplectic matrices
$\int_{U(\ell)} e^{\sqrt{x} tr(M+\bar{M})} dM$	Toeplitz	Toeplitz lattice	longest increasing sequence in random permutations
$\int_{O(\ell)_\pm} e^{x trM} dM$	Hankel	Toda lattice	longest increasing sequence in random involutions
$\int_{U(\ell)} \det(I+M)^k e^{-xtr\bar{M}} dM$	Toeplitz	Toeplitz lattice	longest increasing sequence in random words
$\det(I - K(y, z)I_{(x, \infty)}(z))$ (K sine or Airy kernel)	Fredholm determinant	KdV equation	spectrum of infinite random Hermitian matrices (bulk or edge scaling limit)
$\int \int_{\mathcal{H}_\ell \times \mathcal{H}_\ell(E)} dM_1 dM_2$ $\times e^{-tr(M_1^2 + M_2^2 - cM_1 M_2)}$	general	2d-Toda lattice	coupled random matrices

REFERENCES

- [1] M. Adler and P. van Moerbeke: *Symmetric random matrices and the Pfaff Lattice*, to appear in: *Annals of Mathematics*, sept 2000 (solv-int/99-03009) and *The Hermitian, symmetric and symplectic ensembles and PDE's* (math.-phys./009 001).

- [2] M. Adler and P. van Moerbeke: *The spectrum of coupled random matrices*, Annals of Mathematics, **149**, 921–976 (1999).
- [3] M. Adler and P. van Moerbeke: *Integrals over classical groups, random permutations, Toda and Toeplitz lattices*, Comm. Pure Appl. Math. (2000) (math.CO/99-12143).
- [4] M. Adler, T. Shiota and P. van Moerbeke: *Random matrices, vertex operators and the Virasoro algebra*, Phys. Lett. **A 208**, 101-112, (1995). and *Random matrices, Virasoro algebras and non-commutative KP*, Duke Math. J. **94**, 379-431 (1998).
- [5] D. Aldous and P. Diaconis: *Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem*, Bull. Am. Math. Soc. (new series) **36** (4), 413–432 (1999).
- [6] J. Baik, P. Deift and K. Johansson: *On the distribution of the length of the longest increasing subsequence of random permutations*, Math. Archive, Journal Amer. Math. Soc. **12**, 1119-1178 (1999) (MathCO/9810105).
- [7] J. Baik and E. Rains: *Algebraic aspects of increasing subsequences*, Math.CO/9905083 – (1999).
- [8] D. Bessis, Cl. Itzykson and J.-B. Zuber : *Quantum field theory techniques in graphical enumeration*, Adv. Appl. Math. **1**, 109-157 (1980).
- [9] P. Biane: *Representations of symmetric groups and free probability*, preprint (1998).
- [10] M. Bowick and E. Brézin: *Universal scaling of the tail of the density of eigenvalues in random matrix models*, Phys. Letters **B 268**, 21–28 (1991).
- [11] S.Chadha, G. Mahoux, M.L. Mehta : *A method of integration over matrix variables : II*, J. Phys. A: Math. Gen. **14**, 579-586 (1981).
- [12] C.M. Cosgrove: *Chazy classes IX-XII of third-order differential equations*, Stud. Appl. Math. **104**, 3, 171-228 (2000).
- [13] P. Diaconis, M. Shashahani: *On the eigenvalues of random matrices* J. Appl. Prob., suppl. in honour of Takàcs **31A**, 49-61 (1994).
- [14] P. Erdős and G. Szekeres: *A combinatorial theorem in geometry*, Compositio Math., **2**, 463–470 (1935).
- [15] A.S. Fokas, A.R. Its, A.V. Kitaev: *The isomonodromy approach to matrix models in 2d quantum gravity*, Comm. Math. Phys., **147**, 395–430 (1992).
- [16] P.J. Forrester: *The spectrum edge of random matrix ensembles*, Nucl. Phys. B, (1993).
- [17] I. M. Gessel: *Symmetric functions and P-recursiveness*, J. of Comb. Theory, Ser A, **53**, 257–285 (1990).
- [18] J.M. Hammersley: *A few seedlings of research*, Proc. Sixth. Berkeley Symp. Math. Statist. and Probability, Vol. 1, 345–394, University of California Press (1972).
- [19] M. Hisakado: *Unitary matrix models and Painlevé III*, Mod. Phys. Letters, **A 11** 3001–3010 (1996).

- [20] Cl. Itzykson, J.-B. Zuber: *The planar approximation*, J. Math. Phys. **21**, 411–421 (1980).
- [21] M. Jimbo, T. Miwa, Y. Mori and M. Sato: *Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent*, Physica 1D, 80–158 (1980).
- [22] K. Johansson: *The Longest Increasing Subsequence in a Random Permutation and a Unitary Random Matrix Model*, Math. Res. Lett., **5**, no. 1-2, 63–82 (1998).
- [23] B.F. Logan and L.A. Shepp: *A variational problem for random Young tableaux*, Advances in Math., **26**, 206–222 (1977).
- [24] R.D. Kamien, H.D. Politzer, M.B. Wise: *Universality of random-matrix predictions for the statistics of energy levels* Phys. rev. letters **60**, 1995–1998 (1988).
- [25] M. Kontsevich: *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. **147**, 1-23 (1992).
- [26] D. Knuth: “The Art of Computer programming, Vol III: Searching and Sorting”, Addison-Wesley, Reading, MA, 1973.
- [27] G. Mahoux, M.L. Mehta: *A method of integration over matrix variables: IV*, J. Phys. I (France) **1**, 1093–1108 (1991).
- [28] M.L. Mehta: Random matrices, 2nd ed. Boston: Acad. Press, 1991.
- [29] T. Nagao, M. Wadati: *Correlation functions of random matrix ensembles related to classical orthogonal polynomials*, J. Phys. Soc of Japan, **60** 3298–3322 (1991).
- [30] A.M. Odlyzko: *On the distribution of spacings between the zeros of the zeta function* **48**, 273-308 (1987).
- [31] A. Okounkov: *Random matrices and random permutations* – Math.CO/99-03176 (1999).
- [32] L.A. Pastur: *On the universality of the level spacing distribution for some ensembles of random matrices*, Letters Math. Phys., **25** 259–265 (1992).
- [33] E. M. Rains: *Topics in Probability on compact Lie groups*, Harvard University doctoral dissertation, (1995).
- [34] E. M. Rains: *Increasing subsequences and the classical groups*, Elect. J. of Combinatorics, **5**, R12 (1998).
- [35] C.A. Tracy and H. Widom: *Level-Spacings distribution and the Airy kernel*, Commun. Math. Phys., **159**, 151–174 (1994).
- [36] C.A. Tracy and H. Widom: *Random unitary matrices, permutations and Painlevé*, math.CO /9811154, (1999).
- [37] C.A. Tracy and H. Widom: *On the distribution of the lengths of the longest monotone subsequences in random words*, math.CO/9904042, (1999).
- [38] S.M. Ulam: *Monte Carlo calculations in problems of mathematical physics*, in *Modern Mathematics for the Engineers*, E.F. Beckenbach ed., McGraw-Hill, 261–281 (1961).

- [39] P. van Moerbeke: *Integrable lattices: random matrices and permutations*, MSRI-volume on Random matrices and exactly solvable models, Eds.: P. Bleher, A. Its, Oxford University press, 2000 (arXiv: Math.CO 00-10).
- [40] A.M. Vershik and S.V. Kerov: *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux*, Soviet Math. Dokl., **18**, 527–531 (1977).

Pierre VAN MOERBEKE

Departement de Mathématiques

Université de Louvain

B-1348 Louvain-la-Neuve (Belgium)

E-mail : `vanmoerbeke@geom.ucl.ac.be`

and

Brandeis University

Waltham, Mass 02454, USA

E-mail : `vanmoerbeke@math.brandeis.edu`

FACTORISATION FAIBLE DES APPLICATIONS BIRATIONNELLES
[d'après Abramovich, Karu, Matsuki, Włodarczyk et Morelli]

par **Laurent BONAVERO**

1. INTRODUCTION

Dans tout ce texte, \mathbf{K} désigne un corps algébriquement clos de caractéristique nulle (la nullité de la caractéristique est essentielle dans les parties 2 et 5, les constructions des parties 3 et 4 sont cependant valables en caractéristique arbitraire).

1.1. Rappels sur les éclatements

Si $n \geq 2$ et si V est un \mathbf{K} -espace vectoriel de dimension n , on note $\mathbf{P}_{\mathbf{K}}(V)$ l'espace projectif des droites vectorielles de V . Si $V = \mathbf{K}^n$, on note simplement $\mathbf{P}^{n-1} = \mathbf{P}_{\mathbf{K}}(V)$. Considérons

$$B_0(\mathbf{K}^n) = \{(v, d) \in \mathbf{K}^n \times \mathbf{P}^{n-1} \mid v \in d\}.$$

Si (x_1, \dots, x_n) sont les coordonnées sur \mathbf{K}^n et $[y_1 : \dots : y_n]$ les coordonnées homogènes associées sur \mathbf{P}^{n-1} , alors

$$B_0(\mathbf{K}^n) = \{(x, y) \in \mathbf{K}^n \times \mathbf{P}^{n-1} \mid \forall i, j, x_i y_j = x_j y_i\}.$$

Il en découle que $B_0(\mathbf{K}^n)$ est une sous-variété algébrique fermée lisse de $\mathbf{K}^n \times \mathbf{P}^{n-1}$. La première projection $\pi : B_0(\mathbf{K}^n) \rightarrow \mathbf{K}^n$ est une application régulière birationnelle qui se restreint en un isomorphisme $\pi : B_0(\mathbf{K}^n) \setminus \pi^{-1}(0) \rightarrow \mathbf{K}^n \setminus \{0\}$ avec $\pi^{-1}(0) = \{0\} \times \mathbf{P}^{n-1}$. L'application birationnelle $\pi : B_0(\mathbf{K}^n) \rightarrow \mathbf{K}^n$ s'appelle *l'éclatement de \mathbf{K}^n en 0* ; 0 est le *centre* de π et $\pi^{-1}(0)$ est le *diviseur exceptionnel* de π .

Plus généralement (voir par exemple [Har77]), si Y est une sous-variété fermée lisse d'une variété algébrique lisse X , il y a une variété algébrique lisse $B_Y(X)$ et une application régulière birationnelle $\pi : B_Y(X) \rightarrow X$ qui se restreint en un isomorphisme $\pi : B_Y(X) \setminus \pi^{-1}(Y) \rightarrow X \setminus Y$ et $\pi^{-1}(Y) \simeq \mathbf{P}(N_{Y/X})$ où $N_{Y/X}$ désigne le fibré normal de Y dans X . L'application birationnelle $\pi : B_Y(X) \rightarrow X$ s'appelle *l'éclatement de X le long de Y , ou de centre Y* et $E := \pi^{-1}(Y)$ est le *diviseur exceptionnel* de π . Moralement, on remplace chaque point y de Y par l'espace projectif des directions normales à Y dans X .

passant par y . Si la donnée initiale est $B_Y(X)$, on dit encore que $\pi : B_Y(X) \rightarrow X$ est une *contraction* de centre Y (éclatement et contraction désignent donc *dans ce texte* la même application birationnelle, ils sont utilisés respectivement suivant que la donnée initiale est X (resp. $B_Y(X)$) et la donnée finale $B_Y(X)$ (resp. X)). Enfin, si Z est une sous-variété de X , on appelle *transformée stricte* de Z l'adhérence dans $B_Y(X)$ de $\pi^{-1}(Z) \setminus E$.

1.2. Énoncé du théorème principal

C'est un problème classique depuis une trentaine d'années de savoir s'il est possible de décomposer une application birationnelle entre deux variétés algébriques complètes lisses en une suite d'éclatements et contractions de centres lisses. En dimension 1, la question est vide : toute application birationnelle entre deux courbes algébriques complètes lisses est un isomorphisme. Dans le cas des surfaces, on sait depuis un siècle que toute application birationnelle entre surfaces complètes lisses est une suite d'éclatements et contractions de centre des points (voir par exemple [BPV84]) : un exemple classique est la transformation de Cremona $[x : y : z] \rightarrow [1/x : 1/y : 1/z]$ de \mathbf{P}^2 dans \mathbf{P}^2 qui se décompose en l'éclatement des trois points $[0 : 0 : 1]$, $[0 : 1 : 0]$ et $[1 : 0 : 0]$ suivie de la contraction des transformées strictes des trois droites joignant deux à deux les points précédents. Le problème était ouvert dès la dimension trois.

Nous donnons dans cet exposé les grandes lignes de la démonstration du théorème suivant, dû à Abramovich, Karu, Matsuki, Włodarczyk [AKMW99] et Włodarczyk [Wlod99].

THÉORÈME 1.1. — *Soit $\varphi : X_1 \dashrightarrow X_2$ une application birationnelle entre deux variétés algébriques complètes et lisses X_1 et X_2 sur \mathbf{K} . Alors, φ se factorise en une suite d'éclatements et de contractions de centres lisses. Autrement dit, il y a une suite d'applications birationnelles entre variétés algébriques complètes et lisses*

$$X_1 = V_0 \dashrightarrow K V_1 \dashrightarrow K \cdots \dashrightarrow K V_i \dashrightarrow K V_{i+1} \dashrightarrow K \cdots \dashrightarrow K V_{l-1} \dashrightarrow K V_l = X_2$$

de sorte que $\varphi = \varphi_{l-1} \circ \varphi_{l-2} \circ \cdots \circ \varphi_1 \circ \varphi_0$ et pour tout i , $\varphi_i : V_i \dashrightarrow V_{i+1}$ ou $\varphi_i^{-1} : V_{i+1} \dashrightarrow V_i$ est une application régulière obtenue en éclatant une sous-variété irréductible lisse.

Ce texte est pour l'essentiel une reprise de [AKMW99] et [Mat99], dans lesquels on trouvera beaucoup plus de précisions, une liste de références très complète ainsi qu'une discussion détaillée des extensions ou généralisations du théorème 1.1.

Mentionnons aussi que le théorème de factorisation admet les raffinements fondamentaux suivants :

- si φ est un isomorphisme sur un ouvert U , le centre de chaque φ_i ou φ_i^{-1} peut être choisi disjoint de U ,
- si X_1 et X_2 sont des variétés projectives, chaque V_i peut être choisie projective,
- si X_1 et X_2 ne sont pas supposées complètes, toute application birationnelle propre (voir [Har77] pour cette notion) entre X_1 et X_2 se factorise en une suite d'éclatements et de contractions de centres lisses,

- si X_1 et X_2 sont deux variétés analytiques complexes compactes lisses, toute application biméromorphe propre entre X_1 et X_2 se factorise en une suite d'éclatements et de contractions de centres analytiques lisses,
- le théorème 1.1 est vrai même si \mathbf{K} n'est pas supposé algébriquement clos.

Pour simplifier l'exposition, nous nous contenterons de renvoyer à [AKMW99] et [Mat99] pour une démonstration de ces points.

Faisons aussi les remarques suivantes :

- la factorisation n'est évidemment pas unique (même si l'on impose que pour tout i , φ_i ne soit pas l'inverse de φ_{i-1}),
- dans le cas des surfaces, si φ est régulière, alors toutes les φ_i peuvent être choisies régulières. Ceci est faux en dimension supérieure ou égale à trois,
- la factorisation décrite dans le théorème 1.1 est une solution positive au problème de factorisation faible ; le problème de factorisation forte (au sens où il existe un entier i_0 tel que pour tout $i \leq i_0$, φ_i^{-1} est régulière et pour tout $i \geq i_0 + 1$, φ_i est régulière) est un problème ouvert à ce jour en dimension supérieure ou égale à trois.

1.3. Quelques mots sur la démonstration du théorème 1.1

La démonstration du théorème 1.1 est en certain sens une victoire de la géométrie torique. En effet, la ligne directrice est une réduction (en plusieurs étapes, certaines utilisant fondamentalement des techniques toriques) au cas d'une application birationnelle équivariante entre variétés toriques, cas résolu par Morelli [Mor96] et Włodarczyk [Wlo97], puis étendu au cadre toroïdal par Abramovich, Matsuki et Rashid [AMR99]. Il est bien connu depuis quelques années que la géométrie des variétés toriques (ou des plongements toroïdaux), gouvernée par des objets combinatoires simples issus de la géométrie convexe, donne une bonne vision locale de certaines propriétés des variétés algébriques (voir par exemple l'existence d'altérations due à De Jong). L'exposé qui suit donnera, je l'espère, envie au lecteur de mieux connaître ou de découvrir ces techniques ; ce texte ne remplaçant certainement pas l'énorme effort pédagogique que constitue le texte [Mat99] de K. Matsuki.

1.4. Remerciements

Merci aux collègues qui m'écoutent parler d'éclatements depuis plusieurs années, ils sont trop nombreux pour être tous mentionnés. Merci à Michel Brion, Laurent Manivel, Kenji Matsuki et Emmanuel Peyre pour m'avoir aidé à préparer ce texte, et plus particulièrement à Stéphane Guillermou, infatigable relecteur de multiples versions préliminaires, ainsi qu'à Cinzia Casagrande pour m'avoir signalé les trop nombreuses coquilles de la version distribuée le jour de l'exposé.

Je dédie ce travail à mon fils Alex, et à Marguerite sa maman.

2. COBORDISME BIRATIONNEL ET ACTION DE \mathbf{K}^*

La notion de cobordisme birationnel a été dégagée par Włodarczyk [Wlo97], suite au travail fondamental de Morelli [Mor96] dans le cadre des variétés toriques. L'idée essentielle est la suivante : la théorie de Morse sur les variétés (différentiables réelles) permet de reconstruire topologiquement une variété donnée X à partir d'une fonction de Morse f sur X . Le passage des points critiques de f (qui sont aussi les points fixes du champ de vecteurs $\text{grad}(f)$ lorsque X est munie d'une métrique riemannienne) correspond aux changements de topologie par ajout d'une cellule. À un morphisme birationnel entre X_1 et X_2 de dimension n , on associe une variété algébrique de dimension $n + 1$ munie d'une action de \mathbf{K}^* , avec un ordre sur les composantes connexes de points fixes. L'action de \mathbf{K}^* joue le rôle du champ de vecteurs $\text{grad}(f)$ et à chaque composante connexe de \mathbf{K}^* -points fixes, on associe une application birationnelle élémentaire, dont on montre qu'elle est "localement torique".

2.1. Rappels sur les \mathbf{K}^* -actions

Lorsque le groupe multiplicatif \mathbf{K}^* agit algébriquement sur une variété algébrique X , pour $t \in \mathbf{K}^*$ et $x \in X$, le résultat de l'action de t sur x sera noté $t \cdot x$. On note X/\mathbf{K}^* l'ensemble des orbites de l'action et $X^{\mathbf{K}^*}$ l'ensemble des \mathbf{K}^* -points fixes de X .

Définition. Un *bon quotient* ou *quotient catégorique* $Y = X//\mathbf{K}^*$ est la donnée d'une variété algébrique Y et d'une application régulière $\pi : X \rightarrow Y$ constante sur les \mathbf{K}^* -orbites de sorte que pour tout ouvert affine U de Y , $\pi^{-1}(U)$ est un ouvert affine de X et l'application induite $\pi^* : \mathcal{O}_Y(U) \rightarrow (\mathcal{O}_X(\pi^{-1}(U)))^{\mathbf{K}^*}$ est un isomorphisme (ici, $(\mathcal{O}_X(\pi^{-1}(U)))^{\mathbf{K}^*}$ désigne les éléments \mathbf{K}^* -invariants de $\mathcal{O}_X(\pi^{-1}(U))$). Si de plus les fibres de π sont exactement les orbites, alors Y est appelé *quotient géométrique*.

Rappelons (voir par exemple [MFK94] ou [Dol94]) que \mathbf{K}^* étant un groupe réductif, pour toute variété affine X munie d'une action algébrique de \mathbf{K}^* , le quotient catégorique $X//\mathbf{K}^*$ existe et ses points correspondent aux \mathbf{K}^* -orbites fermées. De plus $X//\mathbf{K}^*$ est normale si X l'est.

Notations. Soit X une variété algébrique sur laquelle \mathbf{K}^* agit algébriquement. Introduisons les deux sous-ensembles localement fermés de X suivant :

$$X_+ = \{x \in X \mid \lim_{t \rightarrow \infty} t \cdot x \text{ n'existe pas dans } X\}$$

et

$$X_- = \{x \in X \mid \lim_{t \rightarrow 0} t \cdot x \text{ n'existe pas dans } X\}.$$

Précisons ici que " $\lim_{t \rightarrow 0} t \cdot x$ existe dans X " signifie que l'application régulière de \mathbf{K}^* dans X qui à t associe $t \cdot x$ s'étend en une application régulière de \mathbf{K} dans X ayant $\lim_{t \rightarrow 0} t \cdot x$ pour valeur en 0.

Premier exemple fondamental. Considérons l'action algébrique de \mathbf{K}^* sur $X = \mathbf{K}^{n+1}$ définie par $t \cdot (x_1, \dots, x_{n+1}) = (t^{a_1}x_1, \dots, t^{a_{n+1}}x_{n+1})$ où les a_i sont des entiers premiers entre eux tels que $a_i < 0$ pour $1 \leq i \leq \alpha$ et $a_i > 0$ pour $\alpha + 1 \leq i \leq n + 1$ pour un entier α tel que $2 \leq \alpha \leq n$. Alors $X_+ = X \setminus (\mathbf{K}^\alpha \times \{0\})$ et $X_- = X \setminus (\{0\} \times \mathbf{K}^{n+1-\alpha})$. Les quotients géométriques X_+/\mathbf{K}^* et X_-/\mathbf{K}^* existent, ainsi que le quotient catégorique $X//\mathbf{K}^*$ et on a un diagramme commutatif :

$$\begin{array}{ccc}
 X_-/\mathbf{K}^* & \overset{\varphi}{\dashrightarrow} & X_+/\mathbf{K}^* \\
 \downarrow \text{JJJJ} & & \downarrow \text{tttt} \\
 X//\mathbf{K}^* & & X//\mathbf{K}^*
 \end{array}$$

Si $\underline{0} \in X//\mathbf{K}^*$ désigne l'unique \mathbf{K}^* -orbite fermée de X , la fibre $\varphi_+^{-1}(\underline{0})$ (resp. $\varphi_-^{-1}(\underline{0})$) est isomorphe à l'espace projectif à poids $\mathbf{P}(a_{\alpha+1}, \dots, a_{n+1})$ (resp. $\mathbf{P}(a_1, \dots, a_\alpha)$). De plus, φ_+ et φ_- se restreignent en des isomorphismes sur l'ouvert $(X_- \cap X_+)/\mathbf{K}^*$, ainsi $\varphi = \varphi_+^{-1} \circ \varphi_- : X_-/\mathbf{K}^* \rightarrow X_+/\mathbf{K}^*$ est une application birationnelle.

2.2. Cobordisme birationnel

La définition suivante sera l'outil fondamental dans toute la suite.

Définition. Soit $\varphi : X_1 \rightarrow X_2$ une application birationnelle entre deux variétés algébriques normales X_1 et X_2 sur \mathbf{K} . Un *cobordisme birationnel* pour φ est la donnée d'une variété algébrique normale B telle que :

- (i) le groupe multiplicatif \mathbf{K}^* agit de façon effective sur B (i.e. $\bigcap_{x \in B} \text{Stab}(x) = 1$),
- (ii) les ensembles B_- et B_+ sont des ouverts (de Zariski) non vides,
- (iii) les quotients géométriques B_-/\mathbf{K}^* et B_+/\mathbf{K}^* existent et sont respectivement isomorphes à X_1 et X_2 ,
- (iv) si $\psi : B_- \rightarrow B_+$ est l'application birationnelle induite par les inclusions $B_- \cap B_+ \subset B_-$ et $B_- \cap B_+ \subset B_+$, le diagramme suivant est commutatif :

$$\begin{array}{ccc}
 B_- & \overset{\psi}{\dashrightarrow} & B_+ \\
 \downarrow \text{ii} & & \downarrow \text{ii} \\
 X_1 & \overset{\varphi}{\dashrightarrow} & X_2
 \end{array}$$

Dans le premier exemple fondamental, $X = \mathbf{K}^{n+1}$ est un cobordisme birationnel pour $\varphi : X_-/\mathbf{K}^* \rightarrow X_+/\mathbf{K}^*$.

Deuxième exemple fondamental (voir aussi [Ful84]).

Soient X une variété algébrique complète et lisse, Y une sous-variété irréductible lisse de X et $\varphi : X_1 \rightarrow X_2 = X$ l'éclatement de X de centre Y . Construisons un cobordisme birationnel pour φ . Soit W la variété algébrique $X_2 \times \mathbf{P}^1$ sur laquelle le groupe \mathbf{K}^* agit algébriquement par multiplication sur le second facteur. Notons $\overline{Y} = Y \times \{0\} \subset W$ et soit \overline{B} la variété algébrique complète et lisse obtenue en éclatant W le long de \overline{Y} . Comme

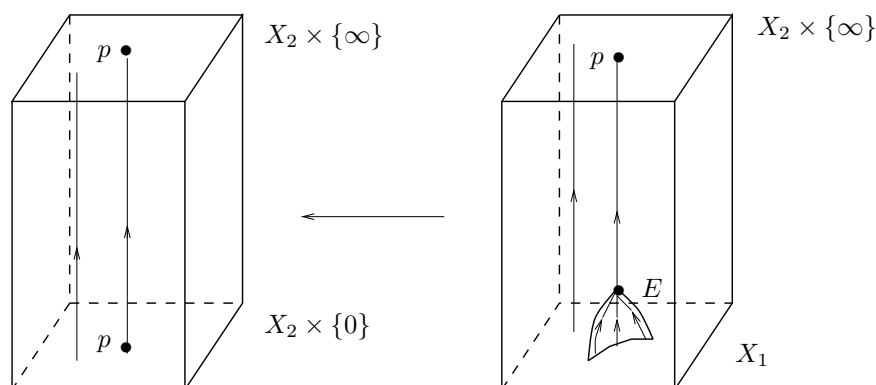


FIGURE 1. Cobordisme birationnel : éclatement d'un point

\bar{Y} est inclus dans l'ensemble des \mathbf{K}^* -points fixes de W , l'action de \mathbf{K}^* sur W se relève en une action sur \bar{B} . La transformée stricte D_1 de $X_2 \times \{0\}$ est isomorphe à X_1 si bien que \bar{B} possède deux diviseurs constitués de \mathbf{K}^* -points fixes, l'un, D_1 , isomorphe à X_1 et l'autre, noté D_2 et égal à l'image inverse dans \bar{B} de $X_2 \times \{\infty\}$, isomorphe à X_2 . Posons alors $B = \bar{B} \setminus (D_1 \cup D_2)$. Soit E le diviseur exceptionnel de l'éclatement $\bar{B} \rightarrow W$. Ce diviseur est isomorphe à $\mathbf{P}(N_{\bar{Y}/W}) = \mathbf{P}(N_{Y/X_2} \oplus \underline{0})$ et est \mathbf{K}^* -invariant ($\underline{0}$ désigne le fibré en droites trivial). Remarquons que l'ensemble des \mathbf{K}^* -points fixes de B correspond à l'image de $\mathbf{P}(N_{X_2 \times \{0\}/W|_{\bar{Y}}})$ dans l'identification précédente et est donc naturellement isomorphe à Y . C'est ensuite un exercice facile de voir que B est un cobordisme birationnel pour φ . La figure 1 représente le cas de l'éclatement d'un point dans une surface et devrait éclairer le lecteur.

2.3. Construction de cobordisme birationnel

On démontre ici le résultat suivant, dû à Włodarczyk [Wlo00] :

THÉORÈME 2.1. — *Soit $\varphi : X_1 \rightarrow X_2$ une application régulière birationnelle entre deux variétés projectives lisses. Alors, il existe une variété projective lisse \bar{B} munie d'une action algébrique effective de \mathbf{K}^* vérifiant :*

- (i) *il existe deux plongements $\iota_1 : X_1 \rightarrow \bar{B}^{\mathbf{K}^*}$ et $\iota_2 : X_2 \rightarrow \bar{B}^{\mathbf{K}^*}$ d'images disjointes,*
- (ii) *la variété $B = \bar{B} \setminus (\iota_1(X_1) \cup \iota_2(X_2))$ est un cobordisme birationnel pour φ .*

Démonstration. Elle suit de près la construction du deuxième exemple fondamental. Comme $\varphi : X_1 \rightarrow X_2$ est une application régulière birationnelle entre deux variétés projectives, il existe un faisceau d'idéaux $\mathcal{I} \subset \mathcal{O}_{X_2}$ cohérent tel que φ soit l'éclatement de \mathcal{I} (voir par exemple [Har77] page 166). Soit W la variété algébrique $X_2 \times \mathbf{P}^1$ sur laquelle le groupe \mathbf{K}^* agit algébriquement par multiplication sur le second facteur. Soit $\mathcal{J} = (p_1^{-1}\mathcal{I} + p_2^{-1}\mathcal{I}_0)\mathcal{O}_W$ où \mathcal{I}_0 est l'idéal du point 0 de \mathbf{P}^1 et soit $\pi : \tilde{W} \rightarrow W$ l'éclatement de \mathcal{J} . Comme \mathcal{J} est \mathbf{K}^* -invariant, l'action de \mathbf{K}^* sur W se relève à \tilde{W} . La variété \tilde{W} est projective et en général singulière. La transformée stricte D_1 de $X_2 \times \{0\}$ est isomorphe à

X_1 et, étant lisse, de Cartier et d'équation locale d'ordre 1, est incluse dans le lieu régulier de \tilde{W} . Soit alors \overline{B} une désingularisation canonique de \tilde{W} (une telle désingularisation est obtenue par une suite d'éclatements le long de centres lisses disjoints du lieu régulier de \tilde{W} et est naturellement munie d'une action algébrique de \mathbf{K}^* relevant l'action de \mathbf{K}^* sur W ; de telles désingularisations existent, voir la partie 5 pour plus de détails). Alors, la pré-image D'_1 de D_1 dans \overline{B} est isomorphe à X_1 , la pré-image D'_2 de $X_2 \times \{\infty\}$ dans \overline{B} est isomorphe à X_2 et $B = \overline{B} \setminus (D'_1 \cup D'_2)$ est un cobordisme birationnel pour φ . ■

2.4. Filtrabilité

La notion de cobordisme filtrable est due à Morelli [Mor96] et Włodarczyk [Wlo97], son origine se situe dans les travaux de Bialynicki-Birula.

Définition. Soit B une variété algébrique munie d'une action algébrique de \mathbf{K}^* . Si F_1 et F_2 sont deux composantes connexes de $B^{\mathbf{K}^*}$, on dit que F_1 précède F_2 , que l'on note $F_1 \prec F_2$, s'il existe $x \in B \setminus B^{\mathbf{K}^*}$ tel que $\lim_{t \rightarrow 0} t \cdot x \in F_1$ et $\lim_{t \rightarrow \infty} t \cdot x \in F_2$.

La relation \prec n'est en général pas transitive.

Définition. Soit B une variété algébrique munie d'une action algébrique de \mathbf{K}^* . On dit que B est *filtrable* s'il n'y a pas de \prec -cycle de composantes connexes de $B^{\mathbf{K}^*}$

$$F_1 \prec F_2 \prec \cdots \prec F_m \prec F_1.$$

En particulier, il n'y a pas de composante connexe F de $B^{\mathbf{K}^*}$ telle que $F \prec F$.

Le lemme suivant est élémentaire mais essentiel :

LEMME 2.2. — Soit B une variété algébrique lisse munie d'une action algébrique de \mathbf{K}^* . Si B est quasi-projective, alors B est filtrable.

Démonstration. Par un résultat de Sumihiro [Sum74] [Sum75], il existe une immersion localement fermée et équivariante de B dans un espace projectif $\mathbf{P}(V)$ où V est un espace vectoriel sur lequel \mathbf{K}^* agit linéairement et algébriquement. Comme l'action de \mathbf{K}^* est algébrique, V se décompose en espaces propres $V = \bigoplus_{k=0}^m V(a_k)$ où les a_k sont des entiers relatifs ordonnés $a_0 < \cdots < a_m$ et \mathbf{K}^* agit sur $x_k \in V(a_k)$ par $t \cdot x_k = t^{a_k} x_k$ (les a_k sont les *poinds* de la représentation V). L'observation suivante est élémentaire mais cruciale : soit $x \in \mathbf{P}(V)$ que l'on relève en $\bar{x} = x_0 \oplus \cdots \oplus x_m$ dans V . Soit $\min(x)$ (resp. $\max(x)$) le plus petit (resp. grand) indice i dans $\{0, \dots, m\}$ tel que $x_i \neq 0$. Alors $\lim_{t \rightarrow 0} t \cdot x$ (resp. $\lim_{t \rightarrow \infty} t \cdot x$) est l'image dans $\mathbf{P}(V)$ de $x_{\min(x)}$ (resp. $x_{\max(x)}$).

Les composantes connexes de $(\mathbf{P}(V))^{\mathbf{K}^*}$ sont les $C_{a_k} = \mathbf{P}(V(a_k))$ et le lemme repose sur le fait suivant : s'il existe x dans $\mathbf{P}(V) \setminus (\mathbf{P}(V))^{\mathbf{K}^*}$ tel que $\lim_{t \rightarrow 0} t \cdot x \in C_{a_i}$ et $\lim_{t \rightarrow \infty} t \cdot x \in C_{a_j}$, alors, d'après l'observation précédente, $a_i < a_j$. De là, si F est une composante connexe de $B^{\mathbf{K}^*}$ et si $a(F)$ est l'unique entier tel que $F \subset C_{a(F)}$, on déduit que $F \prec F'$ implique $a(F) < a(F')$. Il s'ensuit que B est filtrable. ■

2.5. Décomposition d'un cobordisme birationnel et théorie géométrique des invariants

Soit $\varphi : X_1 \rightarrow X_2$ une application régulière birationnelle entre deux variétés projectives lisses et soit B un cobordisme birationnel quasi-projectif (et donc filtrable d'après ce qui précède) pour φ donné par le théorème 2.1. Dans le paragraphe précédent, on a associé à B une suite de poids entiers ordonnés $(a_i)_{i=0, \dots, m}$, et quitte à remplacer V par $\text{Sym}^2(V)$, on peut supposer de plus que tous les poids a_i sont pairs, en particulier $a_i < a_i + 1 < a_{i+1}$. Rappelons aussi qu'à toute composante connexe F de $B^{\mathbf{K}^*}$ correspond l'un de ces poids $a(F)$. Si x appartient à $B^{\mathbf{K}^*}$, x appartient à une unique composante connexe F de $B^{\mathbf{K}^*}$ et on note $a(x)$ le poids $a(F)$ correspondant.

Notations. Soit a_i l'un des poids précédents. On note B_{a_i} le complémentaire dans B de

$$\{x \mid x_\infty = \lim_{t \rightarrow \infty} t \cdot x \text{ existe dans } B \text{ et } a(x_\infty) < a_i\} \cup \\ \{x \mid x_0 = \lim_{t \rightarrow 0} t \cdot x \text{ existe dans } B \text{ et } a(x_0) > a_i\}.$$

Remarquons que chaque B_{a_i} contient l'ouvert $B_- \cap B_+$ et que si x est un point fixe dans B_{a_i} , alors $a(x) = a_i$. En particulier, d'après le paragraphe 2.4, il n'existe pas de point x dans $B_{a_i} \setminus B_{a_i}^{\mathbf{K}^*}$ tel que $\lim_{t \rightarrow 0} t \cdot x$ et $\lim_{t \rightarrow \infty} t \cdot x$ existent dans B_{a_i} . Cette dernière observation joue un rôle crucial dans la construction de "l'atlas torique" de la partie 2.6.

Le lemme suivant est immédiat :

LEMME 2.3. — Avec les notations précédentes, $(B_{a_i})_+ = (B_{a_{i+1}})_-$ pour tout $0 \leq i \leq m - 1$. De plus, $B_- = (B_{a_0})_-$ et $B_+ = (B_{a_m})_+$.

Exemple. Considérons $\mathbf{P}(V)$ où V est un espace vectoriel sur lequel \mathbf{K}^* agit linéairement et algébriquement. Comme précédemment, V se décompose en espaces propres $V = \bigoplus_{k=0}^m V(a_k)$ où les a_k sont les poids de V supposés tels que $a_k < a_k + 1 < a_{k+1}$ pour tout k . Considérons $B = \mathbf{P}(V) \setminus (\mathbf{P}(V(a_0)) \cup \mathbf{P}(V(a_m)))$. Alors, il est très facile de voir que

$$B_{a_i} = \{x \in B \mid \min(x) \leq i \leq \max(x)\}, \quad (B_{a_i})_+ = \{x \in B \mid \min(x) \leq i \leq i + 1 \leq \max(x)\}$$

et

$$(B_{a_i})_- = \{x \in B \mid \min(x) \leq i - 1 \leq i \leq \max(x)\}.$$

Le résultat suivant est dû à Włodarczyk [Wlo00], il voit son origine dans les travaux de Guillemin-Sternberg [GuS89] et Brion-Procesi [BrP90] :

PROPOSITION 2.4. — *Le quotient catégorique $B_{a_i} // \mathbf{K}^*$ et les quotients géométriques $(B_{a_i})_- / \mathbf{K}^*$ et $(B_{a_i})_+ / \mathbf{K}^*$ existent. De plus, B_{a_i} est un cobordisme birationnel pour*

$$\varphi_i : (B_{a_i})_- / \mathbf{K}^* \dashrightarrow (B_{a_i})_+ / \mathbf{K}^*.$$

Démonstration. Les notations étant celles de la démonstration du lemme 2.2, soit ρ_0 l'action de \mathbf{K}^* sur $V = \bigoplus_{k=0}^m V(a_k)$ et pour un entier $r \in \mathbf{Z}$, soit ρ_r l'action de \mathbf{K}^* obtenue en "tordant" ρ_0 par t^{-r} . Autrement dit, si $x_k \in V(a_k)$, on a $\rho_r(t) \cdot x_k = t^{a_k - r} x_k$. Évidemment, l'action de ρ_r sur $\mathbf{P}(V)$ est égale à l'action initiale ρ_0 , mais ρ_r induit un changement de linéarisation sur le \mathbf{K}^* -fibré en droites ample $\mathcal{O}_{\mathbf{P}(V)}(1)$. Sous l'hypothèse que les poids de l'action vérifient $a_i + 1 < a_{i+1}$, on observe que B_{a_i} (resp. $(B_{a_i})_-$, resp. $(B_{a_i})_+$) est le lieu des points semi-stables de B pour le fibré en droites $\mathcal{O}_{\mathbf{P}(V)}(1)$ linéarisé par ρ_{a_i} (resp. ρ_{a_i-1} , resp. ρ_{a_i+1}) - rappelons que x est un point semi-stable d'une variété algébrique munie d'une action de \mathbf{K}^* et d'un fibré en droites L ample et linéarisé par ρ_r si et seulement si il existe une section invariante de $L^{\otimes n}$ non nulle en x pour un certain entier n . La théorie géométrique des invariants [MFK94] [Dol94] assure alors que les quotients catégoriques $B_{a_i} // \mathbf{K}^*$, $(B_{a_i})_- // \mathbf{K}^*$ et $(B_{a_i})_+ // \mathbf{K}^*$ existent, et comme de plus chaque orbite de $(B_{a_i})_-$ (resp. $(B_{a_i})_+$) est fermée dans $(B_{a_i})_-$ (resp. $(B_{a_i})_+$), les quotients $(B_{a_i})_- // \mathbf{K}^*$ et $(B_{a_i})_+ // \mathbf{K}^*$ sont des quotients géométriques. De plus, le quotient catégorique $\pi_i : B_{a_i} \rightarrow B_{a_i} // \mathbf{K}^*$ est un morphisme affine : ceci découle du fait déjà observé qu'il n'existe pas de point x dans $B_{a_i} \setminus B_{a_i}^{\mathbf{K}^*}$ tel que $\lim_{t \rightarrow 0} t \cdot x$ et $\lim_{t \rightarrow \infty} t \cdot x$ existent dans B_{a_i} . Notons enfin qu'il y a un diagramme commutatif :

$$\begin{array}{ccc}
 (B_{a_i})_- // \mathbf{K}^* & \overset{\varphi_i}{\dashrightarrow} & (B_{a_i})_+ // \mathbf{K}^* \\
 \downarrow \text{M} & & \downarrow \text{M} \\
 (\varphi_i)_- & & (\varphi_i)_+ \\
 \downarrow \text{K} & & \downarrow \text{K} \\
 & B_{a_i} // \mathbf{K}^* &
 \end{array}$$

si bien que le cobordisme birationnel $\varphi_i : (B_{a_i})_- // \mathbf{K}^* \dashrightarrow (B_{a_i})_+ // \mathbf{K}^*$ s'interprète comme un changement de linéarisation pour la restriction à B_{a_i} du \mathbf{K}^* -fibré en droites ample $\mathcal{O}_{\mathbf{P}(V)}(1)$. ■

Faisons le point : si $\varphi : X_1 \rightarrow X_2$ est une application régulière birationnelle entre deux variétés projectives lisses, ce qui précède montre que φ se factorise en une suite d'applications birationnelles

$$X_1 = V_0 \overset{\varphi_0}{\dashrightarrow} K V_1 \overset{\varphi_1}{\dashrightarrow} K \cdots \overset{\varphi_{i-1}}{\dashrightarrow} K V_i \overset{\varphi_i}{\dashrightarrow} K V_{i+1} \overset{\varphi_{i+1}}{\dashrightarrow} K \cdots \overset{\varphi_{m-1}}{\dashrightarrow} K V_m \overset{\varphi_m}{\dashrightarrow} K V_{m+1} = X_2$$

où chaque $\varphi_i : V_i \dashrightarrow K V_{i+1}$ est un cobordisme birationnel $\varphi_i : (B_{a_i})_- // \mathbf{K}^* \dashrightarrow (B_{a_i})_+ // \mathbf{K}^*$, la variété B_{a_i} étant une \mathbf{K}^* -variété quasi-projective pour laquelle le quotient catégorique $B_{a_i} // \mathbf{K}^*$ existe. La figure suivante illustre cette décomposition.

2.6. Structures localement toriques

Nous allons expliquer dans ce paragraphe, en montrant que l'on peut munir les variétés B_{a_i} construites précédemment d'un atlas de "cartes étales fortement toriques", que la décomposition obtenue précédemment est une décomposition en applications "localement toriques". Ceci est un point essentiel pour la suite où toutes les constructions élaborées

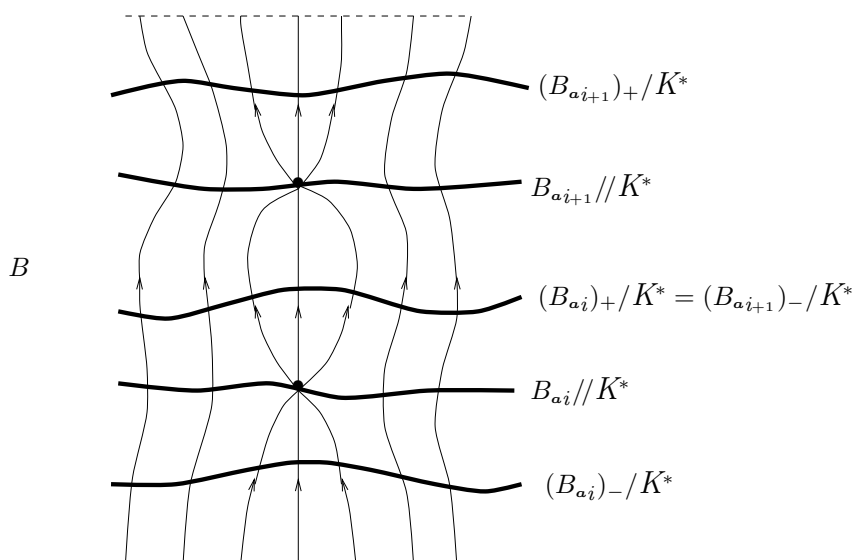


FIGURE 2. Décomposition d'un cobordisme birationnel

(en particulier dans la partie 4) et la vérification de leurs propriétés se feront à l'aide de cet atlas.

Rappelons qu'une application régulière $f : Z \rightarrow X$ entre variétés algébriques est *étale* si elle est lisse de dimension relative 0. De façon équivalente, f est étale si pour tout $z \in Z$, f induit un isomorphisme $f^* : \hat{\mathcal{O}}_{X,f(z)} \rightarrow \hat{\mathcal{O}}_{Z,z}$ entre les complétés des anneaux locaux. Sur le corps des nombres complexes, f est étale si et seulement si f est un biholomorphisme local.

Définition. Soit $f : Z \rightarrow X$ un morphisme étale, \mathbf{K}^* -équivariant entre deux variétés algébriques affines munies d'une action de \mathbf{K}^* . On dit que f est *fortement étale* si et seulement si $f//\mathbf{K}^* : Z//\mathbf{K}^* \rightarrow X//\mathbf{K}^*$ est étale et l'application naturelle $Z \rightarrow Z//\mathbf{K}^* \times_{X//\mathbf{K}^*} X$ est un isomorphisme.

Soit V une variété algébrique lisse munie d'une action algébrique de \mathbf{K}^* et soit $x \in V$ un \mathbf{K}^* -point fixe. Par un résultat de Sumihiro [Sum74] [Sum75], il existe un voisinage \mathbf{K}^* -invariant et affine W_x de x dans V . Comme x est un point fixe, le lemme fondamental de Luna [Lun73] [MFK94] assure qu'il existe un voisinage \mathbf{K}^* -invariant et affine $V_x \subset W_x$ de x , saturé pour la projection $\pi_x : W_x \rightarrow W_x//\mathbf{K}^*$, et un morphisme \mathbf{K}^* -équivariant et fortement étale $\eta_x : V_x \rightarrow X_x := T_x V$ (où $T_x V$ est l'espace tangent de V en x). Comme l'action de \mathbf{K}^* sur $T_x V$ se diagonalise, il existe une base de vecteurs propres de $T_x V$, faisant de $T_x V$ une variété torique affine lisse où \mathbf{K}^* agit comme sous-groupe à un paramètre du tore (nous renvoyons au paragraphe suivant pour les notions de géométrie torique).

En adaptant un peu cette construction, on obtient la proposition suivante, où l'on a repris les notations des paragraphes précédents :

PROPOSITION 2.5. — Soit B_{a_i} comme dans le paragraphe 2.5 et soit $\pi_i : B_{a_i} \rightarrow B_{a_i} // \mathbf{K}^*$ le quotient catégorique. Alors, pour tout $x \in B_{a_i}$, il y a un voisinage \mathbf{K}^* -invariant et affine V_x de x , saturé pour π_i , une variété torique affine lisse X_x et un morphisme \mathbf{K}^* -équivariant et fortement étale $\eta_x : V_x \rightarrow X_x$.

Mentionnons ici que l'on peut choisir les V_x saturés pour π_i car π_i est un morphisme affine d'après la démonstration de la proposition 2.4. Mentionnons aussi que la version en caractéristique non nulle du résultat de Luna donnée par Bardsley et Richardson [BaR85] peut s'appliquer ici si on ne suppose plus \mathbf{K} de caractéristique nulle.

Définition. Dans la proposition 2.5, pour $x \in B_{a_i}$, on appelle *carte torique fortement étale en x* la donnée de $\eta_x : V_x \rightarrow X_x$.

Ces cartes toriques fortement étales ont des propriétés très fortes, en particulier :

- $(B_{a_i})_- \cap V_x = (V_x)_-$ et $(B_{a_i})_+ \cap V_x = (V_x)_+$,
- le morphisme η_x se restreint en des morphismes $(\eta_x)_- := (\eta_x)|_{(V_x)_-} : (V_x)_- \rightarrow (X_x)_-$ et $(\eta_x)_+ := (\eta_x)|_{(V_x)_+} : (V_x)_+ \rightarrow (X_x)_+$ fortement étales et il y a un diagramme commutatif :

$$\begin{array}{ccccc}
 (V_x)_- / \mathbf{K}^* & & & & (V_x)_+ / \mathbf{K}^* \\
 (\eta_x)_- / \mathbf{K}^* \downarrow \scriptstyle i.i. & \rightsquigarrow & & \rightsquigarrow & (\eta_x)_+ / \mathbf{K}^* \\
 (X_x)_- / \mathbf{K}^* & & V_x // \mathbf{K}^* & & (X_x)_+ / \mathbf{K}^* \\
 & \rightsquigarrow & \downarrow \scriptstyle i.i. & \rightsquigarrow & \\
 & & X_x // \mathbf{K}^* & &
 \end{array}$$

La proposition 2.5 signifie donc que le cobordisme birationnel $\varphi_i : (B_{a_i})_- / \mathbf{K}^* \rightsquigarrow (B_{a_i})_+ / \mathbf{K}^*$ est *localement torique* au sens où, au voisinage de tout point, et à morphisme étale près, l'application birationnelle φ_i est torique. Il est essentiel de remarquer que nous sommes encore loin d'avoir démontré le théorème de factorisation : d'une part les $(B_{a_i})_- / \mathbf{K}^*$ et $(B_{a_i})_+ / \mathbf{K}^*$ sont des variétés singulières en général, et d'autre part la structure torique dépend du point x choisi dans B_{a_i} .

Le but de la partie 4 est de décomposer un cobordisme birationnel donné en une suite de cobordismes birationnels (en général singuliers) "toroïdaux", ce qui signifie en un certain sens que la "structure torique" ne dépendra plus du point x choisi.

En attendant, dans la partie 3 suivante, nous nous intéressons au cadre torique.

3. LE THÉORÈME DE MORELLI ET WŁODARCZYK

Nous montrons ici le théorème de factorisation faible des applications birationnelles toriques.

3.1. Rappels de géométrie torique et énoncé du théorème

Des références usuelles de géométrie torique sont [Ewa96], [Ful93] et [Oda88].

Définition. Une *variété torique* (de dimension n) est une variété algébrique normale, contenant le tore $T := (\mathbf{K}^*)^n$ comme ouvert (de Zariski), munie d'une action algébrique de T prolongeant l'action de T sur lui-même.

Si M est un réseau (*i.e.* un groupe abélien libre) de rang n , une variété torique X_Σ de dimension n munie d'une action du tore $T = \text{Hom}(M, \mathbf{K}^*)$ est définie par la donnée d'un éventail Σ , subdivision de l'espace vectoriel dual $N_{\mathbf{Q}} := \text{Hom}(M, \mathbf{Z}) \otimes_{\mathbf{Z}} \mathbf{Q}$ par des cônes rationnels polyédraux.

Si X_Σ est une variété torique d'éventail Σ , à chaque cône σ de Σ correspond naturellement un ouvert affine T -invariant U_σ de X_Σ de sorte que $X_\Sigma = \bigcup_{\sigma \in \Sigma} U_\sigma$. De plus, si $\sigma \subset \tau$ (*i.e.* σ est une face de τ), alors $U_\sigma \subset U_\tau$. Rappelons aussi qu'il y a une correspondance bijective entre les orbites de T de codimension r dans X_Σ et les cônes de dimension r de Σ . Pour $\sigma \in \Sigma$, on note $V(\sigma)$ l'adhérence de l'orbite correspondant à σ ; si $\sigma \subset \tau$, alors $V(\tau) \subset V(\sigma)$. Rappelons enfin qu'une variété torique est lisse si et seulement si tous les cônes de Σ sont *non-singuliers* (*i.e.* engendrés par une famille de $N := \text{Hom}(M, \mathbf{Z})$ pouvant se compléter en une base de N) et complète si et seulement si le support de Σ est $N_{\mathbf{Q}}$.

Le théorème de factorisation faible dans le cadre torique est dû à Morelli [Mor96] et Włodarczyk [Wlo97], voir aussi [AMR99] et [Mat00] :

THÉORÈME 3.1. — *Soit $\varphi : X_1 \dashrightarrow X_2$ une application birationnelle équivariante entre deux variétés toriques complètes et lisses X_1 et X_2 sur \mathbf{K} . Alors, φ se factorise en une suite d'éclatements et de contractions de centres lisses invariants. Autrement dit, il y a une suite d'applications birationnelles équivariantes entre variétés toriques complètes et lisses*

$$X_1 = V_0 \xrightarrow{\varphi_0} K V_1 \xrightarrow{\varphi_1} K \cdots \xrightarrow{\varphi_{i-1}} K V_i \xrightarrow{\varphi_i} K V_{i+1} \xrightarrow{\varphi_{i+1}} K \cdots \xrightarrow{\varphi_{l-2}} K V_{l-1} \xrightarrow{\varphi_{l-1}} K V_l = X_2$$

de sorte que $\varphi = \varphi_{l-1} \circ \varphi_{l-2} \circ \cdots \circ \varphi_1 \circ \varphi_0$ et pour tout i , $\varphi_i : V_i \dashrightarrow K V_{i+1}$ ou $\varphi_i^{-1} : V_{i+1} \dashrightarrow K V_i$ est une application régulière obtenue en éclatant une adhérence d'orbite du tore.

Dans la suite, nous donnons les grandes lignes de la preuve du théorème 3.1. Mentionnons ici que malgré les travaux [Mor96] et [AMR99], le problème de factorisation forte torique est toujours ouvert, même en dimension trois (voir [Mat00] pour une discussion des lacunes de [Mor96] et [AMR99]).

3.2. Cobordisme torique en termes d'éventails, d'après Morelli

Soient M un réseau de rang n et N le réseau dual $\text{Hom}(M, \mathbf{Z})$. Soient N^+ le réseau $(n+1)$ -dimensionnel $N \oplus \mathbf{Z}$, $N_{\mathbf{Q}} := N \otimes_{\mathbf{Z}} \mathbf{Q}$, $N_{\mathbf{Q}}^+ := N^+ \otimes_{\mathbf{Z}} \mathbf{Q}$ et π la projection $N_{\mathbf{Q}}^+ \rightarrow N_{\mathbf{Q}}$.

Dans toute la suite et sauf mention explicite du contraire, tous les cônes rationnels σ de $N_{\mathbf{Q}}^+$ que nous considèrerons seront supposés *simpliciaux* (*i.e.* engendrés par une famille libre d'éléments de N^+) et π -*strictement convexes* (*i.e.* $\pi(\sigma)$ est un cône strictement convexe de $N_{\mathbf{Q}}$).

Définition. Soit σ un cône rationnel de $N_{\mathbf{Q}}^+$. On dit que σ est π -*indépendant* si $\pi|_{\sigma}$ est injective (ceci signifie que $\text{Vect}(\sigma)$ ne contient pas la direction verticale $\{0\} \oplus \mathbf{Q}$). On dit que σ est π -*dépendant* s'il n'est pas π -indépendant. Soit σ un cône rationnel π -indépendant de $N_{\mathbf{Q}}^+$. On dit que σ est π -*non-singulier* si $\pi(\sigma)$ est un cône non-singulier de $N_{\mathbf{Q}}$. On dit que σ est π -*singulier* s'il n'est pas π -non-singulier. On dit qu'un éventail de $N_{\mathbf{Q}}^+$ est π -*non-singulier* si tous ses cônes π -indépendants sont π -non-singuliers.

Nous noterons $\nu = (0, 1) \in N \oplus \mathbf{Z} = N^+$, ce vecteur correspond à un sous-groupe à un paramètre λ_{ν} du tore $N^+ \otimes_{\mathbf{Z}} \mathbf{K}^* = T \times \mathbf{K}^*$.

Soit Σ un éventail simplicial de $N_{\mathbf{Q}}^+$. Définissons les *faces supérieure et inférieure* de Σ

$$\partial_+(\Sigma) = \{x \in \Sigma \mid x + \varepsilon\nu \notin \Sigma \text{ pour } \varepsilon > 0 \text{ petit}\}$$

et

$$\partial_-(\Sigma) = \{x \in \Sigma \mid x - \varepsilon\nu \notin \Sigma \text{ pour } \varepsilon > 0 \text{ petit}\}.$$

Définition. Soit σ un cône rationnel simplicial de $N_{\mathbf{Q}}^+$. Nous dirons que σ est un *circuit* si σ est π -dépendant et si toutes les faces strictes de σ sont π -indépendantes.

Remarquons que tout cône rationnel simplicial π -dépendant de $N_{\mathbf{Q}}^+$ contient (*i.e.* comme face) un unique circuit.

Soit τ un cône rationnel simplicial π -dépendant de $N_{\mathbf{Q}}^+$ et soit σ l'unique circuit inclus dans τ . *Dans ce qui suit, nous allons montrer que la variété affine X_{τ} munie de l'action du sous-groupe à un paramètre λ_{ν} est un cobordisme birationnel, que $V(\sigma)$ est l'unique composante connexe de λ_{ν} -points fixes de X_{τ} et relier $\partial_-(\tau)$ et $\partial_+(\tau)$ à $(X_{\tau})_-$ et $(X_{\tau})_+$.*

Supposons pour simplifier que τ est non-singulier et écrivons $\tau = \langle \rho_1, \dots, \rho_k, \dots, \rho_m \rangle$ où les ρ_i sont les générateurs dans N^+ des faces de dimension 1 de τ numérotés de sorte que $\sigma = \langle \rho_1, \dots, \rho_k \rangle$. Comme τ est non-singulier, on peut compléter la famille $(\rho_1, \dots, \rho_k, \dots, \rho_m)$ en une base $(\rho_1, \dots, \rho_m, \dots, \rho_{n+1})$ de N^+ . La variété torique X_{τ} est isomorphe à $\mathbf{K}^m \times (\mathbf{K}^*)^{n+1-m}$ et contient l'ouvert affine invariant $U_{\sigma} = \mathbf{K}^k \times (\mathbf{K}^*)^{n+1-k}$. Soit (u_1, \dots, u_{n+1}) la base duale de $(\rho_1, \dots, \rho_{n+1})$. Comme $\nu \in \text{Vect}(\sigma)$, on a $\langle u_i, \nu \rangle = 0$ pour $i \geq k + 1$ et quitte à renuméroter les ρ_j , on peut supposer que $\langle u_i, \nu \rangle > 0$ pour $1 \leq i \leq l$ et $\langle u_i, \nu \rangle < 0$ pour $l + 1 \leq i \leq k$ où l est un certain entier vérifiant $1 \leq l \leq k - 1$ (un tel l existe car σ est π -strictement convexe).

Le groupe \mathbf{K}^* agit comme sous-groupe à un paramètre λ_{ν} sur $x \in X_{\tau}$ par

$$t \cdot x = \lambda_{\nu}(t)(x_1, \dots, x_{n+1}) = (t^{\langle u_1, \nu \rangle} x_1, \dots, t^{\langle u_k, \nu \rangle} x_k, x_{k+1}, \dots, x_{n+1}).$$

Pour j vérifiant $1 \leq j \leq k$, notons $\gamma_j = \langle \rho_1, \dots, \check{\rho}_j, \dots, \rho_m \rangle$ où la notation $\check{\rho}_j$ signifie que ρ_j ne figure pas parmi les arêtes de γ_j . C'est une face maximale de τ et les formules précédentes montrent que :

$$(X_\tau)^{\mathbf{K}^*} = \{(0, \dots, 0, x_{k+1}, \dots, x_{n+1})\} = V(\sigma); (X_\tau)_+ = \bigcup_{1 \leq j \leq l} U_{\gamma_j}; (X_\tau)_- = \bigcup_{l+1 \leq j \leq k} U_{\gamma_j}.$$

Par ailleurs, c'est un exercice facile de montrer que

$$\partial_+(\tau) = \bigcup_{1 \leq j \leq l} \gamma_j \text{ et } \partial_-(\tau) = \bigcup_{l+1 \leq j \leq k} \gamma_j.$$

Ainsi, la variété torique $X_{\pi(\partial_+(\tau))}$ (resp. $X_{\pi(\partial_-(\tau))}$) d'éventail $\pi(\partial_+(\tau))$ (resp. $\pi(\partial_-(\tau))$) est la variété torique $(X_\tau)_+/\mathbf{K}^*$ (resp. $(X_\tau)_-/\mathbf{K}^*$) et X_τ est un cobordisme torique birationnel pour $\varphi : X_{\pi(\partial_-(\tau))} \dashrightarrow K X_{\pi(\partial_+(\tau))}$. Tout ceci est illustré par la figure 3.

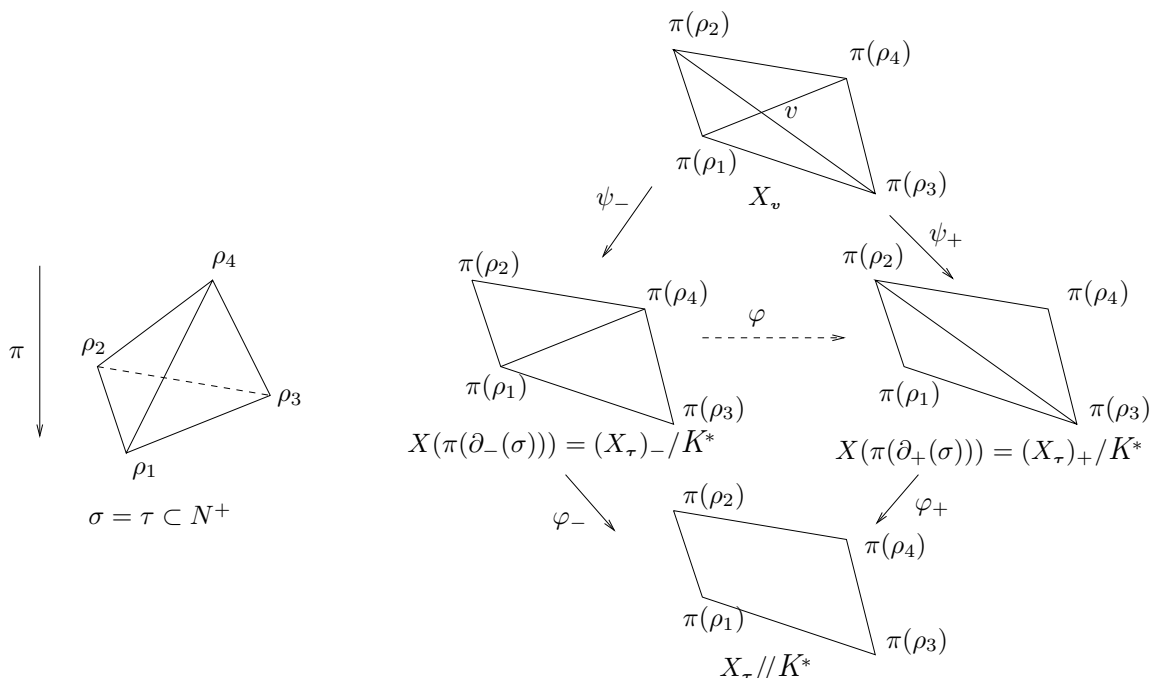


FIGURE 3. Cobordisme torique

Étudions maintenant l'application birationnelle $\varphi : X_{\pi(\partial_-(\tau))} \dashrightarrow K X_{\pi(\partial_+(\tau))}$. Pour chaque ρ_j , soit v_j le générateur dans N de l'arête engendrée par $\pi(\rho_j)$ et w_j le rationnel tel que ρ_j soit proportionnel à (v_j, w_j) : on écrit $\rho_j = c_j(v_j, w_j)$ où c_j est un entier strictement positif. Projetant l'égalité

$$\nu = (0, 1) = \sum_{j=1}^k \langle u_j, \nu \rangle \rho_j$$

sur le facteur $N_{\mathbf{Q}}$ de $N_{\mathbf{Q}}^+$, il vient :

$$\sum_{j=1}^k \langle u_j, \nu \rangle c_j v_j = 0.$$

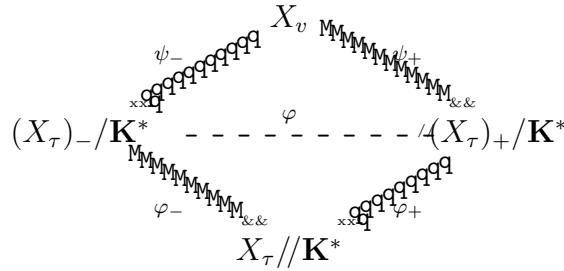
Posons

$$v = \sum_{j=1}^l \langle u_j, \nu \rangle c_j v_j = - \sum_{j=l+1}^k \langle u_j, \nu \rangle c_j v_j \in N.$$

Ce vecteur v est dans l'intérieur relatif des cônes

$$\langle v_1, \dots, v_l \rangle = \bigcap_{j=l+1}^k \pi(\gamma_j) \text{ et } \langle v_{l+1}, \dots, v_k \rangle = \bigcap_{j=1}^l \pi(\gamma_j).$$

Par conséquent, les éventails obtenus par subdivision étoilée de $\pi(\partial_+(\tau))$ et de $\pi(\partial_-(\tau))$ par rapport à v sont égaux, si bien qu'il existe une variété torique X_v et un diagramme commutatif :



Si de plus le cône initial τ est π -non singulier, alors $(X_\tau)_-/\mathbf{K}^*$ et $(X_\tau)_+/\mathbf{K}^*$ sont lisses et

$$v = \sum_{j=1}^l v_j = \sum_{j=l+1}^k v_j \in N,$$

de sorte que X_v est lisse, $\psi_+ : X_v \rightarrow (X_\tau)_+/\mathbf{K}^*$ et $\psi_- : X_v \rightarrow (X_\tau)_-/\mathbf{K}^*$ sont des éclatements le long de sous-variétés lisses invariantes (voir la figure 3).

Insistons ici sur le fait que seule la π -non singularité du cône initial τ est importante (pour obtenir des variétés $(X_\tau)_-/\mathbf{K}^*$ et $(X_\tau)_+/\mathbf{K}^*$ lisses et des éclatements ψ_+ et ψ_- le long de sous-variétés lisses invariantes) mais il n'est pas nécessaire de supposer que τ soit non-singulier, le fait qu'il soit simplicial est suffisant.

3.3. Cobordisme torique

Soit $\varphi : X_{\Delta'} \rightarrow X_\Delta$ une application régulière birationnelle équivariante entre deux variétés toriques projectives lisses de dimension n et d'éventails respectifs Δ' et Δ dans $N_{\mathbf{Q}}$. La construction expliquée au paragraphe 2.3 s'adapte sans difficulté au cas torique (historiquement, rappelons que c'est la construction torique qui a inspiré le cas général), si bien qu'il existe un cobordisme birationnel B torique (au sens où l'action de \mathbf{K}^* sur B est celle d'un sous-groupe à un paramètre du tore de B) quasi-projectif pour φ .

D'après ce qui précède, l'existence d'un tel B en termes d'éventails se traduit de la façon suivante. Il existe un éventail Σ non-singulier (et donc simplicial) dans $N_{\mathbf{Q}}^+$ tel que $\pi(\partial_-(\Sigma)) = \Delta'$, $\pi(\partial_+(\Sigma)) = \Delta$, B est la variété torique d'éventail Σ et φ est le cobordisme birationnel $\varphi : B_-/\mathbf{K}^* = X_{\Delta'} \rightarrow B_+/\mathbf{K}^* = X_\Delta$. Comme B est quasi-projectif, il y a un ordre sur les circuits de Σ (qui correspondent aux composantes connexes des points

fixes pour l'action du sous-groupe à un paramètre $\nu = (0, 1)$). Choisissons un circuit σ_1 minimal pour cet ordre. Soient $\text{Star}(\sigma_1) = \{\tau \in \Sigma \mid \sigma_1 \subset \tau\}$ et $\overline{\text{Star}}(\sigma_1) = \{\tau' \in \Sigma \mid \tau' \subset \tau \text{ pour un } \tau \text{ dans } \text{Star}(\sigma_1)\}$. Alors comme σ_1 est minimal, $\partial_-(\overline{\text{Star}}(\sigma_1)) \subset \partial_-(\Sigma)$. Posons alors $\Sigma_1 = (\Sigma \setminus \partial_-(\overline{\text{Star}}(\sigma_1))) \cup \partial_+(\overline{\text{Star}}(\sigma_1))$. D'après le paragraphe précédent, si B_1 est la variété d'éventail Σ_1 et si X_1 est la variété torique d'éventail $\Delta_1 := \pi(\partial_-(\Sigma_1))$, alors φ se décompose en

$$\varphi : B_-/\mathbf{K}^* = X_{\Delta'} \xleftarrow{\psi_-} X_v \xrightarrow{\psi_+} X_{\Delta_1} = (B_1)_-/\mathbf{K}^* \xrightarrow{\varphi_1} KX_{\Delta} = (B_1)_+/\mathbf{K}^* = B_+/\mathbf{K}^*,$$

et on peut recommencer avec le cobordisme B_1 . Si de plus Σ est π -non singulier, alors X_{Δ_1} et X_v sont lisses, ψ_- et ψ_+ sont des éclatements le long de sous-variétés lisses invariantes et Σ_1 est encore π -non singulier.

Faisons le point : le théorème 3.1 est démontré pour une application régulière $\varphi : X_{\Delta'} \rightarrow X_{\Delta}$ birationnelle équivariante entre deux variétés toriques projectives lisses de dimension n si on peut contruire un cobordisme birationnel \tilde{B} quasi-projectif d'éventail $\tilde{\Sigma}$ simplicial et π -non singulier dans $N_{\mathbf{Q}}^+$ tel que $\pi(\partial_-(\tilde{\Sigma})) = \Delta'$ et $\pi(\partial_+(\tilde{\Sigma})) = \Delta$.

Un tel éventail $\tilde{\Sigma}$ va être obtenu en π -désingularisant l'éventail Σ , son existence est garantie par le théorème de π -désingularisation de Morelli, dont l'énoncé fait l'objet du paragraphe suivant et dont la démonstration est repoussée à la fin de ce texte §6.

3.4. Le théorème de π -désingularisation de Morelli

Les notations sont celles du paragraphe précédent. Le théorème suivant est dû à Morelli [Mor96] avec une démonstration incomplète ; la démonstration a été complétée par Abramovich, Matsuki et Rashid [AMR99] [Mat00].

THÉORÈME 3.2. — *Soit Σ un éventail de $N_{\mathbf{Q}}^+$. Alors il existe un éventail π -non-singulier de $N_{\mathbf{Q}}^+$ obtenu par une suite finie de subdivisions étoilées de Σ . De plus, la procédure de π -désingularisation n'affecte pas les cônes π -non-singuliers de Σ .*

Il apparaîtra clairement lors de la démonstration de ce résultat §6 que l'éventail obtenu n'est en général pas non-singulier.

3.5. Démonstration du théorème de factorisation faible torique

On termine la démonstration du théorème de factorisation, qui repose sur le lemme de Chow torique (voir par exemple [Oda88]) et la levée des indéterminations torique due à De Concini-Procesi [DCP85].

THÉORÈME 3.3. — *Soit X une variété torique complète. Alors il existe une variété torique projective lisse \tilde{X} et une application régulière birationnelle équivariante $\varphi : \tilde{X} \rightarrow X$.*

THÉORÈME 3.4. — Soit $\varphi : X \dashrightarrow X'$ une application birationnelle équivariante entre deux variétés toriques complètes lisses. Alors il existe une suite d'éclatements le long de sous-variétés invariantes de codimension deux $\psi : \bar{X} \rightarrow X$ telle que $\varphi \circ \psi : \bar{X} \rightarrow X'$ soit une application régulière birationnelle équivariante.

De ces deux résultats, déduisons une version précise du théorème de Moishezon torique :

THÉORÈME 3.5. — Soit X une variété torique lisse et complète. Alors il existe une variété torique projective lisse \tilde{X} obtenue à partir de X par suite d'éclatements le long de sous-variétés invariantes de codimension deux $\psi : \tilde{X} \rightarrow X$.

Démonstration. Soit $\varphi : \tilde{X} \rightarrow X$ donnée par le théorème 3.3, où \tilde{X} est projective lisse. Appliquons le théorème 3.4 à φ^{-1} : il y a une suite d'éclatements le long de sous-variétés invariantes de codimension deux $\psi : \bar{X} \rightarrow X$ de sorte que $h = \varphi^{-1} \circ \psi : \bar{X} \rightarrow \tilde{X}$ soit une application régulière birationnelle. Comme ψ est un morphisme projectif et puisque $\varphi \circ h = \psi$, h est aussi un morphisme projectif, et comme \tilde{X} est projective, on en déduit que \bar{X} est projective. ■

Application : soit $\varphi : X_1 \dashrightarrow X_2$ une application birationnelle entre deux variétés toriques complètes et lisses. Par le théorème 3.5, il y a des variétés toriques projectives lisses \tilde{X}_1 et \tilde{X}_2 et deux suites d'éclatements le long de sous-variétés invariantes de codimension deux $\psi_1 : \tilde{X}_1 \rightarrow X_1$ et $\psi_2 : \tilde{X}_2 \rightarrow X_2$. Soit $\tilde{\varphi} = \psi_2^{-1} \circ \varphi \circ \psi_1 : \tilde{X}_1 \dashrightarrow \tilde{X}_2$. Par le théorème 3.4, il existe une suite d'éclatements le long de sous-variétés invariantes de codimension deux $\psi : \bar{X}_1 \rightarrow \tilde{X}_1$ telle que $\tilde{\varphi} \circ \psi : \bar{X}_1 \rightarrow \tilde{X}_2$ soit une application régulière birationnelle équivariante. Comme \tilde{X}_1 et \tilde{X}_2 sont projectives lisses, il y a d'après les paragraphes précédents une factorisation faible pour $\tilde{\varphi} \circ \psi$ et par suite une factorisation faible pour φ . Ceci achève la démonstration du théorème 3.1.

3.6. Extension au cas toroïdal

Il y a une classe de variétés algébriques possédant essentiellement les mêmes propriétés que les variétés toriques.

Définition. Soit V une variété algébrique et U un ouvert de V . Le plongement $U \subset V$ est un *plongement toroïdal* si pour tout $x \in V$, il y a un voisinage V_x de x dans V , une variété torique X_x et une application étale $\eta_x : V_x \rightarrow X_x$ telle que $\eta_x^{-1}(T_x) = V_x \cap U$.

Comme dans le cas torique, il y a une notion naturelle de morphismes toroïdaux et d'applications birationnelles toroïdales entre plongements toroïdaux. De plus, il y a une stratification naturelle de V et les adhérences des strates correspondent formellement aux adhérences d'orbites dans les variétés toriques. Enfin, un plongement toroïdal est décrit par un complexe polyédral conique rationnel qui joue exactement le rôle de l'éventail d'une variété torique (à ceci près que ce complexe est défini abstraitement et qu'il ne se plonge en général pas linéairement dans un espace vectoriel $N_{\mathbf{Q}}$). Les résultats de

la géométrie birationnelle des variétés toriques (existence de désingularisation, levée des indéterminations par éclatements de centres lisses, etc.) s'étendent sans trop de difficultés au cas des plongements toroïdaux (voir [KKMS73] ou [AbO97] p. 69-73). Il en est de même pour le théorème de factorisation comme l'ont remarqué Abramovich, Matsuki et Rashid [AMR99].

THÉORÈME 3.6. — *Soit $\varphi : (U_1 \subset X_1) \xrightarrow{g} K(U_2 \subset X_2)$ une application birationnelle toroïdale entre deux plongements toroïdaux lisses sur \mathbf{K} , X_1 et X_2 complètes. Alors, φ se factorise en une suite d'éclatements et de contractions de centres lisses égaux à des adhérences de strates.*

4. STRUCTURES TOROÏDALES ET COBORDISME BIRATIONNEL

Le but de cette partie est de décomposer un cobordisme birationnel donné en une suite de cobordismes birationnels (en général singuliers) toroïdaux. L'idée est la suivante : si on éclate un faisceau d'idéaux I sur une variété algébrique, on obtient une nouvelle variété algébrique avec un diviseur bien déterminé puisque I devient principal après éclatement. Le complémentaire de son support est alors un candidat pour être un plongement toroïdal. Lorsque de plus la variété algébrique est munie d'une action de \mathbf{K}^* , on peut essayer de construire I de sorte que le plongement toroïdal obtenu soit compatible avec l'action de \mathbf{K}^* .

4.1. Idéal α -toroïdal

Soit V une variété algébrique munie d'une action algébrique effective de \mathbf{K}^* . Pour $v \in V$, on note $\text{Stab}(v)$ le stabilisateur de v ($\text{Stab}(v)$ est isomorphe soit à \mathbf{K}^* , soit à un groupe cyclique $\mathbf{Z}/n\mathbf{Z}$).

Notation. Si α est un entier, on note $J_{\alpha,v}$ l'idéal de $\mathcal{O}_{V,v}$ engendré par les fonctions $f \in \mathcal{O}_{V,v}$ $\text{Stab}(v)$ -semi-invariantes de $\text{Stab}(v)$ -poids α , *i.e.* pour tout $t \in \text{Stab}(v)$, $t^*(f) = t^\alpha f$.

Il n'est pas difficile de voir que si z_1, \dots, z_n sont des générateurs $\text{Stab}(v)$ -semi-invariants de l'idéal maximal $\mathcal{M}_{V,v}$ dont l'existence est assurée par le théorème de Sumihiro [Sum74] [Sum75], alors $J_{\alpha,v}$ est engendré par les monômes $z_1^{m_1} \cdots z_n^{m_n}$ de $\text{Stab}(v)$ -poids α (*i.e.* vérifiant $a_1 m_1 + \cdots + a_n m_n = \alpha$ si a_i désigne le $\text{Stab}(v)$ -poids de z_i). Ceci implique en particulier que si $\eta : V \rightarrow X$ est un morphisme \mathbf{K}^* -équivariant fortement étale entre deux variétés algébriques munies d'une action de \mathbf{K}^* , alors l'image inverse de $J_{\alpha,\eta(v)}$ engendre $J_{\alpha,v}$ pour tous $v \in V$ et $\alpha \in \mathbf{Z}$.

La collection des $J_{\alpha,v}$ lorsque v décrit V ne définit pas un faisceau d'idéaux cohérent en général. Cependant, on a la proposition suivante :

PROPOSITION 4.1. — Soient $\alpha \in \mathbf{Z}$ et B un cobordisme birationnel lisse quasi-projectif de la forme B_{α_i} (les notations sont celles des parties 2.5 et 2.6). Alors, il existe un unique faisceau d'idéaux cohérent \mathbf{K}^* -équivariant I_α non nul tel que pour tout $v \in (B_+ \cap B_-) \cup B^{\mathbf{K}^*}$, on a $(I_\alpha)_v = J_{\alpha, v}$.

L'idéal I_α s'appelle le faisceau α -toroïdal associé à l'action de \mathbf{K}^* sur B . Rappelons que $(B_+ \cap B_-)$ est un ouvert non vide de B et que $(B_+ \cap B_-) \cup B^{\mathbf{K}^*}$ est l'union des \mathbf{K}^* -orbites fermées de B .

Démonstration de la proposition 4.1. Commençons par l'unicité de I_α . Elle est claire sur $(B_+ \cap B_-) \cup B^{\mathbf{K}^*}$. Soit donc v dans le complémentaire de $(B_+ \cap B_-) \cup B^{\mathbf{K}^*}$. L'adhérence de l'orbite de v contient un point fixe. Comme I_α est uniquement déterminé en ce point fixe, il l'est aussi au voisinage puisque I_α est cohérent. Par \mathbf{K}^* -équivariance, I_α est alors uniquement déterminé en v .

Pour l'existence, la proposition 2.5 permet de recouvrir B par des cartes toriques $\eta_x : V_x \rightarrow X_x$ fortement étales. Comme $((B_+ \cap B_-) \cup B^{\mathbf{K}^*}) \cap V_x$ est l'image inverse par η_x de $((X_x)_+ \cap (X_x)_-) \cup (X_x)^{\mathbf{K}^*}$, il suffit de construire un faisceau α -toroïdal sur X_x . Par image inverse, ceci définit un faisceau α -toroïdal sur V_x et l'unicité déjà prouvée entraîne que ces faisceaux ainsi construits se recollent en un faisceau sur B . Le lemme suivant [AKMW99] achève donc la preuve. ■

LEMME 4.2. — Soit X une variété torique affine lisse définie par un cône régulier σ d'un réseau N et soit $a \in N$ un élément primitif correspondant à une action effective de \mathbf{K}^* d'un sous-groupe à un paramètre de T . On suppose de plus que a n'appartient pas à $\sigma \cup -\sigma$ (ceci correspond au fait que $B_+ \cap B_-$ est non vide). Alors, pour tout $\alpha \in \mathbf{Z}$, le faisceau I_α existe, est non nul et engendré par les monômes z^m où $m \in \check{\sigma}$ avec $\langle m, a \rangle = \alpha$.

Remarquons qu'il n'est pas utile d'après ce lemme de calculer explicitement les stabilisateurs $\text{Stab}(v)$. Donnons un exemple concret, que nous utiliserons aussi dans la suite :

Exemple “(2, 1, -1)”. Soit $X = \mathbf{K}^3$ munie de l'action du sous-groupe à un paramètre $a = (2, 1, -1)$ (i.e. $\lambda_a(t)(z_1, z_2, z_3) = (t^2 z_1, t z_2, t^{-1} z_3)$), alors I_2 (resp. I_1 , resp. I_{-1}) est le faisceau d'idéaux engendré par z_1 et z_2^2 (resp. $z_1 z_3$ et z_2 , resp. z_3).

4.2. Plongements toroïdaux avec action toroïdale de \mathbf{K}^*

Soit V une variété algébrique munie d'une action algébrique de \mathbf{K}^* . On suppose que l'action de \mathbf{K}^* est localement torique au sens où on peut recouvrir V par des cartes toriques fortement étales. Soit D un diviseur effectif de V et U l'ouvert $V \setminus \text{Supp}(D)$.

Définition. On dit que U est un *plongement toroïdal avec action toroïdale de \mathbf{K}^** si pour tout $x \in V$, il y a un voisinage \mathbf{K}^* -invariant et affine V_x de x , une variété torique affine X_x avec une action de \mathbf{K}^* d'un sous-groupe à un paramètre et un morphisme \mathbf{K}^* -équivariant étale $\eta_x : V_x \rightarrow X_x$ tel que $\eta_x^{-1}(T_x) = V_x \cap U$ (où T_x désigne le tore dense de X_x). On

appelle *carte toroïdale étale* en x la donnée d'une telle $\eta_x : V_x \rightarrow X_x$. D'après le lemme fondamental de Luna, si une telle carte existe, en restreignant l'ouvert V_x , on peut la supposer de plus fortement étale.

Exemples. Soit X la variété torique \mathbf{K}^2 définie par un cône régulier $\sigma = \langle v_1, v_2 \rangle$ d'un réseau N et soit $a = (a_1, a_2) \in N$ un élément primitif correspondant à une action effective de \mathbf{K}^* d'un sous-groupe à un paramètre de T . Soit $D_{v_1} = \{(z_1, z_2) \in \mathbf{K}^2 \mid z_1 = 0\}$ et $D_{v_2} = \{(z_1, z_2) \in \mathbf{K}^2 \mid z_2 = 0\}$. Alors, $U = \mathbf{K}^2 \setminus (D_{v_1} \cup D_{v_2})$ est un plongement toroïdal avec action toroïdale de \mathbf{K}^* quel que soit le choix du sous-groupe à un paramètre a . En revanche, $U = \mathbf{K}^2 \setminus D_{v_1}$ est un plongement toroïdal avec action toroïdale de \mathbf{K}^* si et seulement si $a = \pm v_1$. Géométriquement, ceci signifie que l'action de \mathbf{K}^* sur \mathbf{K}^2 est produit de l'action triviale sur $\{0\} \times \mathbf{K}$ et d'une action d'un sous-groupe à un paramètre de la variété torique $\mathbf{K} \times \{0\}$ et que le diviseur D_{v_2} est égal à $\mathbf{K} \times \{0\}$.

Cet exemple se généralise en le lemme suivant [AKMW99] :

LEMME 4.3. — Soit X une variété torique définie par un éventail Σ d'un réseau N et soit $a \in N$ un élément primitif correspondant à une action effective de \mathbf{K}^* d'un sous-groupe à un paramètre de T , soient D un diviseur torique effectif de X et $U = X \setminus \text{Supp}(D)$. Les assertions suivantes sont équivalentes :

- (i) pour tout cône σ de Σ et tout diviseur torique E de l'ouvert torique affine $U_\sigma \subset X$, si E n'est pas dans $\text{Supp}(D)$, alors il existe une variété torique affine $X_{\sigma'}$ telle que :

$$U_\sigma \simeq X_{\sigma'} \times \mathbf{K} \text{ et } E \simeq X_{\sigma'} \times \{0\}$$

de sorte que l'action de \mathbf{K}^* sur U_σ soit le produit de l'action d'un sous-groupe à un paramètre sur $X_{\sigma'}$ et de l'action triviale sur \mathbf{K} ,

- (ii) U est un plongement toroïdal avec action toroïdale de \mathbf{K}^* .

Remarquons que l'implication (i) implique (ii) est aisée, c'est la seule utilisée dans la suite.

4.3. Faisceaux toroïdaux et plongements toroïdaux

Soit B un cobordisme birationnel lisse et quasi-projectif de dimension $n + 1$ de la forme B_{a_i} (les notations sont celles des parties 2.5 et 2.6). Fixons nous une famille finie de cartes toriques fortement étales $\eta_x : V_x \rightarrow X_x$ recouvrant B . Dans X_x , l'action de \mathbf{K}^* correspond à celle d'un sous-groupe à un paramètre : concrètement X_x est isomorphe à $\mathbf{K}^m \times (\mathbf{K}^*)^{n+1-m}$ et l'action de \mathbf{K}^* est de la forme $t \cdot (x_1, \dots, x_{n+1}) = (t^{\alpha_1} x_1, \dots, t^{\alpha_{n+1}} x_{n+1})$ pour certains poids entiers $\alpha_1, \dots, \alpha_{n+1}$.

Notation. On note \mathcal{A} une famille finie d'entiers telle que pour toute carte torique fortement étale $\eta_x : V_x \rightarrow X_x$, tous les poids $\alpha_1, \dots, \alpha_{n+1}$ de l'action de \mathbf{K}^* appartiennent à \mathcal{A} . Une telle famille d'entiers sera dite *admissible*.

Définition. Si \mathcal{A} est une famille admissible d'entiers, on note $I_{\mathcal{A}}$ le faisceau d'idéaux sur B défini par $I_{\mathcal{A}} = \prod_{\alpha \in \mathcal{A}} I_{\alpha}$. Un tel faisceau d'idéaux sur B est appelé *faisceau toroïdal*.

Notation. Si $I_{\mathcal{A}}$ est un faisceau toroïdal, on note $\pi_{\mathcal{A}} : B_{I_{\mathcal{A}}}^{tor} \rightarrow B$ la normalisation de l'éclatement de B de centre le faisceau d'idéaux $I_{\mathcal{A}}$. Comme $I_{\mathcal{A}}$ est \mathbf{K}^* -équivariant, la variété $B_{I_{\mathcal{A}}}^{tor}$ est naturellement munie d'une action algébrique de \mathbf{K}^* faisant de $\pi_{\mathcal{A}}$ un morphisme équivariant. On note $D_{I_{\mathcal{A}}}^{tor}$ le diviseur défini par le faisceau inversible $\pi_{\mathcal{A}}^{-1}(I_{\mathcal{A}})\mathcal{O}_{B_{I_{\mathcal{A}}}^{tor}}$ et $U_{I_{\mathcal{A}}}^{tor}$ le complémentaire de son support dans $B_{I_{\mathcal{A}}}^{tor}$.

La proposition suivante est le cœur de cette partie :

PROPOSITION 4.4. — *La variété algébrique $B_{I_{\mathcal{A}}}^{tor}$ est un cobordisme birationnel quasi-projectif. L'ouvert $U_{I_{\mathcal{A}}}^{tor}$ est un plongement toroïdal avec action toroïdale de \mathbf{K}^* . Enfin, $(B_{I_{\mathcal{A}}}^{tor})_+ = \pi_{\mathcal{A}}^{-1}(B_+)$ et $(B_{I_{\mathcal{A}}}^{tor})_- = \pi_{\mathcal{A}}^{-1}(B_-)$.*

Démonstration. Montrons que l'ouvert $U_{I_{\mathcal{A}}}^{tor}$ est un plongement toroïdal avec action toroïdale de \mathbf{K}^* . A nouveau, il suffit de traiter le cas torique, autrement dit de montrer ce résultat dans chacune des cartes toriques fortement étales considérées au début de la construction. Le lemme suivant donne donc le résultat. ■

LEMME 4.5. — Soit X la variété torique affine lisse $\mathbf{K}^m \times (\mathbf{K}^*)^{n+1-m}$ avec une action de \mathbf{K}^* de la forme $t \cdot (x_1, \dots, x_{n+1}) = (t^{\alpha_1} x_1, \dots, t^{\alpha_{n+1}} x_{n+1})$ pour certains poids entiers $\alpha_1, \dots, \alpha_{n+1}$. Soit \mathcal{A} une famille d'entiers contenant $\alpha_1, \dots, \alpha_{n+1}$ et $I_{\mathcal{A}}$ le faisceau toroïdal associé. Soient $\pi_{\mathcal{A}} : X_{I_{\mathcal{A}}}^{tor} \rightarrow X$ la normalisée de l'éclatement de centre $I_{\mathcal{A}}$, $D_{I_{\mathcal{A}}}^{tor}$ le diviseur défini par le faisceau inversible $\pi_{\mathcal{A}}^{-1}(I_{\mathcal{A}})\mathcal{O}_{X_{I_{\mathcal{A}}}^{tor}}$ et $U_{I_{\mathcal{A}}}^{tor}$ le complémentaire de son support dans $X_{I_{\mathcal{A}}}^{tor}$. Alors $U_{I_{\mathcal{A}}}^{tor}$ est un plongement toroïdal avec action toroïdale de \mathbf{K}^* .

Démonstration. Les diviseurs toriques de $X_{I_{\mathcal{A}}}^{tor}$ non inclus dans $\text{Supp}(D_{I_{\mathcal{A}}}^{tor})$ (les seuls à considérer d'après le lemme 4.3) sont des transformées strictes de diviseur $D_i := \{x \in X \mid x_i = 0\}$. Soit donc i dans $\{1, \dots, n+1\}$. Évidemment, le monôme x_i appartient au faisceau α_i -toroïdal I_{α_i} . Si ce dernier est principal engendré par x_i , alors D_i est dans $\text{Supp}(D_{I_{\mathcal{A}}}^{tor})$. Sinon, soit X_i la normalisée de l'éclatement du faisceau α_i -toroïdal I_{α_i} . Il est aisé de vérifier que D_i satisfait l'assertion (i) du lemme 4.3. La suite de la démonstration consiste à remarquer que $X_{I_{\mathcal{A}}}^{tor}$ est obtenu en normalisant l'éclatement de centre I_{α_1} , puis en normalisant l'éclatement de centre l'image inverse de I_{α_2} , etc. ■

Reprenons l'exemple “(2, 1, -1)” : $X = \mathbf{K}^3$ munie de l'action du sous-groupe à un paramètre (2, 1, -1). On prend ici $\mathcal{A} = \{-1, 1, 2\}$. Si X est définie par un cône régulier $\sigma = \langle v_1, v_2, v_3 \rangle$, alors $X_{I_{\mathcal{A}}}^{tor}$ est la variété torique d'éventail Σ obtenu en subdivisant σ en les trois cônes : $\sigma_1 = \langle v_1, 2v_1 + v_2, v_3 \rangle$, $\sigma_2 = \langle v_1 + v_2, v_3, 2v_1 + v_2, v_2 + v_3 \rangle$ et $\sigma_3 = \langle v_1 + v_2, v_2 + v_3, v_2 \rangle$. Le diviseur $D_{I_{\mathcal{A}}}^{tor}$ est le diviseur $D_{v_3} \cup D_{v_1+v_2} \cup D_{2v_1+v_2} \cup D_{v_2+v_3}$. Seul les cônes σ_1 et σ_3 sont à considérer pour vérifier l'assertion (i) du lemme 4.3, avec $E = D_{v_1}$ pour σ_1 et $E = D_{v_2}$ pour σ_3 . L'assertion (i) est satisfaite car $a = (2, 1, -1)$ appartient au réseau engendré par $v_1 + v_2$ et $v_2 + v_3$ et au réseau engendré par $2v_1 + v_2$ et v_3 .

COROLLAIRE 4.6. — Soit B_{a_i} un cobordisme birationnel lisse et quasi-projectif obtenu dans le paragraphe 2.5. Choisissons un faisceau toroïdal I_{A_i} et notons $\pi_{A_i} : B_{a_i}^{tor} \rightarrow B_{a_i}$ la normalisée de l'éclatement de centre I_{A_i} . Alors, il y a un diagramme commutatif :

$$\begin{array}{ccc}
 (B_{a_i}^{tor})_- / \mathbf{K}^* & \xrightarrow{\varphi_i^{tor}} & (B_{a_i}^{tor})_+ / \mathbf{K}^* \\
 \downarrow (\pi_{A_i})_- & & \downarrow (\pi_{A_i})_+ \\
 V_i = (B_{a_i})_- / \mathbf{K}^* & \xrightarrow{\varphi_i} & (B_{a_i})_+ / \mathbf{K}^* = V_{i+1}
 \end{array}$$

Faisons le point : la variété $B_{a_i}^{tor}$ contient l'ouvert $U_{I_{A_i}}^{tor}$ comme plongement toroïdal avec action de \mathbf{K}^* si bien que φ_i^{tor} est une application birationnelle toroïdale au sens où pour tout $x \in B_{a_i}^{tor}$ il y a une carte en x toroïdale fortement étale $\eta_x : V_x^{tor} \rightarrow X_x$ (donc vérifiant $\eta_x^{-1}(T_x) = V_x^{tor} \cap U_{I_{A_i}}^{tor}$) induisant un diagramme commutatif :

$$\begin{array}{ccc}
 (V_x^{tor})_- / \mathbf{K}^* & \xrightarrow{\varphi_i^{tor}} & (V_x^{tor})_+ / \mathbf{K}^* \\
 (\eta_x)_- / \mathbf{K}^* \downarrow & & \downarrow (\eta_x)_+ / \mathbf{K}^* \\
 (X_x)_- / \mathbf{K}^* & \xrightarrow{\quad} & (X_x)_+ / \mathbf{K}^*
 \end{array}$$

où $(X_x)_- / \mathbf{K}^* \rightarrow (X_x)_+ / \mathbf{K}^*$ est une application birationnelle entre variétés toriques. Dans la partie suivante, on explique comment on peut obtenir un résultat analogue en restant dans le cadre des variétés non singulières.

5. DÉSINGULARISATION CANONIQUE ET DÉMONSTRATION DU THÉORÈME DE FACTORISATION

5.1. Désingularisation canonique

Sur un corps algébriquement clos de caractéristique nulle, on sait depuis Hironaka [Hir64] que toute variété algébrique peut être désingularisée par une suite d'éclatements le long de centres lisses. À partir de la dimension trois, il n'y a pas de désingularisation minimale et il n'y a pas de choix naturel d'un modèle non singulier pour une variété algébrique donnée. Cependant, avec les travaux de Bierstone-Milman [BiM97], Encinas-Villamayor [EnV97] et Villamayor [Vil89], on peut parler de désingularisation "canonique".

Définition. Une *résolution canonique des singularités* est un algorithme qui à toute variété algébrique X associe une suite uniquement déterminée d'éclatements le long de centres lisses $r : X^{res} \rightarrow X$ satisfaisant : pour tout morphisme lisse $Y \rightarrow X$, la variété Y^{res} est égale au produit fibré $Y \times_X X^{res}$.

Une résolution canonique des singularités a les propriétés suivantes découlant de la définition :

- les centres des éclatements sont au-dessus du lieu singulier de X ,

- toute famille d'automorphismes $(\theta_g)_{g \in G}$ d'une variété X paramétrée par une variété non singulière G se relève en une famille d'automorphismes de X^{res} . Ceci s'applique en particulier au cas de l'action d'un groupe algébrique.

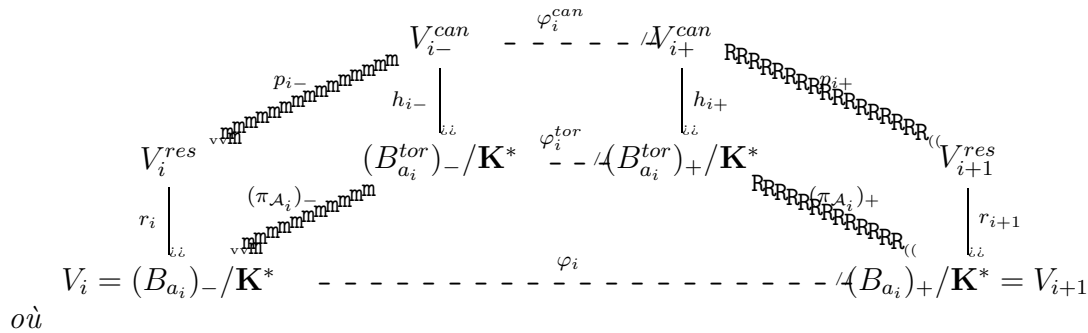
De telles résolutions canoniques des singularités existent d'après Bierstone-Milman, Encinas-Villamayor, Hironaka ou Villamayor.

Les algorithmes connus de résolutions canoniques des singularités ont la propriété supplémentaire suivante : *pour tout faisceau d'idéaux cohérent $I \subset \mathcal{O}_X$ sur une variété non singulière X , il existe une suite uniquement déterminée d'éclatements le long de centres lisses $p : X^{can} \rightarrow X$ de sorte que $p^{-1}(I)\mathcal{O}_{X^{can}}$ soit principal et pour tout morphisme lisse $f : Y \rightarrow X$, la suite correspondante d'éclatements $p' : Y^{can} \rightarrow Y$ de sorte que $(p')^{-1}(f^*I)\mathcal{O}_{Y^{can}}$ soit principal est égale au produit fibré $Y \times_X X^{can} \rightarrow Y$. (Attention, $p : X^{can} \rightarrow X$ dépend de I mais nous n'avons pas inclus I dans la notation pour ne pas alourdir la suite.)*

Dorénavant, on se fixe une résolution canonique des singularités.

La proposition suivante est le cœur de cette partie :

PROPOSITION 5.1. — *Avec les notations du corollaire 4.6, il y a un diagramme commutatif de variétés algébriques :*



- où
- (i) p_{i-} et p_{i+} sont deux suites d'éclatements le long de centres lisses,
 - (ii) si $U_{i-}^{can} = h_{i-}^{-1}((U_{I_{\mathcal{A}_i}}^{tor})_- / \mathbf{K}^*)$ et $U_{i+}^{can} = h_{i+}^{-1}((U_{I_{\mathcal{A}_i}}^{tor})_+ / \mathbf{K}^*)$, alors $(U_{i-}^{can} \subset V_{i-}^{can})$ et $(U_{i+}^{can} \subset V_{i+}^{can})$ sont des plongements toroïdaux et φ_i^{can} est birationnelle toroïdale.

Expliquons la construction de ce diagramme. Il s'agit de montrer dans un premier temps que $(\pi_{\mathcal{A}_i})_-$ est l'éclatement d'un faisceau d'idéaux $(I_i)_-$ sur V_i . C'est plus facile si l'on suppose que la famille admissible \mathcal{A}_i est choisie vérifiant

$$\sum_{\alpha \in \mathcal{A}_i} \alpha = 0,$$

ce qu'il est toujours possible de faire en rajoutant un entier à la famille admissible initialement choisie. Alors, le faisceau toroïdal $I_{\mathcal{A}_i}$ sur B_{a_i} est engendré par des fonctions invariantes d'après le lemme 4.2, donc provient d'un faisceau d'idéaux I_i sur $B_{a_i} // \mathbf{K}^*$. Le faisceau d'idéaux $(I_i)_-$ sur V_i est alors obtenu par image inverse $V_i = (B_{a_i})_- / \mathbf{K}^* \rightarrow B_{a_i} // \mathbf{K}^*$. Le morphisme $p_{i-} : V_{i-}^{can} \rightarrow V_i^{res}$ est la suite d'éclatements uniquement déterminée rendant

$r_i^{-1}((I_i)_-)\mathcal{O}_{V_i^{res}}$ principal et h_{i-} est l'unique morphisme induit par la propriété universelle de l'éclatement d'un faisceau d'idéaux. La vérification du point (ii) se fait à nouveau dans les cartes toriques fortement étales utilisées pour construire le faisceau toroïdal $I_{\mathcal{A}_i}$, on renvoie à [AKMW99] pour les détails.

COROLLAIRE 5.2. — *Soit $\varphi : X_1 \rightarrow X_2$ une application régulière birationnelle entre deux variétés projectives lisses. Alors, φ se factorise en une suite d'éclatements et de contractions de centres lisses.*

Démonstration. C'est une conséquence immédiate de tout ce qui précède : il suffit d'appliquer le théorème 3.6 à chaque φ_i^{can} et de remarquer que puisque X_1 et X_2 sont lisses, alors $X_1 = V_0 = V_0^{res}$ et $X_2 = V_{m+1} = V_{m+1}^{res}$. ■

5.2. Fin de la démonstration du théorème de factorisation

Elle se termine exactement comme dans le cas torique, à l'aide du théorème de Moishezon [Moi67]. La levée des indéterminations, due à Hironaka [Hir75] dans le cadre général se formule ainsi :

THÉORÈME 5.3. — *Soit $\varphi : X \dashrightarrow X'$ une application birationnelle entre deux variétés complètes lisses. Alors il existe une suite d'éclatements le long de sous-variétés lisses $\psi : \tilde{X} \rightarrow X$ telle que $\varphi \circ \psi : \tilde{X} \rightarrow X'$ soit une application régulière birationnelle.*

Le lemme de Chow (voir [Har77] p. 107) est lui aussi valable en général :

THÉORÈME 5.4. — *Soit X une variété algébrique complète. Alors il existe une variété projective lisse \tilde{X} et une application régulière birationnelle $\varphi : \tilde{X} \rightarrow X$.*

On déduit de ces résultats le théorème de Moishezon [Moi67] (dont la preuve initiale, antérieure à la levée des indéterminations est évidemment beaucoup plus difficile) comme dans le cas torique :

THÉORÈME 5.5. — *Soit X une variété algébrique lisse et complète. Alors il existe une variété projective lisse \tilde{X} obtenue à partir de X par suite d'éclatements le long de sous-variétés lisses $\psi : \tilde{X} \rightarrow X$.*

Ceci permet enfin de ramener le théorème de factorisation au corollaire 5.2.

6. APPENDICE : DÉMONSTRATION DU THÉORÈME DE π -DÉSINGULARISATION

6.1. Outils et résultats intermédiaires principaux.

Les notes qui suivent sont une reprise de [AMR99], avec une simplification de la démonstration de la proposition 6.5. *Merci à Adam Parusinski pour sa lecture attentive de cet appendice.*

6.1.1. *Relation de π -dépendance, π -multiplicité et π -profil de multiplicité.* Soit

$$\eta = \langle \rho_1, \dots, \rho_k \rangle$$

un cône simplicial de N^+ et pour $1 \leq i \leq k$, soit v_i le générateur dans N de l'arête engendrée par $\pi(\rho_i)$. Il existe alors un unique w_i dans \mathbf{Q} tel que ρ_i soit proportionnel à (v_i, w_i) .

Définition. Si η est π -indépendant, on définit la π -multiplicité de η , notée $\pi\text{-mult}(\eta)$ comme étant la multiplicité du cône $\pi(\eta)$ (égale à l'indice dans le réseau engendré par $\pi(\eta)$ du sous-groupe $\mathbf{Z}v_1 \oplus \dots \oplus \mathbf{Z}v_k$). Remarquons que η est π -non-singulier si et seulement si $\pi\text{-mult}(\eta) = 1$. On définit le π -profil de multiplicité de η , noté $\pi\text{-mp}(\eta)$, comme étant le quadruplet $\pi\text{-mp}(\eta) := (\pi\text{-mult}(\eta), 0, 0, 0)$.

Si η est π -dépendant, comme le noyau de π est de dimension 1, il existe une unique *relation de π -dépendance*

$$\sum_{i=1}^k r_i v_i = 0 \text{ avec } \max\{|r_i|; 1 \leq i \leq k\} = 1 \text{ et } \sum_{i=1}^k r_i w_i > 0.$$

Pour $1 \leq i \leq k$, on note η_i la face de codimension 1 de η suivante :

$$\eta_i = \langle \rho_1, \dots, \check{\rho}_i, \dots, \rho_k \rangle.$$

Le lemme suivant est élémentaire et essentiel :

LEMME 6.1. — (i) La face η_i est π -indépendante si et seulement si $r_i \neq 0$.

(ii) Si η_i et η_j sont toutes deux π -indépendantes, alors $|r_j| \pi\text{-mult}(\eta_i) = |r_i| \pi\text{-mult}(\eta_j)$.

Notation. Introduisons les ensembles : $I_1(\eta) = \{i; r_i = 1\}$, $I_+(\eta) = \{i; r_i > 0\}$, $I_{-1}(\eta) = \{i; r_i = -1\}$, $I_-(\eta) = \{i; r_i < 0\}$ et $i_1(\eta)$, $i_+(\eta)$, $i_{-1}(\eta)$ et $i_-(\eta)$ leur cardinal respectif.

Définition. Si η est π -dépendant, on définit la π -multiplicité de η comme étant le maximum des π -multiplicités des faces η_i pour $i \in I_+(\eta) \cup I_-(\eta)$. Autrement dit, $\pi\text{-mult}(\eta)$ est le maximum des π -multiplicités des faces π -indépendantes de η . On définit le π -profil de multiplicité de η , noté $\pi\text{-mp}(\eta)$, comme étant le quadruplet

$$\pi\text{-mp}(\eta) := (\pi\text{-mult}(\eta), 0, 0, 0) \text{ si } i_1(\eta) + i_{-1}(\eta) = 1,$$

et

$$\pi\text{-mp}(\eta) := (\pi\text{-mult}(\eta), 1, i_+(\eta) + i_-(\eta), i_1(\eta) + i_{-1}(\eta)) \text{ si } i_1(\eta) + i_{-1}(\eta) \geq 2.$$

Remarquons que la quantité $k_\eta = i_+(\eta) + i_-(\eta)$ est la dimension de l'unique circuit contenu dans η .

Notations. Soit Σ un éventail de $N_{\mathbf{Q}}^+$. On note g_{Σ} le maximum (pour l'ordre lexicographique) des $\pi\text{-mp}(\eta)$ lorsque η décrit tous les cônes maximaux de Σ (*i.e.* non contenus strictement dans un cône de Σ) et s_{Σ} le nombre de cônes maximaux atteignant ce maximum. Alors le π -profil de multiplicité de Σ est le quintuplet :

$$\pi\text{-mp}(\Sigma) := (g_{\Sigma}, s_{\Sigma}).$$

6.1.2. *Faces codéfinies.* Soit $\eta = \langle \rho_1, \dots, \rho_k \rangle$ un cône π -dépendant de $N_{\mathbf{Q}}^+$ et τ une face π -indépendante de η . Si

$$\sum_{i=1}^k r_i v_i = 0$$

est la relation de π -dépendance de η , on dit que τ est *codéfinie* par rapport à η si ses générateurs sont inclus dans $\{\rho_i, r_i \geq 0\}$ ou dans $\{\rho_i, r_i \leq 0\}$. Évidemment, si τ est une face de τ' où τ' est codéfinie par rapport à η , alors τ est codéfinie par rapport à η .

Si $\tau = \langle \rho_1, \dots, \rho_l \rangle$ est un cône π -indépendant de $N_{\mathbf{Q}}^+$, et si v_i est le générateur dans N de l'arête engendrée par $\pi(\rho_i)$, on note :

$$\text{par}(\pi(\tau)) = \{v \in N; v = \sum_{i=1}^l a_i v_i, 0 < a_i < 1\}.$$

Remarquons que si η est un cône π -singulier de $N_{\mathbf{Q}}^+$, et si τ est une face π -indépendante et π -singulière de η de dimension minimale, alors $\text{par}(\pi(\tau)) \neq \emptyset$.

La proposition suivante est le résultat central de ce paragraphe. Elle illustre parfaitement l'utilité des faces codéfinies.

PROPOSITION 6.2. — *Soit η un cône π -dépendant de N^+ et τ une face π -indépendante de η . Soit $v \in \text{par}(\pi(\tau))$ et $\rho \in N_{\mathbf{Q}}^+$ dans l'intérieur relatif de τ tel que $\pi(\rho) = v$. Soit η' l'éventail obtenu par subdivision étoilée de η par rapport à ρ . Si τ est codéfinie par rapport à η , alors $\pi\text{-mp}(\eta') \leq \pi\text{-mp}(\eta)$ et si de plus τ est contenue dans une face γ de codimension 1 de η de π -multiplicité maximale, alors $\pi\text{-mp}(\eta') < \pi\text{-mp}(\eta)$.*

Démonstration. Notons $\tau = \langle \rho_1, \dots, \rho_m \rangle$ et $\eta = \langle \rho_1, \dots, \rho_n \rangle$ avec $n > m$. On note $\rho = \sum_{i=1}^m a_i \rho_i$ et soit $\sum_{i=1}^n r_i v_i = 0$ la relation de π -dépendance de η .

Les cônes maximaux de η' sont les $\eta_{\alpha} = \langle \rho, \rho_1, \dots, \check{\rho}_{\alpha}, \dots, \rho_m, \dots, \rho_n \rangle$ pour $1 \leq \alpha \leq m$. Nous allons estimer $\pi\text{-mp}(\eta_{\alpha})$.

Les faces nouvelles de η_{α} (*i.e.* les faces de η_{α} qui ne sont pas face de η) sont les

$$\gamma_{\alpha\beta} = \langle \rho, \rho_1, \dots, \check{\rho}_{\alpha}, \dots, \check{\rho}_{\beta}, \dots, \rho_n \rangle$$

pour $\beta \neq \alpha$. Supposons que $\gamma_{\alpha\beta}$ est π -indépendante et remarquons que cela implique que $r_{\alpha} \neq 0$ ou $r_{\beta} \neq 0$.

- si $\beta \geq m + 1$, alors

$$\begin{aligned}\pi\text{-mult}(\gamma_{\alpha\beta}) &= |\det(v, v_1, \dots, \check{v}_\alpha, \dots, \check{v}_\beta, \dots, v_n)| \\ &= a_\alpha |\det(v_1, \dots, \check{v}_\beta, \dots, v_n)| \\ &\leq a_\alpha \pi\text{-mult}(\eta) < \pi\text{-mult}(\eta).\end{aligned}$$

- si $\beta \leq m$, alors

$$\begin{aligned}\pi\text{-mult}(\gamma_{\alpha\beta}) &= |a_\alpha \det(v_\alpha, v_1, \dots, \check{v}_\alpha, \dots, \check{v}_\beta, \dots, v_n) \\ &\quad + a_\beta \det(v_\beta, v_1, \dots, \check{v}_\alpha, \dots, \check{v}_\beta, \dots, v_n)|.\end{aligned}$$

Pour $\lambda = \alpha$ ou β , posons $m_\lambda = \det(v_\lambda, v_1, \dots, \check{v}_\alpha, \dots, \check{v}_\beta, \dots, v_n)$. De $\sum_{i=1}^n r_i v_i = 0$, on en déduit que $r_\alpha m_\alpha + r_\beta m_\beta = 0$, et τ étant codéfinie par rapport à η , r_α et r_β sont de même signe donc m_α et m_β sont de signes opposés. De là

$$\begin{aligned}\pi\text{-mult}(\gamma_{\alpha\beta}) &= |a_\alpha m_\alpha + a_\beta m_\beta| \\ &\leq \max(a_\alpha |m_\alpha|, a_\beta |m_\beta|) \\ &\leq \max(a_\alpha, a_\beta) \pi\text{-mult}(\eta) < \pi\text{-mult}(\eta).\end{aligned}$$

Ainsi, pour toute nouvelle face $\gamma_{\alpha\beta}$, on a $\pi\text{-mult}(\gamma_{\alpha\beta}) < \pi\text{-mult}(\eta)$.

Enfin, chaque η_α possède une unique ancienne face de codimension un :

$$\gamma_\alpha = \langle \rho_1, \dots, \check{\rho}_\alpha, \dots, \rho_n \rangle$$

et cette face satisfait évidemment $\pi\text{-mult}(\gamma_\alpha) \leq \pi\text{-mult}(\eta)$.

Le bilan est le suivant : si $\pi\text{-mult}(\gamma_\alpha) = \pi\text{-mult}(\eta)$, alors $\pi\text{-mp}(\eta_\alpha) = (\pi\text{-mult}(\eta), 0, 0, 0)$ et si $\pi\text{-mult}(\gamma_\alpha) < \pi\text{-mult}(\eta)$, alors $\pi\text{-mp}(\eta_\alpha) < (\pi\text{-mult}(\eta), 0, 0, 0)$.

Notons s' le nombre de faces de codimension 1 de η ne contenant pas τ et de π -multiplicité maximale et r_η le nombre de faces de codimension 1 de η de π -multiplicité maximale.

Alors,

- si $s' = 0$, ce qui précède montre que $\pi\text{-mp}(\eta_\alpha) < (\pi\text{-mult}(\eta), 0, 0, 0)$ pour tout α , $1 \leq \alpha \leq m$, et donc $\pi\text{-mp}(\eta') < \pi\text{-mp}(\eta)$.
- si $s' = 1$, on a

$$\pi\text{-mp}(\eta') = ((\pi\text{-mult}(\eta), 0, 0, 0), 1) \leq \pi\text{-mp}(\eta).$$

De plus, cette inégalité est une égalité si et seulement si $r_\eta = s' = 1$, ce qui ne se produit pas dans le cas où τ est contenue dans une face de codimension 1 de η de π -multiplicité maximale.

- si $s' \geq 2$, on a

$$\pi\text{-mp}(\eta') = ((\pi\text{-mult}(\eta), 0, 0, 0), s') < \pi\text{-mp}(\eta) = ((\pi\text{-mult}(\eta), 1, k_\eta, r_\eta), 1).$$

■

La proposition 6.2 nous encourage à fabriquer des faces codéfinies. Ce sera l'objet du paragraphe suivant. Énonçons dès maintenant le lemme immédiat suivant :

LEMME 6.3. — Soient η un cône π -dépendant de N^+ , τ une face π -indépendante de η et σ l'unique circuit contenu dans η . Si $\dim(\sigma) \leq 2$, alors τ est codéfinie par rapport à η .

6.1.3. *Subdivision étoilée positive et négative.* Soient Σ un éventail de $N_{\mathbf{Q}}^+$ et $\sigma = \langle \rho_1, \dots, \rho_k \rangle$ un circuit de Σ . Si

$$\sum_{i=1}^k r_i v_i = 0$$

est la relation de π -dépendance de σ , on pose

$$v_+ = \sum_{i \in I_+(\sigma)} v_i \text{ et } v_- = \sum_{i \in I_-(\sigma)} v_i.$$

(Attention, v_+ et v_- ne sont pas nécessairement primitifs.) Remarquons que comme σ est un circuit, on a $I_+(\sigma) \cup I_-(\sigma) = \{1, \dots, k\}$, autrement dit, tous les r_i de la relation de π -dépendance de σ sont non nuls.

Comme

$$v_+ = \sum_{i \in I_+(\sigma)} v_i = \sum_{i \in I_+(\sigma)} (1 - \varepsilon r_i) v_i + \sum_{i \in I_-(\sigma)} (-\varepsilon r_i) v_i,$$

pour ε petit, v_+ est de la forme $\sum_{i=1}^k c_i v_i$ où tous les c_i sont strictement positifs. Il existe donc ρ_+ dans l'intérieur relatif de σ tel que $\pi(\rho_+) = v_+$. Soit Σ^+ l'éventail obtenu par subdivision étoilée de Σ par rapport à ρ_+ . On l'appelle *subdivision étoilée positive* de Σ par rapport à σ . (Cette construction dépend du choix de ρ_+ mais les propriétés que nous énoncerons n'en dépendent pas.) La même construction à partir de v_- donne lieu à la *subdivision étoilée négative* Σ^- de Σ par rapport à σ .

Malheureusement, le π -profil de multiplicité de Σ^+ (resp. Σ^-) n'est en général pas inférieur ou égal à celui de Σ . Le lemme facile suivant illustre néanmoins l'intérêt des subdivisions étoilées positive et négative :

LEMME 6.4. — Soient η un cône π -dépendant de $N_{\mathbf{Q}}^+$, γ une face de codimension 1 de η de π -multiplicité maximale. Soit η^+ (resp. η^-) l'éventail obtenu par subdivision étoilée positive (resp. négative) de η par rapport à l'unique circuit σ contenu dans η . Alors γ est codéfinie par rapport au cône maximal de η^+ (resp. η^-) contenant γ .

La proposition suivante joue un rôle crucial dans la démonstration du théorème de π -désingularisation.

PROPOSITION 6.5. — Soit σ un circuit de $N_{\mathbf{Q}}^+$ de dimension strictement supérieure à 2. Alors soit l'éventail σ^+ obtenu par subdivision étoilée positive de σ , soit l'éventail σ^- obtenu par subdivision étoilée négative de σ satisfait l'une des deux propriétés suivantes (nous notons σ' l'éventail σ^+ ou l'éventail σ^- qui convient) :

A) pour tout cône maximal δ' de σ' , on a $\pi\text{-mp}(\delta') < \pi\text{-mp}(\sigma)$; en particulier

$$\pi\text{-mp}(\sigma') < \pi\text{-mp}(\sigma).$$

B) il existe un cône maximal κ' de σ' vérifiant $\pi\text{-mp}(\kappa') = \pi\text{-mp}(\sigma)$ et tel que tout cône maximal $\delta' \neq \kappa'$ de σ' vérifie $\pi\text{-mp}(\delta') < \pi\text{-mp}(\sigma)$; en particulier

$$\pi\text{-mp}(\sigma') = \pi\text{-mp}(\sigma).$$

De plus, l'unique ancienne face γ' de codimension 1 de κ' est π -indépendante, et vérifie $\pi\text{-mult}(\gamma') = \pi\text{-mult}(\kappa') = \pi\text{-mult}(\sigma)$. En particulier, γ' est codéfinie par rapport à κ' d'après le lemme 6.4.

Avant de commencer la démonstration proprement dite, commençons par un calcul intermédiaire : notons $\sigma = \langle \rho_1, \dots, \rho_n \rangle$, $\sum_{i=1}^n r_i v_i = 0$ la relation de π -dépendance de σ et $v_+ = \sum_{i \in I_+(\sigma)} v_i = e\bar{v}_+$ avec \bar{v}_+ primitif. Soit enfin ρ_+ dans l'intérieur relatif de σ tel que $\pi(\rho_+) = v_+$. Les cônes maximaux de σ^+ sont de la forme $\sigma_\alpha = \langle \rho_+, \rho_1, \dots, \check{\rho}_\alpha, \dots, \rho_n \rangle$ et nous devons estimer leur π -profil de multiplicité ; chacun d'eux a une face ancienne de codimension un $\gamma_\alpha = \langle \rho_1, \dots, \check{\rho}_\alpha, \dots, \rho_n \rangle$ et de nouvelles faces

$$\gamma_{\alpha\beta} = \langle \rho_+, \rho_1, \dots, \check{\rho}_\alpha, \dots, \check{\rho}_\beta, \dots, \rho_n \rangle.$$

Les calculs de π -multiplicité des faces de σ_α sont faciles :

- si $r_\alpha > 0$ alors

$$\pi\text{-mult}(\gamma_{\alpha\beta}) = \begin{cases} \frac{1}{e} |\pi\text{-mult}(\gamma_\alpha) - \pi\text{-mult}(\gamma_\beta)| & \text{si } r_\beta > 0 \\ \frac{1}{e} \pi\text{-mult}(\gamma_\beta) & \text{si } r_\beta < 0, \end{cases}$$

- si $r_\alpha < 0$ alors

$$\pi\text{-mult}(\gamma_{\alpha\beta}) = \begin{cases} \frac{1}{e} \pi\text{-mult}(\gamma_\alpha) & \text{si } r_\beta > 0 \\ 0 & \text{si } r_\beta < 0. \end{cases}$$

Démonstration de la proposition 6.5.

- supposons qu'une seule face de codimension 1 de σ est de π -multiplicité maximale, autrement dit $\pi\text{-mp}(\sigma) = (\pi\text{-mult}(\sigma), 0, 0, 0)$. Ceci signifie que $i_1(\sigma) + i_{-1}(\sigma) = 1$ et on peut supposer que $i_{-1}(\sigma) = 0$ et $i_1(\sigma) = 1$, on note alors α_0 le seul indice tel que $r_{\alpha_0} = 1$. Montrons que σ^+ satisfait la propriété B).

En effet, les formules précédentes montrent que $\pi\text{-mult}(\sigma_\alpha) < \pi\text{-mult}(\sigma)$ pour $\alpha \neq \alpha_0$ (et donc $\pi\text{-mp}(\sigma_\alpha) < \pi\text{-mp}(\sigma)$ pour $\alpha \neq \alpha_0$) et que $\pi\text{-mult}(\sigma_{\alpha_0}) = \pi\text{-mult}(\sigma)$, la seule face de codimension 1 de σ_{α_0} de π -multiplicité maximale étant γ_{α_0} . On en déduit que $\pi\text{-mp}(\sigma_{\alpha_0}) = \pi\text{-mp}(\sigma)$, et γ_{α_0} , étant de π -multiplicité maximale, est codéfinie par rapport à σ_{α_0} d'après le lemme 6.4.

- supposons que $\pi\text{-mp}(\sigma) = (\pi\text{-mult}(\sigma), 1, i_+(\sigma) + i_-(\sigma), i_1(\sigma) + i_{-1}(\sigma))$. Dans ce cas, $i_1(\sigma) + i_{-1}(\sigma) \geq 2$ et on peut supposer que $i_1(\sigma) \geq 1$. Montrons que σ^+ satisfait l'une des propriétés A) ou B) sauf si $i_{-1}(\sigma) = i_-(\sigma) = 1$.

- (i) si $e > 1$, les formules précédentes montrent que $\pi\text{-mult}(\sigma_\alpha) \leq \pi\text{-mult}(\sigma)$ pour tout α , et que si $\pi\text{-mult}(\sigma_\alpha) = \pi\text{-mult}(\sigma)$, la seule face de codimension 1 de σ_α de π -multiplicité maximale est γ_α . On en déduit que $\pi\text{-mp}(\sigma_\alpha) \leq (\pi\text{-mult}(\sigma), 0, 0, 0) < \pi\text{-mp}(\sigma)$ pour tout α et donc que σ^+ satisfait la propriété A).
- (ii) si $e = 1$, évaluons séparément $\pi\text{-mp}(\sigma_\alpha)$ suivant que $r_\alpha = -1$, $-1 < r_\alpha < 0$, $r_\alpha = 1$ ou $1 > r_\alpha > 0$.

a) si $-1 < r_\alpha < 0$, les formules précédentes montrent que $\pi\text{-mult}(\sigma_\alpha) < \pi\text{-mult}(\sigma)$ et donc que $\pi\text{-mp}(\sigma_\alpha) < \pi\text{-mp}(\sigma)$.

b) si $1 > r_\alpha > 0$, les formules précédentes montrent que

$$\pi\text{-mp}(\sigma_\alpha) \begin{cases} \leq (\pi\text{-mult}(\sigma), 0, 0, 0) < \pi\text{-mp}(\sigma) & \text{si } i_{-1}(\sigma) \leq 1, \\ = (\pi\text{-mult}(\sigma), 1, k_{\sigma_\alpha}, i_{-1}(\sigma)) & \text{si } i_{-1}(\sigma) \geq 2. \end{cases}$$

Dans le deuxième cas, $k_{\sigma_\alpha} \leq k_\sigma = \dim(\sigma)$ et $i_{-1}(\sigma) < i_1(\sigma) + i_{-1}(\sigma)$ donc $\pi\text{-mp}(\sigma_\alpha) < \pi\text{-mp}(\sigma)$.

c) si $r_\alpha = 1$, les formules précédentes montrent que

$$\pi\text{-mp}(\sigma_\alpha) = \begin{cases} (\pi\text{-mult}(\sigma), 1, k_{\sigma_\alpha}, 1 + i_{-1}(\sigma)) & \text{si } i_{-1}(\sigma) \neq 0, \\ (\pi\text{-mult}(\sigma), 0, 0, 0) < \pi\text{-mp}(\sigma) & \text{si } i_{-1}(\sigma) = 0. \end{cases}$$

Dans le premier cas, si $i_1(\sigma) \geq 2$, σ_α possède une face de codimension 1 π -dépendante (car de π -multiplicité nulle) donc $k_{\sigma_\alpha} < k_\sigma$ et par suite $\pi\text{-mp}(\sigma_\alpha) < \pi\text{-mp}(\sigma)$; sinon $i_1(\sigma) = 1$, par suite $\pi\text{-mp}(\sigma_\alpha) = \pi\text{-mp}(\sigma)$ et γ_α , étant de π -multiplicité maximale, est codéfinie par rapport à σ_α d'après le lemme 6.4.

d) si $r_\alpha = -1$, les formules précédentes montrent que

$$\pi\text{-mp}(\sigma_\alpha) = (\pi\text{-mult}(\sigma), 1, k_{\sigma_\alpha}, 1 + i_+(\sigma)).$$

(Remarquons que $i_+(\sigma) \geq i_1(\sigma) \geq 1$). Si $i_-(\sigma) \geq 2$, σ_α possède une face de codimension 1 π -dépendante (car de π -multiplicité nulle) donc $k_{\sigma_\alpha} < k_\sigma$ et par suite $\pi\text{-mp}(\sigma_\alpha) < \pi\text{-mp}(\sigma)$; si $i_-(\sigma) = 1$, alors

$$1 + i_+(\sigma) \geq i_1(\sigma) + i_{-1}(\sigma) = i_1(\sigma) + 1$$

et on ne peut conclure à ce stade de la démonstration.

Le bilan est cependant le suivant : si $i_1(\sigma) + i_{-1}(\sigma) \geq 2$ et $i_1(\sigma) \geq 1$ alors σ^+ satisfait l'une des propriétés A) ou B) sauf si $i_{-1}(\sigma) = i_-(\sigma) = 1$. De façon symétrique, si $i_1(\sigma) + i_{-1}(\sigma) \geq 2$ et $i_{-1}(\sigma) \geq 1$ alors σ^- satisfait l'une des propriétés A) ou B) sauf si $i_1(\sigma) = i_+(\sigma) = 1$.

Le seul cas restant est donc celui où

$$i_{-1}(\sigma) = i_-(\sigma) = i_1(\sigma) = i_+(\sigma) = 1,$$

mais alors $\dim(\sigma) = 2$ ce qui est exclu par l'hypothèse.

■

Nous aurons besoin du lemme suivant qui précise la proposition 6.5 :

LEMME 6.6. — Soit Σ un éventail de N^+ , σ un circuit de Σ de dimension strictement supérieure à 2. Si $\overline{\text{Star}(\sigma)'}$ désigne l'éventail obtenu par la subdivision étoilée positive ou négative de $\overline{\text{Star}(\sigma)}$ donnée par la proposition précédente, alors

$$\pi\text{-mp}(\overline{\text{Star}(\sigma)'}) < \pi\text{-mp}(\overline{\text{Star}(\sigma)}) \text{ dans le cas A)}$$

et

$$\pi\text{-mp}(\overline{\text{Star}(\sigma)'}) = \pi\text{-mp}(\overline{\text{Star}(\sigma)}) \text{ dans le cas B) .}$$

Indication de démonstration. On montre que si

$$\pi\text{-mp}(\sigma) = (\pi\text{-mult}(\sigma), b_\sigma, k_\sigma, r_\sigma, 1) \text{ et } \pi\text{-mp}(\sigma') = (\pi\text{-mult}(\sigma'), b_{\sigma'}, k_{\sigma'}, r_{\sigma'}, s'),$$

alors il existe des entiers e et s supérieurs ou égaux à 1 tels que

$$\pi\text{-mp}(\overline{\text{Star}(\sigma)}) = (\pi\text{-mult}(\sigma)e, b_\sigma, k_\sigma, r_\sigma, s)$$

et

$$\pi\text{-mp}(\overline{\text{Star}(\sigma)'}) = (\pi\text{-mult}(\sigma')e, b_{\sigma'}, k_{\sigma'}, r_{\sigma'}, s's).$$

■

6.2. Démonstration du théorème de π -désingularisation.

Soit Σ un éventail de N^+ . La stratégie est claire : si Σ est π -non-singulier, il n'y a rien à faire. Sinon, il suffit de construire un éventail Σ_1 obtenu par une suite finie de subdivisions étoilées de Σ n'affectant pas les cônes π -non-singuliers de Σ et vérifiant $\pi\text{-mp}(\Sigma_1) < \pi\text{-mp}(\Sigma)$. Si Σ_1 est π -non-singulier, c'est fini, sinon on recommence. Ce procédé doit s'arrêter après un nombre fini d'étapes et l'éventail obtenu est π -non-singulier. La construction de Σ_1 se fait en trois étapes, illustrées par la figure 4.

Étape 1. Posons $\pi\text{-mp}(\Sigma) = (g_\Sigma, s_\Sigma)$ et choisissons η un cône maximal de Σ de π -profil de multiplicité maximal (*i.e.* égal à g_Σ) et soit σ l'unique circuit contenu dans η .

(i) Supposons que $\dim(\sigma) > 2$ et appliquons alors la proposition 6.5 ; on note Σ' l'éventail ainsi obtenu et η' le sous-éventail de Σ' obtenu en subdivisant η .

- Dans le cas A), le lemme 6.6 assure que $\pi\text{-mp}(\Sigma') < \pi\text{-mp}(\Sigma)$. On pose alors $\Sigma_1 := \Sigma$.
- Dans le cas B), le lemme 6.6 assure que $\pi\text{-mp}(\Sigma') = \pi\text{-mp}(\Sigma)$. Soit γ l'unique face de codimension 1 de η telle que $\gamma \cap \sigma = \gamma'$ (γ' donnée par la proposition 6.5). Alors γ est codéfinie par rapport à l'unique cône maximal ν de η' dont elle est face, elle est de plus de π -multiplicité maximale.

- (ii) Supposons que $\dim(\sigma) \leq 2$ et choisissons une face γ de codimension 1 de η , de π -multiplicité maximale. Comme $\dim(\sigma) \leq 2$, γ est codéfinie par rapport à η .

Le bilan de l'Étape 1 est le suivant : nous avons construit un éventail Σ' , subdivision étoilée de Σ , vérifiant $\pi\text{-mp}(\Sigma') \leq \pi\text{-mp}(\Sigma)$ avec

- (i) un cône maximal ν de π -profil de multiplicité maximal,
- (ii) une face γ de codimension 1 de ν , de π -multiplicité maximale, codéfinie par rapport à ν .

Choisissons alors τ une face de γ (donc codéfinie par rapport à ν), π -singulière de dimension minimale, $v \in \text{par}(\pi(\tau))$ et $\rho \in N_{\mathbf{Q}}^+$ dans l'intérieur relatif de τ tel que $\pi(\rho) = v$.

Le problème à ce stade est que τ , qui est codéfinie par rapport à ν , ne l'est en général pas par rapport aux autres cônes maximaux qui la contiennent.

Étape 2. Énonçons la proposition suivante, dont la démonstration sera donnée à la fin de ce paragraphe :

PROPOSITION 6.7. — *Soit Σ un éventail de N^+ , σ un circuit de Σ . Soit τ dans $\overline{\text{Star}(\sigma)}$, π -indépendante. Alors il y a une subdivision de $\overline{\text{Star}(\sigma)}$, notée $\overline{\text{Star}(\sigma)'}$, obtenue par une suite finie de subdivisions étoilées positives ou négatives par rapport à des circuits successifs contenus dans σ telle que :*

- (i) $\pi\text{-mp}(\overline{\text{Star}(\sigma)'}) \leq \pi\text{-mp}(\overline{\text{Star}(\sigma)})$,
- (ii) τ est une face de $\overline{\text{Star}(\sigma)'}$ (i.e. τ n'est pas affectée par les subdivisions) et τ est codéfinie par rapport à tous les cônes maximaux de $\overline{\text{Star}(\sigma)'}$ qui la contiennent.

Appliquons cette proposition de la façon suivante à notre situation : on considère l'ensemble des circuits σ' différents de celui contenu dans ν tels que τ est contenu dans $\overline{\text{Star}(\sigma')}$ et on applique la proposition précédente à chacun d'eux (le résultat obtenu ne dépend pas de l'ordre dans lequel on a considéré les différents σ').

Le bilan de l'Étape 2 est alors le suivant : nous avons construit un éventail Σ'' , obtenu par une suite de subdivisions étoilées de Σ vérifiant $\pi\text{-mp}(\Sigma'') \leq \pi\text{-mp}(\Sigma)$, avec

- (i) un cône maximal ν de π -profil de multiplicité maximal,
- (ii) une face γ de codimension 1 de ν , de π -multiplicité maximale, codéfinie par rapport à ν ,
- (iii) une face τ de γ (donc codéfinie par rapport à ν), π -singulière de dimension minimale, $v \in \text{par}(\pi(\tau))$ et $\rho \in N_{\mathbf{Q}}^+$ dans l'intérieur relatif de τ tel que $\pi(\rho) = v$. De plus, τ est codéfinie par rapport à tous les cônes maximaux de Σ'' qui la contiennent.

Étape 3. On note Σ_1 l'éventail obtenu par subdivision étoilée de Σ'' par rapport à ρ . Comme τ est codéfinie par rapport à tous les cônes maximaux de Σ'' qui la contiennent, on a $\pi\text{-mp}(\Sigma_1) \leq \pi\text{-mp}(\Sigma'')$ d'après la proposition 6.2, et puisque τ est contenue dans une face γ de π -multiplicité maximale, on a en fait $\pi\text{-mp}(\Sigma_1) < \pi\text{-mp}(\Sigma'')$, ce qui termine la démonstration du théorème de π -désingularisation. ■

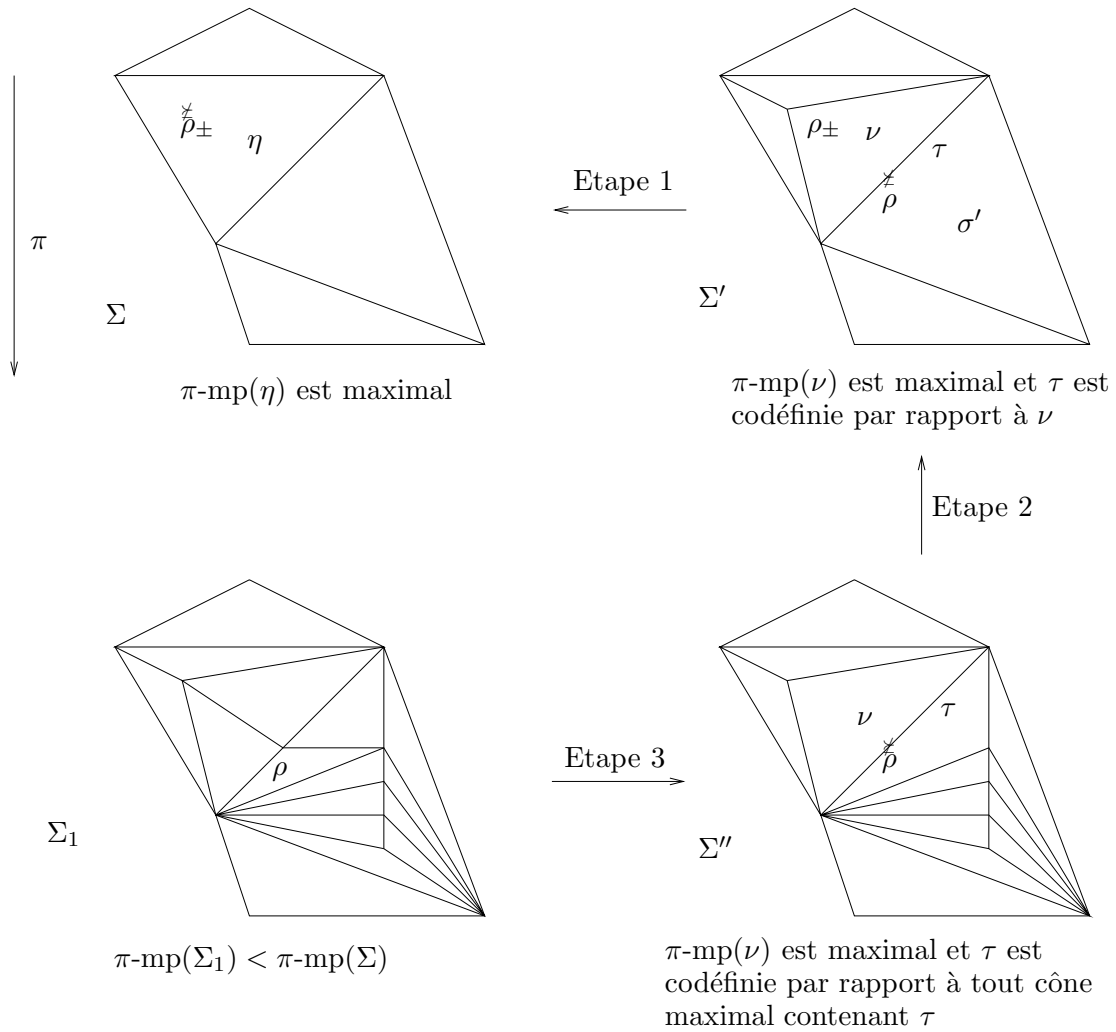


FIGURE 4. Le théorème de π -désingularisation

Démonstration de la proposition 6.7. Si $\dim(\sigma) \leq 2$, il n'y a rien à faire : τ est codéfinie par rapport à tous les cônes maximaux de $\overline{\text{Star}(\sigma)}$ qui la contiennent. On suppose dorénavant que $\dim(\sigma) > 2$.

Si $\pi\text{-mult}(\sigma) = 1$, remarquons que si η est un cône maximal contenant σ , alors toutes les faces de codimension un π -indépendantes de η ont même π -multiplicité. Soit $\overline{\text{Star}(\sigma)}^+$ la subdivision étoilée positive de $\overline{\text{Star}(\sigma)}$ par rapport à σ . Alors $\pi\text{-mp}(\overline{\text{Star}(\sigma)}^+) \leq \pi\text{-mp}(\overline{\text{Star}(\sigma)})$ par le lemme 6.6 et la face τ est codéfinie par rapport à tous les cônes maximaux de $\overline{\text{Star}(\sigma)}^+$ qui la contiennent d'après le lemme 6.4.

La démonstration de la proposition 6.7 se fait alors par récurrence sur $\pi\text{-mult}(\sigma)$: subdivisons une première fois $\overline{\text{Star}(\sigma)}$ à l'aide de la proposition 6.5. On a $\pi\text{-mp}(\overline{\text{Star}(\sigma)}^+) \leq \pi\text{-mp}(\overline{\text{Star}(\sigma)})$ par le lemme 6.6.

Si le cas A) de la proposition 6.5 se produit, les circuits contenus dans $\overline{\text{Star}(\sigma)'}^{\prime}$ sont tous de π -multiplicité strictement inférieure à celle de σ et on conclut par l'hypothèse de récurrence.

Si le cas B) de la proposition 6.5 se produit, seul le circuit $\bar{\sigma}$ contenu dans κ' pose un problème *a priori* car on ne peut pas lui appliquer l'hypothèse de récurrence. Mais si τ est face d'un cône maximal contenant κ' , alors $\tau \cap \sigma$ est incluse dans γ' . En effet : écrivons $\sigma = \langle \rho_1, \dots, \rho_k \rangle$ et quitte à renuméroter les ρ_i , on peut supposer que $\tau = \langle \rho_l, \dots, \rho_k, \dots, \rho_m \rangle$ pour certains indices l et m tels que $2 \leq l \leq k$ et $m \geq k$ (si $\tau \cap \sigma = \emptyset$, alors τ est codéfinie par rapport à tous les cônes maximaux de $\overline{\text{Star}(\sigma)}$ qui la contiennent). Le cône κ' est de la forme $\langle \rho, \rho_1, \dots, \check{\rho}_i, \dots, \rho_k \rangle$ pour un certain i tel que $1 \leq i \leq k$ et alors $\gamma' = \langle \rho_1, \dots, \check{\rho}_i, \dots, \rho_k \rangle$. Comme τ est face d'un cône maximal contenant κ' , on a $i \leq l - 1$, c'est-à-dire $\gamma' = \langle \rho_1, \dots, \check{\rho}_i, \dots, \rho_l, \dots, \rho_k \rangle$ donc

$$\tau \cap \sigma = \langle \rho_l, \dots, \rho_k \rangle \subset \gamma' = \langle \rho_1, \dots, \check{\rho}_i, \dots, \rho_l, \dots, \rho_k \rangle.$$

Finalement, comme γ' est codéfinie par rapport à κ' , τ est codéfinie par rapport à tout cône maximal contenant $\bar{\sigma}$. ■

BIBLIOGRAPHIE

- [AKMW99] D. ABRAMOVICH, K. KARU, K. MATSUKI, J. WŁODARCZYK – *Tori-fication and factorization of birational maps*. Preprint math.AG/9904135, 1999.
- [AMR99] D. ABRAMOVICH, K. MATSUKI, S. RASHID – *A note on the factorization theorem of toric birational maps after Morelli and its toroidal extension*. Tohoku Math. J. (2), 51, no. 4, (489–537) 1999.
- [AbO97] D. ABRAMOVICH, F. OORT – *Alterations and resolution of singularities*. In *Resolution of Singularities*, Progr. Math. 181, Birkhäuser, (39-108) 2000.
- [BaR85] P. BARDSLEY, R.W. RICHARDSON – *Etale slices for algebraic transformation groups in characteristic p*. Proc. London Math. Soc. (3), 51, (295-317) 1985.
- [BiM97] E. BIERSTONE, P. MILMAN – *Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant*. Invent. Math. 128 (2), (207-302) 1997.
- [BPV84] W. BARTH, C. PETERS, A. VAN DE VEN – *Compact complex surfaces*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 4, Springer-Verlag, 1984.

- [BrP90] M. BRION, C. PROCESI – *Action d'un tore dans une variété projective*. In *Operator algebras, unitary representations, enveloping algebras, and invariant theory (Paris, 1989)*, Progr. Math. 92, Birkhäuser, (509-539) 1990.
- [DCP85] C. DE CONCINI, C. PROCESI – *Complete Symmetric Varieties II*. In *Algebraic Groups and Related Topics* (R. Hotta, ed.), Adv. Studies in Pure Math. 6, (481-513) 1985.
- [Dol94] I.V. DOLGACHEV – *Introduction to geometric invariant theory*. Lecture Notes Series, 25, Seoul, 1994.
- [EnV97] S. ENCINAS, O. VILLAMAYOR – *Constructive desingularization*. In *Resolution of Singularities*, Progr. Math., 181, Birkhäuser, (39-108) 2000.
- [Ewa96] G. EWALD – *Combinatorial convexity and algebraic geometry*. Graduate Texts in Mathematics, 168, Springer-Verlag, 1996.
- [Ful84] W. FULTON – *Intersection theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 2, Springer-Verlag, 1984.
- [Ful93] W. FULTON – *Introduction to toric varieties*. Annals of mathematics studies, 131, Princeton University Press, 1993.
- [GuS89] V. GUILLEMIN, S. STERNBERG – *Birational equivalence in the symplectic category*. Invent. Math. 97, No. 3, (485-522) 1989.
- [Har77] R. HARTSHORNE – *Algebraic Geometry*. Graduate Texts in Mathematics, 52, Springer-Verlag, 1977.
- [Hir64] H. HIRONAKA – *Resolution of singularities of an algebraic variety over a field of characteristic zero*. Annals of Math. vol 79, (109-326) 1964.
- [Hir75] H. HIRONAKA – *Flattening theorem in complex analytic geometry*. Amer. J. of Math. vol. 97, no. 2, (503-547) 1975.
- [KKMS73] G. KEMPF, F. KNUDSEN, D. MUMFORD, B. SAINT-DONAT – *Toroidal embeddings I*. Lecture Notes in Mathematics, 339, Springer-Verlag, 1973.
- [Lun73] D. LUNA – *Slices étales*. Bull. Soc. Math. France, Suppl., Mem. No. 33, (81-105) 1973.
- [Mat99] K. MATSUKI – *Lectures on factorization of birational maps*. RIMS preprint math.AG/0002084, 2000.
- [Mat00] K. MATSUKI – *Erratum to the paper "A note on the factorization theorem of toric birational maps after Morelli and its toroidal extension"*. Preprint math.AG/0006038, 2000.
- [Moi67] B. MOISHEZON – *On n -dimensional compact varieties with n algebraically independent meromorphic functions*. Amer. Math. Soc. Trans. 63, (51-177) 1967.
- [Mor96] R. MORELLI – *The birational geometry of toric varieties*. J. Alg. Geom. 5, (751-782) 1996.

- [MFK94] D. MUMFORD, J. FOGARTY, F. KIRWAN – *Geometric Invariant Theory (Third Enlarged Edition)*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 034, Springer-Verlag, 1994.
- [Oda88] T. ODA – *Convex bodies and algebraic geometry : an introduction to the theory of toric varieties*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, 015, Springer-Verlag, 1988.
- [Sum74] H. SUMIHIRO – *Equivariant Completion I*. J. Math. Kyoto Univ. 14, (1-28) 1974.
- [Sum75] H. SUMIHIRO – *Equivariant Completion II*. J. Math. Kyoto Univ. 15, (573-605) 1975.
- [Vil89] O. VILLAMAYOR – *Constructiveness of Hironaka's resolution*. Ann. Sci. Ecole Norm. Sup. (4), 22, no. 1, (1-32) 1989.
- [Wlo97] J. WŁODARCZYK – *Decomposition of birational toric maps in blow-ups and blow-downs. A proof of the weak Oda conjecture*. Transactions of the AMS 349, (373-411) 1997.
- [Wlo00] J. WŁODARCZYK – *Birational cobordism and factorization of birational maps*. J. Alg. Geom. 9, no. 3, (425-449) 2000.
- [Wlod99] J. WŁODARCZYK – *Combinatorial structures on toroidal varieties and a proof of the weak factorization theorem*. Preprint math.AG/9904076, 1999.

Laurent BONAVERO

Université Grenoble I

Institut Fourier

Lab. de Mathématiques Pures

URA 5582 du CNRS

F-38402 SAINT-MARTIN D'HÈRES Cedex

E-mail : bonavero@ujf-grenoble.fr

COURBES ENTIÈRES DANS LES SURFACES ALGÈBRIQUES COMPLEXES

[d'après McQuillan, Demailly–El Goul,...]

par Marco BRUNELLA

1. QUELQUES CONJECTURES ET QUELQUES THÉORÈMES SUR LES COURBES ENTIÈRES

Soit X une variété complexe. Une *courbe entière* dans X est une application holomorphe non constante de la droite complexe \mathbf{C} à valeurs dans X . La variété X est *hyperbolique* (au sens de Brody [Kob]) si elle ne contient aucune courbe entière. Les variétés complexes compactes qui possèdent une métrique Kählérienne à courbure sectionnelle holomorphe négative sont des exemples classiques de variétés hyperboliques [Kob].

Rappelons qu'une variété projective complexe X est *de type général* si son fibré canonique $K_X = \det(\Omega_X^1)$ est *gros*, c'est-à-dire il existe $m \in \mathbf{N}$ tel que $K_X^{\otimes m}$ possède $n + 1$ ($n = \dim X$) sections holomorphes globales algébriquement indépendantes s_0, \dots, s_n (i.e., les fonctions méromorphes $\frac{s_1}{s_0}, \dots, \frac{s_n}{s_0}$ sont algébriquement indépendantes). La propriété " K_X gros" est très proche (mais un peu plus faible) de " K_X ample", ce qui équivaut d'ailleurs, du point de vue métrique, à " K_X possède une métrique hermitienne à courbure positive" et donc à " X possède une métrique Kählérienne à courbure de Ricci négative" (Aubin, Yau). Il est donc naturel de s'attendre à des contraintes sur les courbes entières dans les variétés de type général. Vers 1980, Green et Griffiths proposaient dans [GGr] la conjecture suivante.

CONJECTURE 1.1. — (*Green–Griffiths*). *Soit X une variété projective complexe de type général, et soit $f : \mathbf{C} \rightarrow X$ une courbe entière dans X . Alors l'image de f n'est pas Zariski-dense dans X : il existe une hypersurface algébrique $Y \subset X$ telle que $f(\mathbf{C}) \subset Y$.*

Ici l'hypersurface Y dépend, a priori, de la courbe f . On trouve dans les travaux de Lang (par exemple [Lan]) la conjecture un peu plus forte selon laquelle, étant donné X comme ci-dessus, il existe une sous-variété propre $Z \subset X$ telle que *toute* courbe entière dans X est contenue dans Z . En plus, selon cette même conjecture, Z coïncide avec l'adhérence de Zariski de l'union de toutes les sous-variétés de X qui sont images de

variétés abéliennes par des applications rationnelles non constantes. Ces conjectures de Lang sont intimement liées à certaines conjectures de nature arithmétique, que nous ne discuterons pas et pour lesquelles nous renvoyons à [Lan].

Si $\dim X = 1$, la conjecture 1.1 ne présente aucune difficulté : “type général” signifie dans ce cas que le genre de X est au moins 2, et donc toute application holomorphe de \mathbf{C} dans X est constante (Liouville).

Si $\dim X = n \geq 2$, la situation est par contre beaucoup plus compliquée, et même pour $n = 2$ la conjecture 1.1 n’a pas encore été entièrement démontrée. Avant d’énoncer le théorème de McQuillan, qui constitue l’objet central de cet exposé, nous allons expliquer une approche de la conjecture 1.1 (et d’autres conjectures sur les courbes entières) qui s’est révélée fructueuse.

Le point de départ est le résultat suivant, qui se trouve dans [GGr] (avec une lacune dans la preuve) et dans [Dm1] et [SY2]. Voir aussi [Dm2] pour un exposé très agréable de ces idées. Soit X une variété projective complexe et soit $f : \mathbf{C} \rightarrow X$ une courbe entière. Soit P un opérateur différentiel algébrique sur X à valeurs dans le dual A^* d’un fibré linéaire ample $A \in \text{Pic}(X)$ (nous renvoyons aux articles mentionnés pour la définition exacte de ces opérateurs différentiels, mais le lecteur peut bien s’imaginer ce que cela signifie). Alors f satisfait l’équation différentielle associée à P :

$$P(f', f'', \dots, f^{(k)}) \equiv 0.$$

La preuve de ce théorème est basée sur des arguments de courbure négative (lemme d’Ahlfors–Schwarz), ou bien sur des estimations de la théorie de Nevanlinna.

Supposons, pour fixer les idées, qu’on veut démontrer qu’une variété X est hyperbolique. D’après le résultat ci-dessus, il suffit de construire sur X un nombre assez grand d’opérateurs différentiels algébriques “indépendants” (et à valeurs dans les duaux de fibrés amples) : chacun de ces opérateurs donne une équation différentielle satisfaite par toute application holomorphe $f : \mathbf{C} \rightarrow X$, et si l’on dispose d’un nombre assez grand d’équations différentielles “indépendantes” on en déduit la constance de f . Nous expliquerons plus loin, dans des cas concrets, la notion d’indépendance requise.

La difficulté dans cette approche est qu’il n’est pas du tout facile de construire des opérateurs différentiels algébriques “indépendants”. Construire des opérateurs différentiels algébriques revient à construire des sections holomorphes globales de certains fibrés, ce qu’on peut faire à l’aide de la formule de Riemann–Roch (à condition de savoir calculer certaines classes de Chern, de savoir démontrer certains théorèmes d’annulation de cohomologie,...). Par exemple, dans [GGr] on montre que si X est une surface de type général alors de tels opérateurs différentiels existent toujours. Mais la condition d’“indépendance” est souvent difficile, sinon impossible, à vérifier : la formule de Riemann–Roch permet (parfois) de montrer qu’un fibré a des sections non triviales, voire *beaucoup* de telles sections, mais ne permet pas de contrôler les lieux d’annulation de ces sections, et c’est justement ces lieux d’annulation qui interviennent dans cette condition d’“indépendance”.

Il y a toutefois des cas où la stratégie qu'on vient de décrire marche assez bien. C'est par exemple le cas du théorème de Bloch (1926) qui affirme que la conjecture 1.1 est vraie si $\dim H^0(X, \Omega_X^1) > \dim X$ et qu'on redémontre dans [GGr] et [Dm1] (entre autres). Ici, grosso modo, ce sont les 1-formes holomorphes sur X qui fournissent les opérateurs différentiels recherchés. Un autre exemple remarquable, sur lequel nous reviendrons, est le théorème de Lu et Yau [LuY] qui prouve la conjecture 1.1 dans le cas des surfaces de type général avec signature strictement positive ($c_1^2 > 2c_2$). Enfin, Siu et Yeung montrent dans [SY1], par ces mêmes techniques, l'hyperbolicité du complémentaire dans le plan projectif d'une courbe générique de degré assez grand.

Voici maintenant le théorème de McQuillan, qui prouve la conjecture de Green–Griffiths dans le cas des surfaces dont la classe de Segre $c_1^2 - c_2$ est strictement positive.

THÉORÈME 1.2. — [MQ1]. *Soit X une surface projective complexe de type général dont les nombres de Chern satisfont l'inégalité $c_1^2(X) > c_2(X)$. Alors aucune courbe entière dans X n'est Zariski-dense.*

C'est-à-dire, une courbe entière dans une telle surface a son image contenue dans une courbe algébrique, courbe qui doit être, évidemment, rationnelle ou elliptique. D'autre part, dans les années '70 Bogomolov avait démontré que, sous les mêmes hypothèses (et nous verrons que ce n'est pas par hasard), X contient un nombre fini de courbes rationnelles ou elliptiques [Bog] (voir aussi [Des] pour un exposé dans ce séminaire). On obtient ainsi (McQuillan + Bogomolov) une confirmation partielle de la conjecture de Lang : une surface X de type général avec $c_1^2(X) > c_2(X)$ contient une courbe algébrique C (dont les composantes irréductibles sont rationnelles ou elliptiques) telle que toute courbe entière dans X est en fait contenue dans C .

Il nous conviendra de découper le théorème 1.2 en deux morceaux :

THÉORÈME 1.3. — [MQ1] (voir aussi [LuY], [GGr], [Dm1]). *Soit X une surface projective complexe de type général avec $c_1^2(X) > c_2(X)$ et soit $f : \mathbf{C} \rightarrow X$ une courbe entière. Il existe alors une surface projective complexe Z , une application holomorphe surjective $\pi : Z \rightarrow X$ et un feuilletage holomorphe (singulier) \mathcal{F} sur Z tels que :*

- i) f se relève sur Z en une courbe entière g (i.e., il existe une courbe entière $g : \mathbf{C} \rightarrow Z$ telle que $\pi \circ g = f$) ;*
- ii) g est tangente à \mathcal{F} (i.e., si ω est une 1-forme holomorphe locale qui définit \mathcal{F} , on a alors $g^*(\omega) \equiv 0$).*

THÉORÈME 1.4. — [MQ1]. *Soit Z une surface projective complexe de type général, soit \mathcal{F} un feuilletage holomorphe sur Z et soit $g : \mathbf{C} \rightarrow Z$ une courbe entière tangente à \mathcal{F} . Alors $g(\mathbf{C})$ n'est pas Zariski-dense dans Z .*

Le théorème 1.2 est corollaire des théorèmes 1.3 et 1.4. Remarquons que la surface Z qui apparaît dans le théorème 1.3 est forcément de type général (car X l'est), même si sa

classe de Segre $c_1^2(Z) - c_2(Z)$ n'est plus nécessairement positive. Heureusement, dans le théorème 1.4 il n'y a aucune hypothèse sur les nombres de Chern de Z .

Le théorème 1.3 n'est pas, à vrai dire, entièrement nouveau. En effet, c'est un cas spécial du résultat de [GGr], [Dm1] et [SY2] rappelé ci-dessus, via des arguments de Bogomolov [Bog] (voir aussi [LuY]). Plus exactement, ces arguments de Bogomolov (qu'on expliquera en détail dans la suite) permettent de montrer l'existence sur X d'opérateurs différentiels algébriques *du premier ordre* (i.e., sections holomorphes globales non triviales de $(Sym^m \Omega_X^1) \otimes A^*$, pour m entier assez grand et A ample), et dire qu'une courbe entière est solution de l'équation différentielle associée à un tel opérateur revient à dire que, modulo un revêtement $Z \rightarrow X$ (de degré m), la courbe entière est tangente à un feuilletage holomorphe. D'ailleurs, c'est en développant ces arguments que Lu et Yau démontrent le théorème 1.2 dans le cas $c_1^2(X) > 2c_2(X)$: en utilisant un résultat de Miyaoka [Miy], ils montrent (grosso modo) que X possède *deux* opérateurs différentiels algébriques du premier ordre et "indépendants", donc la courbe entière est tangente à *deux* feuilletages holomorphes, ce qui entraîne bien sûr qu'elle n'est pas Zariski-dense.

La preuve de McQuillan du théorème 1.3 est toutefois indépendante de [GGr], [Dm1], [SY2] : les arguments de Bogomolov sont combinés avec une *inégalité tautologique* pour les courbes entières, démontrée au début de [MQ1] et qui est, d'une certaine façon, la contrepartie globale des arguments infinitésimaux (lemme d'Ahlfors–Schwarz) de [GGr], [Dm1], [SY2]. Cette inégalité joue un rôle fondamental aussi dans la preuve du théorème 1.4.

À côté de la conjecture de Green–Griffiths, une autre conjecture sur les courbes entières qui a été beaucoup étudiée est celle de Kobayashi, formulée autour de 1970.

CONJECTURE 1.5. — (*Kobayashi*). *i) Pour tout $n \geq 1$ il existe un entier $d(n)$ tel que toute hypersurface générique dans \mathbf{CP}^{n+1} de degré $d \geq d(n)$ est hyperbolique.*

ii) Pour tout $n \geq 1$ il existe un entier $e(n)$ tel que pour toute hypersurface générique D dans \mathbf{CP}^n de degré $e \geq e(n)$ le complémentaire $\mathbf{CP}^n \setminus D$ est hyperbolique.

Ici *générique* signifie qu'il faut exclure, dans l'espace des hypersurfaces de degré donné, celles qui appartiennent à une union dénombrable de sous-ensembles algébriques.

Pour $n = 1$ la conjecture 1.5 est facile, avec $d(1) = 4$ et $e(1) = 3$. Pour $n \geq 2$ le résultat optimal devrait être avec $d(n) = e(n) = 2n + 1$ [Zai]. Moralement, la partie ii) devrait être réductible à la partie i) (avec $e(n) = d(n)$), en considérant dans \mathbf{CP}^{n+1} l'hypersurface donnée comme revêtement cyclique de \mathbf{CP}^n , de degré e et ramifié le long de D , et en relevant sur cette hypersurface les courbes entières dans $\mathbf{CP}^n \setminus D$. Il y a toutefois une lacune dans ce raisonnement, car les hypersurfaces dans \mathbf{CP}^{n+1} obtenues par cette construction ne sont pas génériques par rapport à la totalité des hypersurfaces. Remarquons cependant que ces revêtements cycliques forment un sous-ensemble algébrique dans l'ensemble des hypersurfaces. Pour réparer la lacune, il suffit alors de montrer que ce sous-ensemble

algébrique n'est pas entièrement contenu dans un des sous-ensembles algébriques exclus par la partie i) de la conjecture (voir [DEG] pour un exemple de cette démarche).

Dans le cas $n = 2$, la conjecture de Kobayashi est conséquence de la conjecture de Green–Griffiths, avec $d(2) = e(2) = 5$. En effet, une (hyper)surface lisse X dans $\mathbf{C}P^3$ de degré $d \geq 5$ est de type général, donc (si la conjecture 1.1 est vraie) toute courbe entière dans X est contenue dans une courbe rationnelle ou elliptique. Mais un théorème de Clemens [Cle] (de Xu [XuG] si $d = 5$) affirme que, si de plus X est générique, alors X ne contient *aucune* courbe rationnelle ou elliptique, d'où son hyperbolicité. Un raisonnement similaire s'applique au cas du complémentaire $\mathbf{C}P^2 \setminus D$, via le revêtement cyclique mentionné ci-dessus. Le théorème de Clemens est dans ce cas remplacé par la constatation facile suivante : une courbe générique $D \subset \mathbf{C}P^2$ de degré $e \geq 5$ coupe *toute* courbe elliptique en au moins 1 point et *toute* courbe rationnelle en au moins 3 points.

Cette réduction de la conjecture de Kobayashi à celle de Green–Griffiths semble envisageable aussi dans le cas $n \geq 3$.

On a déjà mentionné le théorème de Siu et Yeung [SY1], qui prouve la conjecture 1.5.ii) pour $n = 2$ (avec $e(2)$ de l'ordre de 10^{13}). Suite à [MQ1], Demailly et El Goul ont prouvé la conjecture 1.5.i), toujours pour $n = 2$, tout en améliorant considérablement le résultat de [SY1]. Le même théorème se trouve aussi dans [MQ2], mais avec des estimations sur le degré un peu moins fines.

THÉORÈME 1.6. — [DEG] (voir aussi [MQ2]). *Pour $n = 2$ la conjecture de Kobayashi 1.5 est vraie, avec $d(2) = e(2) = 21$.*

La partie ii) de la conjecture est analysée dans [DEG] par la construction de revêtements ramifiés le long de D (idem pour [SY1]). On peut toutefois étudier cette partie sans cette construction mais en utilisant des techniques “logarithmiques”, à pôles sur D . Par cette voie, dans un travail plus récent El Goul a prouvé la conjecture 1.5.ii) ($n = 2$) avec $e(2) = 15$ [ElG].

Comme le théorème 1.2, le théorème 1.6 se démontre en deux étapes. D'abord, on montre que sur un revêtement de la surface X toute courbe entière devient tangente à un feuilletage holomorphe. Ensuite on applique le théorème 1.4 et on conclut avec [Cle]. La différence importante avec le théorème 1.2 est que la première partie de cette preuve (la construction du feuilletage) est bien plus compliquée. La raison est qu'une hypersurface X dans $\mathbf{C}P^3$ (de degré au moins 3) ne satisfait *jamais* l'inégalité $c_1^2(X) > c_2(X)$, ainsi les arguments de Bogomolov utilisés dans le théorème 1.3 ne sont plus suffisants. Plus précisément, ces arguments montrent, cette fois, qu'il existe sur X un opérateur différentiel algébrique *du second ordre* (et non plus du premier ordre). Un tel opérateur n'engendre pas un feuilletage sur un revêtement Z de X , mais il engendre plutôt un feuilletage (par courbes) sur une 3-variété W au-dessus de X ; la courbe entière se relève sur W et devient tangente à ce feuilletage. Il faut ensuite des arguments additionnels et assez délicats pour “réduire” la dimension de W de 3 à 2.

2. L'INÉGALITÉ TAUTOLOGIQUE

Soit X une variété complexe compacte et soit $f : \mathbf{C} \rightarrow X$ une courbe entière dans X . On va d'abord rappeler comment on peut associer à f un courant positif fermé $\Phi \in A^{1,1}(X)'$, de bidegré $(1, 1)$. Grosso modo, Φ sera le “courant d'intégration sur $f(\mathbf{C})$ ”, quitte à surmonter le problème de la non-compacité de \mathbf{C} . C'est une construction très classique, qui dans son essence remonte à Ahlfors [Ahl] et qui se résume dans l'affirmation suivante : le bord à l'infini de \mathbf{C} est négligeable.

Fixons une métrique hermitienne (i.e., une $(1, 1)$ -forme positive) ω sur X . Pour tout $r \in \mathbf{R}^+$ définissons le courant positif (non fermé) $\Phi_r \in A^{1,1}(X)'$ par la formule

$$\Phi_r(\eta) = \frac{\int_0^r \frac{dt}{t} \int_{\mathbf{D}(t)} f^*(\eta)}{\int_0^r \frac{dt}{t} \int_{\mathbf{D}(t)} f^*(\omega)} \quad \forall \eta \in A^{1,1}(X)$$

où $\mathbf{D}(t) \subset \mathbf{C}$ est le disque de rayon t (le lecteur pourrait se demander la raison de la double intégration $\int_0^r \frac{dt}{t} \int_{\mathbf{D}(t)}$, à la place de $\int_{\mathbf{D}(r)}$; cette raison est technique, et sera expliquée plus loin).

Un argument de compacité dans l'espace des courants montre qu'on peut choisir une suite divergente r_n , $n \in \mathbf{N}$, telle que Φ_{r_n} converge, pour $n \rightarrow +\infty$, vers un courant positif Φ . Nous voulons toutefois que cette limite soit aussi fermée. En appliquant le théorème de Stokes, on voit facilement que cette dernière propriété est satisfaite dès que la condition suivante est remplie : si $a(t) = \text{aire}(\mathbf{D}(t))$ (par rapport à la (pseudo)métrique $f^*(\omega)$), $l(t) = \text{longueur}(\partial\mathbf{D}(t))$, $A(r) = \int_0^r a(t) \frac{dt}{t}$, $L(r) = \int_0^r l(t) \frac{dt}{t}$, alors

$$\lim_{n \rightarrow +\infty} \frac{L(r_n)}{A(r_n)} = 0.$$

Or, c'est justement le “lemme d'Ahlfors” (ou, plus précisément, une variation sur ce lemme, voir e.g. [Br1] lemme 0) qui garantit l'existence de suites divergentes $\{r_n\}$ remplissant cette condition.

Dorénavant Φ désignera un courant positif fermé associé à f par la construction précédente. Remarquons que Φ ne peut pas être trivial, car $\Phi(\omega) = 1$. Un tel courant n'est pas forcément unique, car il peut dépendre de la suite $\{r_n\}$ (le choix de ω est par contre inessentiel, à constante multiplicative près). On peut toutefois s'arranger pour que certaines propriétés fonctorielles soient satisfaites : par exemple, si $\pi : X_1 \rightarrow X_2$ est un morphisme surjectif entre variétés de même dimension et si f_1 et f_2 sont deux courbes entières dans X_1 et X_2 telles que $f_2 = \pi \circ f_1$ et f_1 n'est pas entièrement contenue dans le lieu critique de π , on peut alors construire Φ_1 et Φ_2 tels que $\pi_*(\Phi_1) = \Phi_2$.

Soit Y une hypersurface complexe compacte dans X , et soit $[Y]$ sa classe de cohomologie dans $H^2(X, \mathbf{R})$. Le courant fermé Φ détermine lui aussi une classe de cohomologie $[\Phi]$, dans $H^{2n-2}(X, \mathbf{R})$ (où $n = \dim X$). On peut donc faire le produit d'intersection $[\Phi] \cdot [Y] \in \mathbf{R}$; concrètement, il s'agit de prendre une 2-forme fermée lisse η qui représente

[Y] et d'évaluer Φ sur η . La propriété suivante sera essentielle dans la suite :

$$\text{si } f(\mathbf{C}) \not\subset Y \quad \text{alors} \quad [\Phi] \cdot [Y] \geq 0.$$

La preuve (voir, e.g., [Br1] lemme 1 ou [Dm2]) est une simple application des formules de Poincaré–Lelong et de Jensen, et c'est exactement ici que la “double intégration” ci-dessus montre ses qualités. Observons que l'hypothèse “ $f(\mathbf{C})$ n'est pas contenue dans Y ” n'exclut pas que le support de Φ , $\text{Supp } \Phi$, soit entièrement contenu dans Y : on a évidemment $\text{Supp } \Phi \subset \overline{f(\mathbf{C})}$, mais cette inclusion peut bien être stricte. Par exemple, il peut arriver que Φ soit le courant d'intégration sur une courbe compacte $C \subset Y$, et dans ce cas l'inégalité ci-dessus dit quelque chose de non trivial (tandis que si $C \not\subset Y$ la même inégalité est triviale).

Après ces préliminaires, nous pouvons maintenant passer à l'*inégalité tautologique* de [MQ1].

Soit PTX le fibré tangent projectivisé de X , $\pi : PTX \rightarrow X$ la projection canonique, et $\mathcal{O}_{PTX}(-1)$ le fibré tautologique sur PTX (qui a degré -1 sur chaque fibre de π). On peut relever la courbe entière f sur PTX , à travers sa dérivée projectivée $f' : \mathbf{C} \rightarrow PTX$. À f' on peut associer un courant positif fermé $\Phi' \in A^{1,1}(PTX)'$, et ensuite on peut considérer le produit d'intersection entre $[\Phi']$ et la classe de Chern de $\mathcal{O}_{PTX}(-1)$, $c_1(\mathcal{O}_{PTX}(-1)) \in H^2(PTX, \mathbf{R})$.

PROPOSITION 2.1. — [MQ1] *theorem A*.

$$c_1(\mathcal{O}_{PTX}(-1)) \cdot [\Phi'] \geq 0.$$

On peut comprendre cette inégalité de la manière suivante. Soit C une courbe compacte lisse dans X , et soit C' son relevé dans PTX . Alors $\mathcal{O}_{PTX}(-1)|_{C'}$ n'est rien d'autre que le fibré tangent de C (par tautologie), et donc $c_1(\mathcal{O}_{PTX}(-1)) \cdot [C'] = \chi(C)$. Le produit qui apparaît dans la proposition 2.1 pourrait alors, à raison, être appelé “caractéristique d'Euler de Φ ”, et la proposition 2.1 dirait que cette caractéristique est non-négative ; ce qui n'est pas inattendu, car Φ est “uniformisé” par \mathbf{C} .

La preuve de McQuillan de la proposition 2.1 est, dans ses grandes lignes, la suivante. La méthode est inspirée des travaux de Faltings et Vojta sur les conjectures (arithmétiques) de Mordell–Lang [Lan] (on a d'ailleurs déjà noté que ces conjectures arithmétiques sont, conjecturellement, liées aux courbes entières).

Soient $Z = X \times X$, $\Delta \subset Z$ la diagonale, W l'éclatement de Z le long de Δ , $Y \subset W$ le diviseur exceptionnel. On a $\Delta \simeq X$, $Y \simeq PTX$, et le fibré $\mathcal{O}_W(Y)$ restreint à Y coïncide avec $\mathcal{O}_{PTX}(-1)$. La courbe f se relève sur Z , en posant $\hat{f} : \mathbf{C} \rightarrow X \times X$, $\hat{f}(t) = (f(t), f(t))$. On ne peut pas relever \hat{f} sur W , car $\hat{f}(\mathbf{C}) \subset \Delta$. Toutefois, on peut déformer \hat{f} de façon telle que son image ne soit plus contenue dans Δ : pour tout $s \in \mathbf{C}$ (proche de 1) on définit $\hat{f}_s : \mathbf{C} \rightarrow X \times X$, $\hat{f}_s(t) = (f(t), f(st))$. Pour $s \neq 1$ (proche de 1) on peut maintenant relever \hat{f}_s sur W , en obtenant ainsi des courbes entières $\tilde{f}_s : \mathbf{C} \rightarrow W$ qui

ne sont pas contenues dans l'hypersurface Y . Comme d'habitude, on associe aux courbes entières \tilde{f}_s des courants positifs fermés $\tilde{\Phi}_s$. Puisque $\tilde{f}_s(\mathbf{C}) \not\subset Y$, on a alors $[\tilde{\Phi}_s] \cdot [Y] \geq 0$, c'est-à-dire $c_1(\mathcal{O}_W(Y)) \cdot [\tilde{\Phi}_s] \geq 0$. Ensuite on passe à la limite pour $s \rightarrow 1$: les courbes $\tilde{f}_s : \mathbf{C} \rightarrow W$ "convergent" vers la dérivée $f' : \mathbf{C} \rightarrow PTX \simeq Y \subset W$. On en déduit que $c_1(\mathcal{O}_W(Y)) \cdot [\Phi'] \geq 0$. Mais $\mathcal{O}_W(Y)|_Y \simeq \mathcal{O}_{PTX}(-1)$, d'où la conclusion de la proposition 2.1.

On trouve dans [Vo1] une autre preuve de la proposition 2.1, ainsi que sa généralisation au cas logarithmique (déguisée dans [MQ1] sous la forme d'inégalité tautologique *raffinée*). Les arguments de Vojta sont plus "classiques", ils sont basés sur des estimations de la théorie de Nevanlinna.

Voici cette généralisation logarithmique. Supposons que X contient une hypersurface D à croisements normaux, telle que $f(\mathbf{C}) \cap D = \emptyset$. Soit $TX(\log D)$ le fibré tangent logarithmique de X à pôles sur D (i.e., le dual de $\Omega_X^1(\log D)$; il est donc engendré localement par les champs de vecteurs holomorphes tangents à D). Soit $PTX(\log D)$ son projectivisé et $\mathcal{O}_{PTX(\log D)}(-1)$ son fibré tautologique. Comme auparavant, on peut encore relever f sur $PTX(\log D)$ et associer à ce relevé f' un courant positif fermé Φ' . On a alors

PROPOSITION 2.2. — [Vo1] *proposition 5.1 (voir aussi [MQ1] theorem II.3.3.2.bis)*.

$$c_1(\mathcal{O}_{PTX(\log D)}(-1)) \cdot [\Phi'] \geq 0.$$

Pour terminer cette section, signalons que dans [MQ1] et [Vo1] on trouve en réalité des inégalités plus précises et générales que celles ci-dessus. Tout d'abord, McQuillan et Vojta ne travaillent pas avec le courant Φ (ou Φ'), mais directement avec la courbe entière f (ou f'). Ils obtiennent donc des estimations sur la croissance de $\int_0^r \frac{dt}{t} \int_{\mathbf{D}(t)} (f')^*(\eta)$, où η est une 2-forme qui représente la classe de Chern du fibré tautologique, dont les propositions 2.1 et 2.2 sont des corollaires. En plus, McQuillan ne se limite pas aux courbes entières, mais il considère aussi des courbes "ramifiées sur \mathbf{C} ", et Vojta ne demande pas $f(\mathbf{C}) \cap D = \emptyset$, mais seulement $f(\mathbf{C}) \not\subset D$. Par conséquent dans leurs estimations apparaissent des termes correctifs, qui sont toutefois identiquement nuls dans les cas que nous avons traités.

3. CONSTRUCTION DE FEUILLETAGES

Dans cette section nous expliquerons d'abord la preuve de McQuillan du théorème 1.3, basée sur des idées de Bogomolov [Bog] [Des] et l'inégalité tautologique de la section précédente. Voir aussi [DEG] section 2 pour une preuve légèrement différente, qui utilise le lemme de Ahlfors–Schwarz de [Dm1] ou [SY2] à la place de l'inégalité tautologique.

Soit X une surface projective complexe lisse de type général. Comme dans la section précédente, soit $\mathcal{O}_{PTX}(-1)$ le fibré tautologique sur PTX (qui est une 3-variété), soit

$\mathcal{O}_{PTX}(1)$ son dual, et soit $\mathcal{O}_{PTX}(m) = \mathcal{O}_{PTX}(1)^{\otimes m}$. Rappelons le fait suivant : pour tout fibré linéaire L sur X , la cohomologie de X à coefficients dans $(Sym^m \Omega_X^1) \otimes L$ est isomorphe à celle de PTX à coefficients dans $\mathcal{O}_{PTX}(m) \otimes \pi^*(L)$.

L'hypothèse du théorème 1.3 sur les nombres de Chern de X est utilisée exclusivement à travers le lemme suivant.

LEMME 3.1. — (Bogomolov). Si $c_1^2(X) > c_2(X)$, alors le fibré $\mathcal{O}_{PTX}(1)$ est gros.

PREUVE

Un calcul standard donne $c_1^3(\mathcal{O}_{PTX}(1)) = c_1^2(X) - c_2(X) > 0$, et ainsi d'après la formule de Riemann–Roch appliquée à $\mathcal{O}_{PTX}(m)$ on a

$$\dim H^0(PTX, \mathcal{O}_{PTX}(m)) + \dim H^2(PTX, \mathcal{O}_{PTX}(m)) > cm^3$$

pour un certain $c > 0$ et pour tout $m \gg 0$. D'autre part, par dualité de Serre et l'isomorphisme $K_X \otimes TX \simeq \Omega_X^1$, on a

$$H^2(PTX, \mathcal{O}_{PTX}(m)) \simeq H^2(X, Sym^m \Omega_X^1) \simeq H^0(X, K_X^{\otimes(1-m)} \otimes Sym^m \Omega_X^1).$$

Mais X est de type général, donc $K_X^{\otimes(m-1)}$ a des sections holomorphes globales non triviales pour $m \gg 0$, et par conséquent

$$\dim H^0(X, K_X^{\otimes(1-m)} \otimes Sym^m \Omega_X^1) \leq \dim H^0(X, Sym^m \Omega_X^1) = \dim H^0(PTX, \mathcal{O}_{PTX}(m))$$

pour $m \gg 0$. Donc $\dim H^0(PTX, \mathcal{O}_{PTX}(m)) \geq \dim H^2(PTX, \mathcal{O}_{PTX}(m))$ pour $m \gg 0$, d'où la conclusion cherchée : $\dim H^0(PTX, \mathcal{O}_{PTX}(m))$ a une croissance cubique en m , ce qui équivaut à dire que $\mathcal{O}_{PTX}(1)$ est gros. □

On pourrait remarquer que la preuve ci-dessus n'a pas réellement besoin que X soit de type général : il suffit que $K_X^{\otimes m}$ soit effectif pour $m \gg 0$, ce qui est bien plus faible que K_X gros. Mais, d'après la classification des surfaces, toute surface X avec $K_X^{\otimes m}$ effectif pour $m \gg 0$ et $c_1^2(X) > c_2(X)$ est en fait de type général.

Fixons sur X un fibré ample A . Puisque $\mathcal{O}_{PTX}(1)$ est gros, le fibré $\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)$ est aussi gros pour $m \gg 0$ (la propriété “gros” est ouverte), et en particulier il possède des sections holomorphes globales non triviales (en fait, pour tout fibré linéaire L , il y a équivalence entre “ L est gros” et “pour tout fibré linéaire B le fibré $L^{\otimes m} \otimes B$ est effectif pour $m \gg 0$ ”). Soit s une telle section et soit $D = (s = 0) \subset PTX$ son diviseur des zéros, ainsi $\mathcal{O}_{PTX}(D)$ est isomorphe à $\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)$.

Soient $f : \mathbf{C} \rightarrow X$ une courbe entière, $f' : \mathbf{C} \rightarrow PTX$ son relevé, $\Phi' \in A^{1,1}(PTX)'$ un courant positif fermé associé à f' . Notons que $\Phi = \pi_*(\Phi')$ est un courant positif fermé non trivial (en fait, c'est un courant associé à f) et donc $c_1(\pi^*(A)) \cdot [\Phi'] = c_1(A) \cdot [\Phi] > 0$ car A est ample. On a alors

$$[\Phi'] \cdot [D] = c_1(\mathcal{O}_{PTX}(m)) \cdot [\Phi'] - c_1(\pi^*(A)) \cdot [\Phi'] < 0$$

car $c_1(\mathcal{O}_{PTX}(m)) \cdot [\Phi'] \leq 0$ d'après la proposition 2.1. De cette inégalité et du fait que D est un diviseur positif on déduit qu'il existe une composante irréductible Z de $\text{Supp } D$ qui contient toute la courbe entière f' :

$$f'(\mathbf{C}) \subset Z.$$

Distinguons alors deux situations possibles :

1) Z est *verticale*, i.e. $\pi(Z)$ est une courbe algébrique C dans X : on a évidemment $f(\mathbf{C}) \subset C$ et le théorème 1.3 (ainsi que le théorème 1.2) est démontré.

2) Z est *horizontale*, i.e. $\pi(Z) = X$. Une telle surface est munie naturellement d'un feuilletage tautologique \mathcal{F} : si $z \in Z$ est un point générique, où Z est transverse à la fibration π , alors un voisinage U de z dans Z induit un feuilletage sur un voisinage V de $x = \pi(z)$ dans X (U est une section de PTX au-dessus de V), et ce feuilletage se relève sur U à travers l'isomorphisme $U \xrightarrow{\pi} V$ (les feuilles dans U sont donc les dérivées des feuilles dans V) ; ensuite, ce feuilletage (non singulier) sur un ouvert de Zariski de Z s'étend à un feuilletage (singulier) sur toute la surface Z . Bien sûr, Z peut être singulière, et dans ce cas il vaut mieux la remplacer par sa désingularisation, qu'on notera encore par Z ; le feuilletage \mathcal{F} et la courbe f' se relèvent, évidemment, sur cette désingularisation. Or, le fait que f' soit à valeurs dans Z et que f' soit la dérivée d'une courbe dans X impliquent, tautologiquement, que f' est *tangente* à \mathcal{F} . Voici donc démontré le théorème 1.3 (avec $g = f'$).

En principe, ces arguments pourraient être considérablement améliorés. Plutôt que prendre *une seule* section s de $\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)$ et son diviseur $D = (s = 0)$, il convient d'introduire le lieu base $B_m \subset PTX$ de ce même fibré, défini comme l'intersection des lieux d'annulation de *toutes* les sections du fibré. Les arguments qui précèdent montrent alors que $f'(\mathbf{C}) \subset B_m$. Si on a la chance de démontrer que $\dim B_m \leq 1$ pour m assez grand, on aura alors une preuve directe du théorème 1.2 qui ne nécessitera pas l'usage des feuilletages. Et si, sous certaines hypothèses, on arrive à montrer que $\dim B_m \leq 0$, on aura alors démontré l'hyperbolicité de X (c'est le cas, par exemple, d'une variété Kählérienne à courbure sectionnelle négative, car $\mathcal{O}_{PTX}(1)$ est alors ample et donc $B_m = \emptyset$ pour m assez grand). Mais, malheureusement, le contrôle de ce lieu base B_m est très difficile : le lemme 3.1 garantit que $\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)$ a beaucoup de sections, mais ne dit rien sur les lieux d'annulation de ces sections.

Il y a toutefois un cas particulier où cette approche est effective, grâce à un théorème de Miyaoka [Miy] [LuY] (voir aussi [MQ1] theorem II.0.2.2 et [DEG] theorem 2.3).

LEMME 3.2. — (Miyaoka). Soit X une surface projective complexe de type général avec $c_1^2(X) > 2c_2(X)$. Soit $Z \subset PTX$ une surface irréductible horizontale. Alors le fibré $\mathcal{O}_{PTX}(1)|_Z$ est gros.

PREUVE

L'argument est proche de celui du lemme 3.1, mais en plus il faut utiliser la semi-stabilité de Ω_X^1 et les inégalités qui en découlent [Bog]. Sans perdre de généralité, on peut supposer X minimale, c'est-à-dire K_X numériquement effectif (rappelons qu'un fibré est *numériquement effectif* s'il a degré non-négatif sur toute courbe).

Calculons $c_1^2(\mathcal{O}_{PTX}(1)|_Z)$. Le fibré $\mathcal{O}_{PTX}(Z)$ est isomorphe à $\mathcal{O}_{PTX}(m) \otimes \pi^*(L)$ pour un certain $m \in \mathbf{N}$ (le degré de $Z \rightarrow X$) et un certain $L \in Pic(X)$. Ce fibré étant effectif, on en déduit que $(Sym^m \Omega_X^1) \otimes L$ a une section globale non triviale, i.e. L^* a un morphisme non trivial vers $Sym^m \Omega_X^1$. La semi-stabilité de Ω_X^1 , et donc de $Sym^m \Omega_X^1$, donne alors l'inégalité

$$c_1(L) \cdot c_1(X) \leq \frac{m}{2} c_1^2(X).$$

On obtient ainsi

$$\begin{aligned} c_1^2(\mathcal{O}_{PTX}(1)|_Z) &= c_1^2(\mathcal{O}_{PTX}(1)) \cdot [c_1(\mathcal{O}_{PTX}(m)) + c_1(\pi^*(L))] = \\ &= m(c_1^2(X) - c_2(X)) - c_1(X) \cdot c_1(L) \geq \frac{m}{2}(c_1^2(X) - 2c_2(X)) > 0. \end{aligned}$$

La formule de Riemann–Roch et la dualité de Serre impliquent que $\mathcal{O}_{PTX}(1)|_Z$ ou bien son dual $\mathcal{O}_{PTX}(-1)|_Z$ est gros. Mais la deuxième possibilité est exclue, car on obtient aussi

$$c_1(\mathcal{O}_{PTX}(-1)|_Z) \cdot c_1(\pi^*(K_X)|_Z) = c_1(X) \cdot c_1(L) - m c_1^2(X) \leq -\frac{m}{2} c_1^2(X) < 0$$

et le produit d'intersection entre un fibré gros et un fibré numériquement effectif, tel que K_X et donc $\pi^*(K_X)|_Z$, ne peut pas être strictement négatif.

□

Ce lemme ne dit pas que le lieu base de $\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)$ est de dimension 1 (au plus), car les sections de ce fibré sur Z ne s'étendent pas nécessairement à toute la variété PTX . Mais les arguments précédents s'appliquent encore sans aucun problème : si $f : \mathbf{C} \rightarrow X$ est telle que $f' : \mathbf{C} \rightarrow PTX$ est à valeurs dans une surface horizontale Z , et si s est une section non triviale de $[\mathcal{O}_{PTX}(m) \otimes \pi^*(A^*)]|_Z$, alors l'image $f'(\mathbf{C})$ est en fait contenue dans la courbe $C = (s = 0)$. En effet, on peut considérer Φ' comme courant sur Z , et évidemment $c_1(\mathcal{O}_{PTX}(m)|_Z) \cdot [\Phi'] = c_1(\mathcal{O}_{PTX}(m)) \cdot [\Phi'] \leq 0$, d'où $[\Phi'] \cdot [C] < 0$. Cela prouve, d'après Lu et Yau [LuY], la conjecture de Green–Griffiths dans le cas $c_1^2 > 2c_2$.

Pour terminer cette discussion autour du théorème 1.3, rappelons deux résultats de Bogomolov et Vojta qui avaient exploité précédemment les mêmes techniques.

Dans [Bog] (voir aussi [Des]) on montre qu'une surface X de type général avec $c_1^2(X) > c_2(X)$ contient un nombre fini de courbes rationnelles ou elliptiques. D'après ce qui précède, ces courbes (qui sont entières...) sont tangentes à un feuilletage, quitte à passer sur un revêtement Z de X . Leur finitude provient alors d'un théorème de Jouanolou [Jou] qui affirme qu'un feuilletage sur une surface possède un nombre fini de feuilles algébriques, sauf dans le cas où le feuilletage est donné par les niveaux d'une fonction rationnelle. Mais

dans ce dernier cas on déduit encore la finitude des courbes rationnelles ou elliptiques sur X , car une surface de type général ne possède aucune fibration rationnelle ou elliptique.

Dans [Vo2] on montre qu'une fibration $X \xrightarrow{F} C$, où X est une surface algébrique, C est une courbe algébrique, la fibre générique de F est de genre ≥ 2 , et F n'est pas isotriviale, possède un nombre fini de sections ("conjecture de Mordell géométrique", prouvée originellement par Manin et ensuite par beaucoup de monde, voir [Lan] pour une discussion approfondie ; le résultat de Vojta fournit des estimations différentes de celles des résultats antérieurs). La preuve consiste à construire un feuilletage (sur un revêtement...) qui contient ces sections comme feuilles, et ensuite on applique à nouveau le théorème de Jouanolou. La construction du feuilletage est similaire à celle de Bogomolov.

Passons maintenant aux constructions de [DEG] et [MQ2], visant à la conjecture de Kobayashi.

Si X est une hypersurface lisse dans $\mathbf{C}P^3$, de degré $d \geq 5$ afin qu'elle soit de type général, le lemme 3.1 ne s'applique pas, car $c_1^2(X) < c_2(X)$ (et, en fait, on peut même montrer que $\mathcal{O}_{PTX}(1)$ n'est jamais gros dans ce cas). On a toutefois l'inégalité $c_1^2(X) > \frac{9}{13}c_2(X)$ dès que $d \geq 15$, ce qui sera utile dans la suite.

Pour obtenir des équations différentielles satisfaites par les courbes entières dans X il faut prendre en considération les dérivées secondes de ces courbes. Soit donc $J_2 \rightarrow X$ le fibré des 2-jets des courbes holomorphes dans X , et soit $E_2 \rightarrow X$ le quotient de J_2 par l'action du groupe des 2-jets de difféomorphismes de $(\mathbf{C}, 0)$ (voir [Dm1] ou [DEG] pour plus de détails sur ces fibrés). Il faut penser à E_2 comme étant le fibré des 2-jets des courbes holomorphes "non paramétrées". On a une $\mathbf{C}P^1$ -fibration $E_2 \xrightarrow{\hat{\pi}} PTX$, induite par 2-jet \mapsto 1-jet, et un fibré tautologique $\mathcal{O}_{E_2}(-1) \in Pic(E_2)$. On a aussi une fibration $\tilde{\pi} = \pi \circ \hat{\pi} : E_2 \rightarrow X$. Remarquons que $\dim E_2 = 4$.

LEMME 3.3. — [Dm1] (voir aussi [DEG] et [MQ2]). Soit X une surface de type général avec $c_1^2(X) > \frac{9}{13}c_2(X)$. Alors le fibré $\mathcal{O}_{E_2}(1)$ est gros.

La preuve imite celle du lemme 3.1 : on calcule $c_1^4(\mathcal{O}_{E_2}(1))$, qui se trouve être positivement proportionnel à $13c_1^2(X) - 9c_2(X)$, ensuite on applique la formule de Riemann–Roch, et la dualité de Serre pour contrôler la cohomologie en degré positif. À noter que la cohomologie de E_2 à coefficients dans $\mathcal{O}_{E_2}(m)$ (y comprise celle en degré 0) s'exprime en termes de celle de X à coefficients dans certains fibrés vectoriels, les fibrés des opérateurs différentiels algébriques du second ordre.

En particulier, sous les hypothèses du lemme 3.3 (satisfaites par les hypersurfaces de degré ≥ 15) le fibré linéaire $\mathcal{O}_{E_2}(m) \otimes \tilde{\pi}^*(A^*)$, où $A \in Pic(X)$ est ample, a des sections holomorphes globales non triviales pour $m \gg 0$. Si $f : \mathbf{C} \rightarrow X$ est une courbe entière et $f'' : \mathbf{C} \rightarrow E_2$ son relevé, on en déduit que l'image $f''(\mathbf{C})$ est contenue dans une hypersurface $W \subset E_2$, composante irréductible du lieu d'annulation d'une section de

$\mathcal{O}_{E_2}(m) \otimes \tilde{\pi}^*(A^*) :$

$$f''(\mathbf{C}) \subset W.$$

Cette dernière affirmation est démontrée dans [DEG] comme conséquence du lemme de Ahlfors–Schwarz de [GGr], [Dm1] et [SY2], et dans [MQ2] à l’aide d’une inégalité tautologique pour les dérivées d’ordre supérieur.

Comme dans le cas $c_1^2 > c_2$, on a donc deux situations possibles. Si W est *verticale*, i.e. $\hat{\pi}(W) \subset PTX$ est une surface Z , on en déduit que $f''(\mathbf{C}) \subset Z$ et donc soit $f(\mathbf{C})$ est dans une courbe $C \subset X$ (si Z est verticale) soit f' est tangente au feuilletage tautologique de Z (si Z est horizontale). Si, par contre, W est *horizontale*, i.e. $\hat{\pi}(W) = PTX$, on en déduit seulement que f'' est tangente au feuilletage tautologique (par courbes) de W . Puisque $\dim W = 3$ et puisqu’on ne connaît pas grand chose sur les feuilletages en dimension ≥ 3 , dans ce deuxième cas il faut des arguments supplémentaires pour se ramener à la dimension 2.

Une démarche naturelle est alors la suivante, inspirée du résultat de [LuY] expliqué ci-dessus : essayer de démontrer que, si $W \subset E_2$ est horizontale, le fibré $\mathcal{O}_{E_2}(1)|_W$ est encore gros. Cela donnera une surface $\hat{Z} \subset W$ telle que $f''(\mathbf{C}) \subset \hat{Z}$, et donc une surface $Z = \hat{\pi}(\hat{Z}) \subset PTX$ telle que $f'(\mathbf{C}) \subset Z$. Si l’on cherche un résultat “général” dans cette direction, avec des hypothèses concernant seulement les nombres de Chern de X , il faudra alors imposer sur $c_1^2(X)$ et $c_2(X)$ une inégalité plus forte que celle qui apparaît dans le lemme 3.3. Dans [DEG] on fait un tel calcul, qui est encore une fois basé sur la semi-stabilité de Ω_X^1 , et on trouve que l’inégalité à imposer est $c_1^2(X) > \frac{9}{7}c_2(X)$. Ce résultat “général” n’est pas très utile à ce point, car nous savons déjà construire des feuilletages sur des surfaces sous l’hypothèse plus faible $c_1^2(X) > c_2(X)$...

Toutefois, les hypersurfaces génériques dans $\mathbf{C}P^3$ ont une propriété remarquable : leur groupe de Picard est isomorphe à \mathbf{Z} (théorème de Noether–Lefschetz). En utilisant cette propriété, Demailly et El Goul parviennent à démontrer que pour avoir $\mathcal{O}_{E_2}(1)|_W$ gros il suffit d’avoir une inégalité du type

$$c_1^2(X) > \frac{9}{13 + 12\theta_2} c_2(X)$$

où $\theta_2 = \theta_2(X)$ est l’infimum des nombres rationnels $\frac{l}{m}$ (avec $m \in \mathbf{N}$, $l \in \mathbf{Z}$) tels que le fibré $\mathcal{O}_{E_2}(m) \otimes \tilde{\pi}^*(K_X^{\otimes l})$ a des sections holomorphes globales non triviales. On a $\theta_2 < 0$ car $\mathcal{O}_{E_2}(1)$ est gros, et $\theta_2 > -\infty$ car K_X est gros. Bien sûr, cette inégalité est plus forte que $c_1^2(X) > \frac{9}{13}c_2(X)$. Mais Demailly et El Goul arrivent aussi à donner des estimations sur θ_2 et, enfin, ils montrent par cette voie que l’inégalité $c_1^2(X) > \frac{9}{13+12\theta_2}c_2(X)$ est satisfaite par les hypersurfaces génériques de $\mathbf{C}P^3$ dès que leur degré est ≥ 21 .

Tous ces calculs sont assez compliqués, et nous ne pouvons que renvoyer à [DEG] pour les détails. Observons toutefois l’aspect suivant. Prouver une estimation du type “ $\theta_2 > \dots$ ” signifie prouver que certains fibrés n’ont aucune section non triviale. Or, si on a une famille (connexe) de surfaces (e.g., la famille des hypersurfaces lisses de $\mathbf{C}P^3$ de

degré fixé) il suffit de faire cela dans le cas d'une seule surface pour en déduire la même propriété pour toutes les surfaces *génériques* de la famille (théorème de semi-continuité). C'est donc ici que le fait de travailler avec des hypersurfaces de $\mathbf{C}P^3$, et pas avec des surfaces abstraites, se révèle très utile : on choisit une hypersurface spéciale dans $\mathbf{C}P^3$ et on fait les calculs sur elle.

En conclusion, donc, Demailly et El Goul montrent que si $X \subset \mathbf{C}P^3$ est générique de degré $d \geq 21$ et si $f : \mathbf{C} \rightarrow X$ est une courbe entière, il existe une surface $Z \rightarrow X$, un feuilletage \mathcal{F} sur Z , un relevé $g : \mathbf{C} \rightarrow Z$ de f sur Z tangent à \mathcal{F} . Le théorème 1.4 et le théorème de Clemens [Cle] conduisent alors à la preuve du théorème 1.6 : X est hyperbolique. Cela se trouve aussi dans [MQ2], suivant le même schéma mais avec des détails qui semblent un peu plus compliqués (d'ailleurs, McQuillan étudie une situation plus générale que celle de la conjecture de Kobayashi).

Terminons cette section avec quelques remarques sur le cas logarithmique.

Soit X une surface projective complexe et soit $D \subset X$ une courbe à croisements normaux. Pour étudier les courbes entières dans $X \setminus D$ il est alors naturel de travailler avec des objets à singularités logarithmiques le long de D . Soit donc $PTX(\log D)$ le fibré tangent logarithmique projectivisé, et soit $\mathcal{O}_{PTX(\log D)}(-1)$ son fibré tautologique.

Comme dans le lemme 3.1, on obtient que si (X, D) est de type log-général (i.e., $K_X \otimes \mathcal{O}_X(D)$ est gros) et si $c_1^2(TX(\log D)) > c_2(TX(\log D))$, alors $\mathcal{O}_{PTX(\log D)}(1)$ est gros. Cela fournit donc un feuilletage sur un revêtement Z de X , et les courbes entières dans $X \setminus D$ se relèvent en feuilles de ce feuilletage. Bien sûr, la proposition 2.1 est ici remplacée par sa version logarithmique 2.2.

Sans surprise, cela ne marche jamais si $X = \mathbf{C}P^2$ et D est une courbe lisse : l'inégalité ci-dessus entre nombres de Chern logarithmiques n'est jamais satisfaite. Il faut à nouveau prendre en compte les dérivées secondes, etc.. On trouve tout cela dans [ElG], où l'auteur prouve la deuxième partie de la conjecture de Kobayashi avec $e(2) = 15$. Bien sûr, le théorème 1.4 est, dans ce contexte, remplacé lui aussi par sa version logarithmique.

4. FEUILLES ENTIÈRES

Dans cette section nous expliquerons les idées qui apparaissent dans la preuve du théorème 1.4.

Soit X une surface projective complexe lisse et \mathcal{F} un feuilletage sur X , à singularités isolées. La donnée d'un tel feuilletage équivaut à la donnée d'une suite exacte

$$0 \rightarrow N_{\mathcal{F}}^* \rightarrow \Omega_X^1 \rightarrow \mathcal{I}_S \cdot K_{\mathcal{F}} \rightarrow 0$$

où $N_{\mathcal{F}}^*$ et $K_{\mathcal{F}}$ sont des fibrés linéaires sur X (le *fibré conormal* et le *fibré canonique* de \mathcal{F}), S est un sous-schéma de X de codimension 2 (le *schéma singulier* de \mathcal{F}), et \mathcal{I}_S est son

idéel. Plus exactement, $N_{\mathcal{F}}^*$ (resp. $T_{\mathcal{F}} = K_{\mathcal{F}}^*$) est le fibré linéaire engendré par les 1-formes holomorphes locales (resp. les champs de vecteurs holomorphes locaux) qui définissent \mathcal{F} . En prenant le déterminant de la suite exacte ci-dessus, on obtient

$$K_X = K_{\mathcal{F}} \otimes N_{\mathcal{F}}^*.$$

On peut associer à \mathcal{F} son *graphe* $F \subset PTX$: c'est une surface irréductible, qui coupe chaque fibre de $\pi : PTX \rightarrow X$ au-dessus d'un point non singulier x en un seul point v_x (= la droite tangente à \mathcal{F} en x) et qui contient chaque fibre de π au-dessus d'un point singulier. En général, F peut avoir des singularités, au-dessus de certains points singuliers de \mathcal{F} .

Remarquons que l'énoncé du théorème 1.4 est invariant par transformations birationnelles : on peut donc faire tous les éclatements qu'on veut. Cela permet, grâce à un théorème classique de Seidenberg [Sei], de supposer que toutes les singularités de \mathcal{F} sont *réduites*. Nous ne voulons pas rentrer ici dans les détails de cette théorie (voir, e.g., [Br2] pour un exposé plus complet), et pour simplifier l'exposition nous allons dans la suite faire les hypothèses suivantes :

- i) \mathcal{F} a une seule singularité, p (ce qui évitera l'introduction de trop d'indices) ;
- ii) p est une singularité *réduite* et *non dégénérée*, i.e. \mathcal{F} est engendré au voisinage de p par un champ de vecteurs holomorphe dont la partie linéaire en p a valeurs propres λ_1, λ_2 avec $\lambda_1 \cdot \lambda_2 \neq 0$ et $\frac{\lambda_1}{\lambda_2} \notin \mathbf{Q}^+$ (ce qui exclura les singularités de type "nœud-col", qui donnent quelque difficulté technique en plus).

Bien sûr, les deux propositions ci-dessus restent vraies dans le cas général des feuilletages à singularités réduites, voir [MQ1] et [Br1].

Le graphe $F \subset PTX$ contient la fibre $E = \pi^{-1}(p)$, et l'hypothèse ii) implique que F est lisse au voisinage de E (et donc partout). En fait, la projection $\pi : F \rightarrow X$ s'identifie à l'éclatement de X en p et $E \subset F$ est son diviseur exceptionnel. La restriction du fibré tautologique $\mathcal{O}_{PTX}(-1)$ à F coïncide avec $\pi^*(T_{\mathcal{F}})$ hors de E (par tautologie) et avec $\mathcal{O}_F(E)$ le long de E (car $\deg(\mathcal{O}_{PTX}(-1)|_E) = -1 = \deg(\mathcal{O}_F(E)|_E)$). C'est-à-dire :

$$\mathcal{O}_{PTX}(-1)|_F = \pi^*(T_{\mathcal{F}}) \otimes \mathcal{O}_F(E).$$

Soit $f : \mathbf{C} \rightarrow X$ une *feuille entière* de \mathcal{F} , i.e. une courbe entière tangente à \mathcal{F} . La dérivée $f' : \mathbf{C} \rightarrow PTX$ est donc à valeurs dans F . Si $\Phi' \in A^{1,1}(PTX)'$ est un courant positif fermé associé à f' (qu'on peut aussi identifier à un courant sur F), l'inégalité tautologique (proposition 2.1) donne alors l'inégalité

$$c_1(\pi^*(T_{\mathcal{F}})) \cdot [\Phi'] + [\Phi'] \cdot [E] \geq 0$$

i.e.

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq -[\Phi'] \cdot [E]$$

où $\Phi = \pi_*(\Phi')$ est un courant positif fermé associé à f . Notons que $[\Phi'] \cdot [E] \geq 0$, car $f'(\mathbf{C}) \not\subset E$. Le résultat suivant de McQuillan est alors une amélioration significative de l'inégalité ci-dessus.

PROPOSITION 4.1. — [MQ1] *theorem II.3.3.2. Si $f(\mathbf{C})$ est Zariski-dense dans X , on a*

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq 0.$$

PREUVE

Supposons d'abord que f ne passe pas par p : $f(\mathbf{C}) \cap \{p\} = \emptyset$ (mais, bien sûr, p peut être dans $\text{Supp } \Phi$, sinon tout est trivial).

Soit alors $X^{(1)} \xrightarrow{\pi^{(1)}} X$ l'éclaté de X en p , $E^{(1)} \subset X^{(1)}$ le diviseur exceptionnel, $f^{(1)} : \mathbf{C} \rightarrow X^{(1)}$ le relevé de f . Le feuilletage \mathcal{F} se relève en un feuilletage $\mathcal{F}^{(1)}$ sur $X^{(1)}$, tangent à $E^{(1)}$. On peut donc considérer son *graphe logarithmique* $F^{(1)} \subset PTX^{(1)}(\log E^{(1)})$. Or, $\mathcal{F}^{(1)}$ a deux singularités $p^{(1)}, q^{(1)}$ sur $E^{(1)}$, qui sont encore réduites et non dégénérées. Mais du point de vue logarithmique ces deux points $p^{(1)}$ et $q^{(1)}$ *ne sont pas* singuliers : si v est un champ de vecteurs holomorphe local qui engendre $\mathcal{F}^{(1)}$ au voisinage de $p^{(1)}$, alors v s'annule en $p^{(1)}$ en tant que section de $TX^{(1)}$, mais il ne s'annule pas en tant que section de $TX^{(1)}(\log E^{(1)})$. Cela signifie que le graphe $F^{(1)}$ est une section de $PTX^{(1)}(\log E^{(1)})$ même au-dessus de $p^{(1)}$ et $q^{(1)}$, et donc partout. On obtient alors la formule tautologique

$$\mathcal{O}_{PTX^{(1)}(\log E^{(1)})}(-1)|_{F^{(1)}} = \pi^*(T_{\mathcal{F}^{(1)}}).$$

Notons, à ce propos, que le “fibré tangent logarithmique” du feuilletage $\mathcal{F}^{(1)}$ coïncide avec son fibré tangent ordinaire $T_{\mathcal{F}^{(1)}}$, puisqu'un champ de vecteurs local qui engendre le feuilletage est forcément logarithmique, i.e. tangent à $E^{(1)}$. Par contre, le “fibré conormal logarithmique” de $\mathcal{F}^{(1)}$ est égal à $N_{\mathcal{F}^{(1)}}^* \otimes \mathcal{O}_{X^{(1)}}(E^{(1)})$ et diffère donc du fibré conormal ordinaire.

La feuille entière $f^{(1)} : \mathbf{C} \rightarrow X^{(1)}$ a son image disjointe de $E^{(1)}$, on peut donc lui appliquer l'inégalité tautologique logarithmique (proposition 2.2), ce qui donne

$$c_1(T_{\mathcal{F}^{(1)}}) \cdot [\Phi^{(1)}] \geq 0$$

où $\Phi^{(1)}$ est un courant positif fermé associé à $f^{(1)}$. On peut supposer que $\Phi^{(1)}$ se projette sur Φ par $\pi^{(1)}$, et d'autre part $T_{\mathcal{F}^{(1)}}$ coïncide avec $(\pi^{(1)})^*(T_{\mathcal{F}})$ car p est réduite. On obtient ainsi l'inégalité cherchée :

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq 0.$$

Le cas général, où f passe par p , nécessite une “suite infinie” d'éclatements. On peut faire à nouveau la construction précédente, mais alors $f^{(1)}(\mathbf{C})$ n'est plus disjointe de $E^{(1)}$ car elle peut contenir $p^{(1)}$ et/ou $q^{(1)}$. Il faut donc éclater $p^{(1)}$ et $q^{(1)}$, en obtenant une surface $X^{(2)}$, une projection $X^{(2)} \xrightarrow{\pi^{(2)}} X$ avec $E^{(2)} = (\pi^{(2)})^{-1}(p) =$ chaîne de 3 courbes rationnelles, un feuilletage $\mathcal{F}^{(2)}$ tangent à $E^{(2)}$ et une feuille entière $f^{(2)} : \mathbf{C} \rightarrow X^{(2)}$. Le feuilletage $\mathcal{F}^{(2)}$ est singulier aux points de croisement de $E^{(2)}$ et, en plus, en deux

points $p^{(2)}$ et $q^{(2)}$, qui se trouvent aux extrémités de $E^{(2)}$ et qui sont susceptibles d'être dans l'image de $f^{(2)}$. On éclate $p^{(2)}$ et $q^{(2)}$, on itère la construction, et pour tout $n \in \mathbf{N}$ on obtient ainsi une surface $X^{(n)} \xrightarrow{\pi^{(n)}} X$, une chaîne de $(2n - 1)$ courbes rationnelles $E^{(n)} = (\pi^{(n)})^{-1}(p)$, un feuilletage $\mathcal{F}^{(n)}$ tangent à $E^{(n)}$, une feuille entière $f^{(n)} : \mathbf{C} \rightarrow X^{(n)}$, et deux singularités $p^{(n)}$ et $q^{(n)}$ aux extrémités de $E^{(n)}$.

Soit alors $D^{(n)} \subset X^{(n)}$ la chaîne de $(2n - 3)$ courbes rationnelles qui s'obtient de $E^{(n)}$ en ôtant les deux courbes rationnelles extrémales. On a $f^{(n)}(\mathbf{C}) \cap D^{(n)} = \emptyset$, et on peut maintenant appliquer l'inégalité tautologique logarithmique à $f^{(n)} : \mathbf{C} \rightarrow X^{(n)} \setminus D^{(n)}$. Du point de vue logarithmique (à pôles sur $D^{(n)}$) $\mathcal{F}^{(n)}$ est singulier *seulement* en $p^{(n)}$ et $q^{(n)}$. On obtient ainsi l'inégalité

$$c_1(T_{\mathcal{F}^{(n)}}) \cdot [\Phi^{(n)}] \geq -[(\Phi^{(n)})'] \cdot ([E_{p^{(n)}}] + [E_{q^{(n)}}])$$

où $E_{p^{(n)}}$ et $E_{q^{(n)}}$ sont les fibres de $PTX^{(n)}(\log D^{(n)})$ au-dessus de $p^{(n)}$ et $q^{(n)}$, fibres contenues dans le graphe logarithmique de $\mathcal{F}^{(n)}$. Puisque $T_{\mathcal{F}^{(n)}} = (\pi^{(n)})^*(T_{\mathcal{F}})$ et $\pi_*^{(n)}(\Phi^{(n)}) = \Phi$, on a

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq -[(\Phi^{(n)})'] \cdot ([E_{p^{(n)}}] + [E_{q^{(n)}}]).$$

D'autre part, un calcul simple (voir [Br1] page 206) montre que

$$[\Phi^{(n)}]^2 = [\Phi]^2 - [\Phi'] \cdot [E] - \sum_{j=1}^{n-1} \{[(\Phi^{(j)})'] \cdot [E_{p^{(j)}}] + [(\Phi^{(j)})'] \cdot [E_{q^{(j)}}]\}.$$

C'est à ce point que l'hypothèse de Zariski-densité de f (et donc de $f^{(n)}$ pour tout n) intervient : cette hypothèse implique que les classes des courants $\Phi^{(n)}$ sont numériquement effectives, et par conséquent $[\Phi^{(n)}]^2 \geq 0$. La somme $\sum_{j=1}^{n-1} \{...\}$ ci-dessus est donc majorée, pour tout n , par $[\Phi]^2$. Cela implique que $[(\Phi^{(n)})'] \cdot [E_{p^{(n)}}]$ et $[(\Phi^{(n)})'] \cdot [E_{q^{(n)}}]$ (qui sont non-négatifs) tendent vers 0 pour $n \rightarrow +\infty$. L'inégalité

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq 0$$

s'obtient alors par passage à la limite. □

À coté de la proposition 4.1, qui estime $c_1(T_{\mathcal{F}}) \cdot [\Phi]$, on a la proposition suivante, qui estime $c_1(N_{\mathcal{F}}) \cdot [\Phi]$. Nous suivrons ici [Br1], qui donne un résultat un peu meilleur que celui de [MQ1] et qui permettra donc d'abrégier la conclusion de la preuve du théorème 1.4.

PROPOSITION 4.2. — [Br1] *théorème 2. Si $f(\mathbf{C})$ est Zariski-dense dans X , on a*

$$c_1(N_{\mathcal{F}}) \cdot [\Phi] \geq 0.$$

PREUVE

On a déjà observé dans la section 2 que, bien que f soit Zariski-dense, le courant Φ peut contenir le courant d'intégration sur une courbe algébrique. Dans la suite, pour

simplifier, on va supposer que Φ ne contient pas de telle composante (voir [Br1] pour le cas général, qui demande simplement des arguments logarithmiques). Cela revient à dire que les nombres de Lelong de Φ hors des singularités de \mathcal{F} sont nuls.

Le produit $c_1(N_{\mathcal{F}}) \cdot [\Phi]$ est égal à $\Phi(\Omega)$, où Ω est une 2-forme fermée qui représente $c_1(N_{\mathcal{F}})$. On va donc construire une telle 2-forme, en suivant les idées de Baum et Bott [BBo]. Remarquons que le fibré $N_{\mathcal{F}}$ est, hors de la singularité p , plat le long des feuilles de \mathcal{F} . Cela permet de construire une 2-forme fermée Ω telle que $[\Omega] = c_1(N_{\mathcal{F}})$ et $\Omega|_{\mathcal{F}} \equiv 0$ hors d'un voisinage U de p . Un calcul standard (voir, e.g., [Br1] ou [Br2] chapter 3) donne aussi la recette suivante pour $\Omega|_{\mathcal{F}}$ dans U . Soit ω une 1-forme holomorphe locale qui engendre \mathcal{F} dans U et soit β une $(1,0)$ -forme C^∞ sur U telle que $d\omega = \beta \wedge \omega$ dans $U \setminus V$, où $V \subset\subset U$ est un autre voisinage de p . Une telle β existe toujours (voir ci-dessous) ; en général, on ne peut pas la choisir holomorphe. On peut alors supposer que

$$\Omega|_{\mathcal{F}} = \frac{1}{2\pi i} d\beta|_{\mathcal{F}}$$

dans U . Remarquons que $d\beta|_{\mathcal{F}} \equiv 0$ dans $U \setminus V$, car $d\omega = \beta \wedge \omega \implies d\beta \wedge \omega \equiv 0$, et donc $\text{Supp}(\Omega|_{\mathcal{F}}) \subset \overline{V} \subset\subset U$. Remarquons aussi que, une fois fixé ω , la restriction $\beta|_{\mathcal{F}}$ est uniquement définie dans $U \setminus V$; et si on remplace ω par $F \cdot \omega$, $F \in \mathcal{O}^*(U)$, la restriction $\beta|_{\mathcal{F}}$ est remplacée par $(\beta + \frac{dF}{F})|_{\mathcal{F}}$. On peut donc dire que la “classe de cohomologie relative” de $\beta|_{\mathcal{F}}$ dans $U \setminus V$ est intrinsèquement définie par \mathcal{F} .

La valeur de Φ sur une 2-forme η dépend seulement de $\eta|_{\mathcal{F}}$, car $\Phi(\eta)$ se calcule en intégrant η sur la feuille $f : \mathbf{C} \rightarrow X$. Puisque $\Omega|_{\mathcal{F}} \equiv 0$ hors de U , on a ainsi localisé le calcul : $\Phi(\Omega)$ est égal à $\Phi_U(\frac{1}{2\pi i} d\beta)$, où Φ_U est la restriction de Φ à U (ou, si l'on préfère, $\Phi_U = \chi_U \cdot \Phi$). On peut toujours supposer que U est une boule autour de p et β est définie au voisinage de \overline{U} . On a alors

$$\Phi(\Omega) = \Phi_U(\frac{1}{2\pi i} d\beta) = \Psi(\frac{1}{2\pi i} \beta)$$

où $\Psi \in A^1(\partial U)'$ est le courant bord de Φ_U (ou, si l'on préfère, $\Psi = d\chi_U \wedge \Phi$). Dorénavant on supposera que $p \in \text{Supp} \Phi$, sinon on a trivialement $\Phi(\Omega) = 0$.

Pour fixer les idées, supposons que \mathcal{F} soit linéarisable au voisinage de p , i.e. engendré par la 1-forme

$$\omega = zdw - \lambda wdz$$

avec $\lambda \neq 0$, $\lambda \notin \mathbf{Q}^+$ (car p est réduite et non dégénérée). Supposons aussi que $\lambda \notin \mathbf{R}^-$ (voir [Br1] pour le cas $\lambda \in \mathbf{R}^-$, qui est en effet un peu plus délicat). Alors, quitte à restreindre U , la 3-sphère ∂U est transverse à \mathcal{F} et donc \mathcal{F} induit sur ∂U un feuilletage non singulier réel \mathcal{L} de dimension 1, orienté par $\mathcal{F}|_U$:

$$\mathcal{L} = \partial(\mathcal{F}|_U).$$

Le courant Φ est associé à une *mesure transverse* à \mathcal{F} , invariante par holonomie [Sul] : la valeur de Φ sur η se calcule en intégrant η le long des feuilles et ensuite en intégrant la

fonction ainsi obtenue (sur l'espace des feuilles) par rapport à la mesure transverse. Cette mesure transverse à \mathcal{F} induit une mesure transverse à \mathcal{L} , invariante par holonomie. Et, réciproquement, cette dernière est associée à un courant $\hat{\Psi}$ sur ∂U , de degré 1 (= dim \mathcal{L}), positif (i.e. $\hat{\Psi}(\gamma) > 0$ si $\gamma|_{\mathcal{L}} > 0$), fermé. Bien sûr, ce courant $\hat{\Psi}$ coïncide avec le bord Ψ de Φ_U introduit ci-dessus : $\hat{\Psi} = \Psi$.

L'hypothèse d'annulation des nombres de Lelong de Φ hors de p se traduit dans l'absence d'atomes pour les mesures transverses en jeu. Cela permet d'exclure le cas $\lambda \notin \mathbf{R}$. En effet, dans ce cas le flot \mathcal{L} a deux orbites périodiques ($\{z = 0\} \cap \partial U$ et $\{w = 0\} \cap \partial U$) et toute autre orbite a ces deux orbites comme ensemble limite, ce qui force toute mesure transverse invariante à être concentrée sur ces deux orbites et donc à être atomique. On a donc $\lambda \in \mathbf{R}^+ \setminus \mathbf{Q}^+$.

Pour ce qui concerne la $(1, 0)$ -forme β , on peut choisir

$$\beta = h \cdot \frac{1 + \lambda}{|z|^2 + |\lambda w|^2} (\bar{z} dz + \bar{\lambda} \bar{w} dw)$$

où $h \in C^\infty(\bar{U})$, $h \equiv 0$ au voisinage de p , $h \equiv 1$ sur $\bar{U} \setminus V$. On a alors

$$\frac{1}{2\pi i} \beta|_{\mathcal{L}} = \frac{(1 + \lambda)}{2\pi i} \frac{dz}{z}|_{\mathcal{L}} > 0$$

et par conséquent $\Psi(\frac{1}{2\pi i} \beta) > 0$.

Le cas $\lambda \in \mathbf{R}^-$ conduit par contre à $\Psi(\frac{1}{2\pi i} \beta) = 0$, voir [Br1]. En tout cas on a donc $\Psi(\frac{1}{2\pi i} \beta) \geq 0$, c'est-à-dire

$$c_1(N_{\mathcal{F}}) \cdot [\Phi] = \Psi(\frac{1}{2\pi i} \beta) \geq 0.$$

□

Grâce aux propositions 4.1 et 4.2, la preuve du théorème 1.4 est maintenant immédiate. En effet, d'après $K_X = T_{\mathcal{F}}^* \otimes N_{\mathcal{F}}^*$ on obtient

$$c_1(K_X) \cdot [\Phi] \leq 0$$

pour tout courant Φ associé à une feuille entière Zariski-dense d'un feuilletage à singularités réduites. Mais la classe $[\Phi]$ est numériquement effective et non triviale, et donc

$$c_1(L) \cdot [\Phi] > 0$$

pour tout fibré gros $L \in \text{Pic}(X)$ (c'est une conséquence du théorème de l'indice de Hodge). Ainsi K_X n'est pas gros, i.e. X n'est pas de type général.

Dans [MQ1] la conclusion de la preuve est un peu plus élaborée, car on dispose seulement d'une version "faible" de la proposition 4.2. Grosso modo, cette version faible permet de conclure qu'on a l'inégalité de la proposition 4.2 si $[\Phi]^2 = 0$. Si par contre $[\Phi]^2 > 0$, McQuillan fait appel à un théorème de Miyaoka [ShB] qui affirme que le fibré canonique d'un feuilletage non-réglé est pseudoeffectif (sa classe dans $\text{Pic}(X) \otimes \mathbf{R}$ est dans la clôture

du cône effectif). Dans notre situation, cela donne en particulier l'inégalité $c_1(T_{\mathcal{F}}) \cdot [\Phi] \leq 0$, car $[\Phi]$ est numériquement effective, et donc

$$c_1(T_{\mathcal{F}}) \cdot [\Phi] = 0.$$

Ensuite, des techniques standard en géométrie des surfaces (théorème de l'indice de Hodge, décomposition de Zariski des diviseurs,...) montrent (grâce à : $[\Phi]$ numériquement effective, $[\Phi]^2 > 0$, $K_{\mathcal{F}}$ pseudoeffectif, $c_1(K_{\mathcal{F}}) \cdot [\Phi] = 0$) que $T_{\mathcal{F}}$ est “presque” trivial, i.e. \mathcal{F} est “presque” engendré par un champ de vecteurs holomorphe global. Cela implique à nouveau que X n'est pas de type général, essentiellement parce qu'une surface de type général ne possède aucun champ de vecteurs holomorphe global.

D'ailleurs, tout cela a été le point de départ d'un travail postérieur de McQuillan [MQ3] (voir aussi [Br2]) dans lequel on classe les feuilletages \mathcal{F} des surfaces projectives tels que $K_{\mathcal{F}}$ n'est pas gros, ce qui en particulier contient la classification des feuilletages ayant une feuille entière Zariski-dense, grâce à la proposition 4.1. À vrai dire, à l'heure actuelle cette classification des feuilletages avec fibré canonique non-gros présente encore une lacune (le cas “dimension de Kodaira $-\infty$ ”), mais cette lacune est remplissable dans le cas spécial des feuilletages avec une feuille entière Zariski-dense (en utilisant la proposition 4.2, ou sa version faible dans [MQ1]).

Il serait évidemment de grand intérêt de généraliser ces résultats, même partiellement, en dimension supérieure. Dans l'étude de la conjecture de Green–Griffiths pour les surfaces X de type général, on rencontre la situation suivante : une n -variété projective Y , une projection surjective $\pi : Y \rightarrow X$, un feuilletage par courbes \mathcal{F} sur Y , et une feuille entière $f : \mathbf{C} \rightarrow Y$. On voudrait démontrer que $\pi \circ f : \mathbf{C} \rightarrow X$ n'est pas Zariski-dense. Si, par contradiction, elle l'était, on aurait alors $c_1(\pi^*(K_X)) \cdot [\Phi] > 0$. D'autre part, $\pi^*(K_X)$ se décompose comme $K_{\mathcal{F}} \otimes L_{\mathcal{F},\pi}^*$, où $L_{\mathcal{F},\pi}$ est, grosso modo, le quotient du fibré normal à \mathcal{F} par le fibré tangent à $\text{Ker } \pi$. Il semble encore possible de démontrer une inégalité du type $c_1(T_{\mathcal{F}}) \cdot [\Phi] \geq 0$, à éclatements près, à partir de l'inégalité tautologique. Par contre, une inégalité du type $c_1(L_{\mathcal{F},\pi}) \cdot [\Phi] \geq 0$ semble plus difficile à prouver. Un obstacle dans tout ça est dû à l'absence d'un théorème de résolution des singularités des feuilletages en dimension ≥ 3 . Notons toutefois que, dans le cas bidimensionnel, la proposition 4.2 reste vraie sans aucune hypothèse sur les singularités de \mathcal{F} , moyennant une faible hypothèse sur les singularités de $[\Phi]$ (voir [Br2] chapter 3).

BIBLIOGRAPHIE

- [Ahl] L. Ahlfors, *Zur Theorie der Überlagerungsflächen*, Acta Math. 65 (1935).
- [BBo] P. Baum, R. Bott, *On the zeroes of meromorphic vector fields*, Essais en l'honneur de De Rham, Springer Verlag (1970), 29-74.

- [Bog] F. Bogomolov, *Families of curves on a surface of general type*, Soviet Math. Dokl. 18 (1977), 1294-1297.
- [Br1] M. Brunella, *Courbes entières et feuilletages holomorphes*, Ens. Math. 45 (1999), 195-216.
- [Br2] M. Brunella, *Birational geometry of foliations*, Congr. UMALCA, notas de curso, IMPA (2000).
- [Cle] H. Clemens, *Curves on generic hypersurfaces*, Ann. Sci. ENS 19 (1986), 629-636.
- [Dm1] J.-P. Demailly, *Algebraic criteria for Kobayashi hyperbolic projective varieties and jet differentials*, Proc. Symp. Pure Math. 62 (1997), 285-360.
- [Dm2] J.-P. Demailly, *Variétés hyperboliques et équations différentielles algébriques*, Gazette des math. 73 (1997), 3-23.
- [DEG] J.-P. Demailly, J. El Goul, *Hyperbolicity of generic surfaces of high degree in projective 3-space*, Amer. J. Math. 122 (2000), 515-546.
- [Des] M. Deschamps, *Courbes de genre géométrique borné sur une surface de type général (d'après F. Bogomolov)*, Sémin. Bourbaki, exp. n° 519 (juin 1978), Lecture Notes in Mathematics 710 (1979), 233-247.
- [ElG] J. El Goul, *Logarithmic jets and hyperbolicity*, article en préparation.
- [GGr] M. Green, Ph. Griffiths, *Two applications of algebraic geometry to entire holomorphic mappings*, The Chern Symposium 1979, Springer Verlag (1980), 41-74.
- [Jou] J.-P. Jouanolou, *Hypersurfaces solutions d'une équation de Pfaff analytique*, Math. Ann. 332 (1978), 239-248.
- [Kob] S. Kobayashi, *Hyperbolic complex spaces*, Springer Verlag (1998).
- [Lan] S. Lang, *Survey of diophantine geometry (Number Theory III)*, EMS vol. 60, Springer Verlag (1991).
- [LuY] S.S.Y. Lu, S.T. Yau, *Holomorphic curves in surfaces of general type*, Proc. Nat. Acad. Sci. USA 87 (1990), 80-82.
- [MQ1] M. McQuillan, *Diophantine approximations and foliations*, Publ. IHES 87 (1998), 121-174.
- [MQ2] M. McQuillan, *Holomorphic curves on hyperplane sections of 3-folds*, Geom. funct. anal. 9 (1999), 370-392.
- [MQ3] M. McQuillan, *Noncommutative Mori theory*, prépubl. IHES M/00/15 (2000).
- [Miy] Y. Miyaoka, *Algebraic surfaces of positive index*, Classification of algebraic and analytic manifolds, Progr. in Math. 39, Birkhäuser (1983), 281-301.
- [Sei] A. Seidenberg, *Reduction of singularities of the differential equation $Ady = Bdx$* , Amer. J. Math. 89 (1968), 248-269.
- [ShB] N.I. Shepherd-Barron, *Miyaoka's theorems on the generic seminegativity of T_X and on the Kodaira dimension of minimal regular threefolds*, Flips and abundance for algebraic threefolds, Astérisque 211 (1992), 103-114 .

- [SY1] Y.T. Siu, S.K. Yeung, *Hyperbolicity of the complement of a generic smooth curve of high degree in the complex projective plane*, Inv. Math. 124 (1996), 573-618.
- [SY2] Y.T. Siu, S.K. Yeung, *Defects for ample divisors of abelian varieties, Schwarz lemma, and hyperbolic hypersurfaces of low degree*, Amer. J. Math. 119 (1997), 1139-1172.
- [Sul] D. Sullivan, *Cycles for the dynamical study of foliated manifolds and complex manifolds*, Inv. Math. 36 (1976), 225-255.
- [Vo1] P. Vojta, *On the ABC conjecture and diophantine approximations by rational points*, Amer. J. Math. 122 (2000), 843-872.
- [Vo2] P. Vojta, *On algebraic points on curves*, Comp. Math. 78 (1991), 29-36.
- [XuG] G. Xu, *Subvarieties of general hypersurfaces in projective space*, J. Diff. Geom. 39 (1994), 139-172.
- [Zai] M. Zaidenberg, *The complement of a generic hypersurface of degree $2n$ in $\mathbf{C}P^n$ is not hyperbolic*, Siberian Math. J. 28 (1987), 425-432.

Marco BRUNELLA

Université de Bourgogne

Laboratoire de Topologie

UMR 5584 du CNRS

9, Avenue Savary - BP 47870

F-21078 DIJON

E-mail : brunella@satie.u-bourgogne.fr

**ESPACES PROBABILISÉS FILTRÉS :
DE LA THÉORIE DE VERSHIK AU MOUVEMENT BROWNIEN,
VIA DES IDÉES DE TSIRELSON**

par Michel ÉMERY

D'infinis tirages ne nécessitent pas, comme les ignorants le supposent, un temps infini; il suffit en réalité que le temps soit infiniment subdivisible.

J. L. BORGES, La loterie à Babylone

La cinématique étudie géométriquement tous les mouvements possibles, dans le cadre d'un espace-temps donné. De même, la théorie des martingales, véritable cinématique des probabilités, s'intéresse à toutes les évolutions possibles, au cours du temps, des probabilités d'événements futurs; et l'on sait, depuis sa fondation par J. L. Doob il y a soixante ans, que le cadre naturel pour cette théorie est un *espace probabilisé filtré*. Il s'agit d'un espace probabilisé muni d'une famille croissante $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathcal{T}}$ de tribus, appelée *filtration*. Le symbole \mathcal{T} désigne un ensemble ordonné qui joue le rôle d'axe des temps; la filtration formalise l'acquisition d'information, les événements de \mathcal{F}_n s'interprétant comme ceux dont, à l'instant n , on sait avec certitude s'ils sont ou non réalisés.

En théorie des martingales à temps discret, on prend habituellement pour axe des temps \mathcal{T} l'ensemble \mathbb{N} des entiers naturels. Une autre situation est le cas, dit des martingales inverses, où l'on prend $\mathcal{T} = -\mathbb{N}$; Doob a montré que le comportement des martingales inverses est plus simple que celui des martingales, mais A. M. Vershik a découvert, il y a une trentaine d'années, que, au contraire du cas où $\mathcal{T} = \mathbb{N}$, la description des filtrations lorsque $\mathcal{T} = -\mathbb{N}$ fait apparaître, au voisinage de $-\infty$, des phénomènes tout à fait inattendus et qui heurtent l'intuition.

Prenons pour fixer les idées une filtration $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{Z}}$ telle que la tribu $\bigcap_n \mathcal{F}_n$ soit grossière (tous ses événements ont probabilité zéro ou un), et telle que, pour chaque n , \mathcal{F}_n soit engendrée par \mathcal{F}_{n-1} et par un événement indépendant de \mathcal{F}_{n-1} et de probabilité $\frac{1}{2}$ (en d'autres termes, aucun événement non trivial n'est connu depuis toujours, et l'univers des événements connus s'enrichit, à chaque instant entier, du résultat d'un coup de pile ou face indépendant du passé). Sous ces hypothèses, on s'attendrait à l'existence d'un jeu de pile ou face joué à chaque instant entier depuis un temps infini, et tel que \mathcal{F}_n décrive les résultats observés jusqu'à l'instant n . Mais Vershik a, entre autres,

- exhibé des contre-exemples (les espaces probabilisés filtrés qu'il a appelés non standard);
- donné un critère nécessaire et suffisant pour reconnaître si un espace probabilisé filtré est standard;
- démontré le *théorème d'isomorphisme lacunaire* : tout espace probabilisé filtré du type ci-dessus mais non standard peut être rendu standard par changement de temps, c'est-à-dire en regroupant en paquets les apports successifs d'information : il existe une sous-suite $\nu : -\mathbb{N} \rightarrow -\mathbb{N}$ strictement croissante, et des variables aléatoires *indépendantes* $(U_n)_{n \in \mathbb{N}}$ telles que, pour chaque n , $\mathcal{F}_{\nu(n)}$ soit engendrée par la famille $\{U_m, m \leq n\}$.

Ces idées de Vershik sont restées inconnues des probabilistes occidentaux jusqu'aux années 90, quand B. Tsirelson et d'autres les ont utilisées pour construire, à la stupéfaction générale, une probabilité équivalente à la mesure de Wiener, mais telle que le nouvel espace probabilisé filtré ne soit pas isomorphe à l'espace de Wiener. Ceci a relancé l'intérêt pour des questions sur les filtrations browniennes, soulevées par M. Yor voici vingt ans; et a aussi poussé Vershik à rédiger la synthèse [24] de ses anciens travaux sur ce thème, dans laquelle nous avons largement puisé.

Tsirelson a ensuite introduit dans [20] un nouvel invariant, le *confort*, qui lui a permis de montrer que l'espace probabilisé filtré associé à un processus de Walsh (sorte de mouvement brownien à valeurs dans une étoile formée d'au moins trois demi-droites issues d'une même origine) n'est pas non plus isomorphe à l'espace de Wiener. Ces idées diffusent maintenant dans les milieux probabilistes, et commencent à donner lieu à une floraison de travaux dont nous parlerons un peu en fin d'exposé.

1. NOTATIONS ET DÉFINITIONS

Un *espace probabilisé filtré* est la donnée d'un ensemble Ω , d'une famille croissante $\mathcal{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}}$ de tribus de parties de Ω , appelée *filtration*, et d'une probabilité P sur la tribu \mathcal{F}_0 (et donc aussi, par restriction, sur chaque \mathcal{F}_n).¹ Nous supposons toujours que \mathcal{F}_0 (et donc aussi chaque \mathcal{F}_n) est essentiellement séparable, c'est-à-dire engendrée, aux événements négligeables près, par une famille dénombrable, ou encore par une v.a. (D'une façon générale, lorsque nous dirons qu'une tribu ou une filtration est engendrée par quelque chose, ce sera toujours modulo les événements négligeables.)

L'ensemble $-\mathbb{N}$ joue le rôle d'axe des temps; sa structure discrète mais non bien ordonnée est la plus simple qui puisse donner lieu aux phénomènes découverts par Vershik. En théorie du mouvement brownien, le temps serait \mathbb{R}_+ au lieu de $-\mathbb{N}$.

Intuitivement, un espace probabilisé filtré décrit l'évolution chronologique des probabilités au fur et à mesure de l'acquisition des connaissances. Cette évolution étant elle-même aléatoire, la probabilité à un instant n d'un événement E est une variable aléatoire, égale à la probabilité conditionnelle² $P[E|\mathcal{F}_n]$; les événements de \mathcal{F}_n sont

1. Les probabilistes définissent traditionnellement P sur une tribu, dite « ambiante », pouvant être plus grosse que \mathcal{F}_0 ; la définition simplifiée nous suffira ici.

2. Techniquement, la probabilité conditionnelle $P[E|\mathcal{F}_n]$ peut être définie comme la densité de Radon-Nikodym de la mesure $F \mapsto P(E \cap F)$ sur \mathcal{F}_n , par rapport à la mesure P restreinte à \mathcal{F}_n .

ceux dont la probabilité à l'instant n ne peut prendre que les valeurs extrêmes 0 et 1, à l'exclusion des valeurs intermédiaires; et la probabilité « brute » P , non aléatoire, représente l'état de nos connaissances — ou de notre ignorance — avant le début de l'évolution.

Si (Ω, \mathcal{F}, P) est un espace probabilisé filtré, l'ensemble \mathcal{F}_0/P , formé des classes d'équivalences d'événements de \mathcal{F}_0 pour l'égalité presque sûre, est une tribu (abstraite) probabilisée; un *isomorphisme* entre deux espaces probabilisés filtrés (Ω, \mathcal{F}, P) et $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ est une bijection Ψ entre \mathcal{F}_0/P et $\bar{\mathcal{F}}_0/\bar{P}$, qui préserve les structures de tribus probabilisées, et telle que pour tous $n \in -\mathbb{N}$ et $E \in \mathcal{F}_0$ on ait, aux négligeables près, $E \in \mathcal{F}_n \Leftrightarrow \Psi(E) \in \bar{\mathcal{F}}_n$. Cet isomorphisme s'étend alors aux espaces $L^0(\Omega, \mathcal{F}_0, P)$ et $L^0(\bar{\Omega}, \bar{\mathcal{F}}_0, \bar{P})$, formés des classes d'équivalences de variables aléatoires pour l'égalité p.s.

Pour simplifier, restreignons-nous à la plus élémentaire des situations auxquelles s'intéresse Vershik, en considérant un espace probabilisé filtré (Ω, \mathcal{F}, P) satisfaisant aux deux hypothèses suivantes³ :

Grossièreté asymptotique : la « tribu initiale » $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ est grossière (elle ne contient que les événements négligeables et les événements presque certains; ou encore, pour tout événement E , $P[E|\mathcal{F}_{-\infty}] = P[E]$);

Dyadicité : il existe pour chaque $n \leq 0$ une v.a. X_n , prenant chacune des deux valeurs -1 et 1 avec probabilité $\frac{1}{2}$, indépendante de \mathcal{F}_{n-1} , et telle que \mathcal{F}_n soit la plus petite tribu contenant \mathcal{F}_{n-1} et les événements $\{X_n = -1\}$ et $\{X_n = 1\}$.

La question que s'est posée Vershik est : Ces deux hypothèses suffisent-elles à caractériser l'espace probabilisé filtré (Ω, \mathcal{F}, P) à isomorphisme près? L'exemple évident vérifiant ces conditions est l'espace naturel du jeu de pile ou face; mais y en a-t-il d'autres?

En remplaçant l'axe des temps $-\mathbb{N}$ par $\{0, 1, \dots, N\}$ ou par \mathbb{N} , on obtiendrait un problème analogue, et beaucoup plus facile, auquel la réponse serait positive, simplement parce que la filtration \mathcal{F} serait alors engendrée par le processus X . Mais lorsque le temps est $-\mathbb{N}$, cet argument est en défaut. Prendre par exemple un jeu de pile ou face $Y = (Y_n)_{n \in \mathbb{Z}}$ (c'est-à-dire des v.a. Y_n indépendantes et valant ± 1 avec probabilité $\frac{1}{2}$), appeler \mathcal{F} la filtration engendrée par Y , et poser $X_n = Y_{n-1}Y_n$. Le processus X est aussi un jeu de pile ou face, les deux hypothèses ci-dessus sont satisfaites par \mathcal{F} et X , et cependant X n'engendre pas \mathcal{F} , puisque la v.a. Y_0 , qui est mesurable pour \mathcal{F}_0 , est indépendante du processus X et ne peut donc être une fonctionnelle des X_n .

Dans cet exemple, bien que X n'engendre pas \mathcal{F} , l'espace probabilisé filtré (Ω, \mathcal{F}, P) engendré par Y est néanmoins isomorphe à celui engendré par X , puisque les processus X et Y ont la même loi (ce sont tous deux des jeux de pile ou face). Voici maintenant un exemple, beaucoup plus surprenant, emprunté à Vershik [24], d'un espace probabilisé filtré (Ω, \mathcal{F}, P) qui satisfait aux deux hypothèses ci-dessus, sans être engendré par aucun jeu de pile ou face.

3. Prises ensemble, elles disent que F possède la propriété de représentation prévisible par rapport à un jeu de pile ou face.

EXEMPLE. — Considérer une suite $(M_n)_{n \in \mathbb{N}_0}$ de v.a. qui suivent chacune la loi uniforme sur $[0, 1]$, et qui se déduisent les unes des autres de la façon suivante : Ayant observé toutes les M_m pour $m \leq n$, on obtient M_{n+1} en jetant une pièce de monnaie et en ne conservant, dans le développement décimal de M_n , que les décimales de rang pair ou impair selon que l'on a obtenu pile ou face. En termes plus savants, M est un processus de Markov stationnaire, de loi uniforme sur $[0, 1]$ et de probabilités de transition

$$0, d_1 d_2 d_3 d_4 d_5 d_6 \dots \begin{cases} \nearrow & 0, d_1 d_3 d_5 \dots \text{ avec probabilité } \frac{1}{2} \\ \searrow & 0, d_2 d_4 d_6 \dots \text{ avec probabilité } \frac{1}{2}. \end{cases}$$

Si l'on se donne $m \leq n \in \mathbb{N}$ et si l'on pose $r = 2^{n-m}$ et $M_m = 0, e_1 e_2 e_3 e_4 \dots$, la loi conditionnelle de M_n à l'instant m consiste en le choix au hasard de l'un parmi les r nombres $0, e_1 e_{r+1} e_{2r+1} \dots$, $0, e_2 e_{r+2} e_{2r+2} \dots$, et ainsi de suite jusqu'à $0, e_r e_{2r} e_{3r} \dots$, chacun avec probabilité $1/r$; puisque ces r nombres sont des v.a. indépendantes et de même loi uniforme, la loi faible des grands nombres entraîne que cette loi conditionnelle est proche de la loi uniforme quand r est grand, c'est-à-dire quand m tend vers $-\infty$, et la grossièreté asymptotique résulte alors de la propriété de Markov. Le caractère dyadique est évident; il est clair également que le jeu de pile ou face qui a servi à construire M n'engendre pas tout l'espace probabilisé filtré : la première décimale de M_0 , par exemple, est indépendante de ce jeu de pile ou face. Mais il est bien plus délicat de montrer qu'*aucun* jeu de pile ou face ne peut engendrer cet espace probabilisé filtré; Vershik le déduit de son « critère de standardité » (et d'estimations de nature combinatoire sur M). Pour pouvoir énoncer ce critère, nous aurons besoin de quelques définitions.

DÉFINITION 1. — On appelle sous-filtration d'une filtration \mathcal{F} une filtration \mathcal{E} telle que $\mathcal{E}_n \subset \mathcal{F}_n$ pour tout instant $n \in \mathbb{N}$.

On dit qu'une filtration \mathcal{E} est immergée dans un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ si \mathcal{E} est une sous-filtration de \mathcal{F} telle que, pour tout événement $E \in \mathcal{E}_0$ et tout temps $n \in \mathbb{N}$, les probabilités conditionnelles $\mathbb{P}[E|\mathcal{E}_n]$ et $\mathbb{P}[E|\mathcal{F}_n]$ sont (p.s.) égales.

DÉFINITION 2. — Soient $(\Omega, \mathcal{E}, \mathbb{P})$ et $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ deux espaces probabilisés filtrés. On dit que $(\Omega, \mathcal{E}, \mathbb{P})$ est immersible dans $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ s'il existe une filtration $\bar{\mathcal{E}}$ immergée dans $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ telle que les espaces probabilisés filtrés $(\Omega, \mathcal{E}, \mathbb{P})$ et $(\bar{\Omega}, \bar{\mathcal{E}}, \bar{\mathbb{P}})$ soient isomorphes.

L'immersibilité formalise rigoureusement l'idée très simple souvent ainsi informellement exprimée : « Quitte à enrichir l'espace, on peut supposer que... ».

DÉFINITION 3. — Un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ est dit standard diffus s'il existe une suite $(U_n)_{n \in \mathbb{N}_0}$ de v.a. indépendantes et de lois diffuses⁴ telle que, pour chaque $n \in \mathbb{N}$, la tribu \mathcal{F}_n soit engendrée par la famille $(U_m)_{m \leq n}$.

Un espace probabilisé filtré est dit standard s'il est immersible dans un espace probabilisé filtré standard diffus.

4. C'est-à-dire $\mathbb{P}[U_n = a] = 0$ pour tout réel a .

On vérifie sans peine que deux espaces probabilisés filtrés standard diffus sont toujours isomorphes (car on peut choisir les U_n de loi uniforme sur $[0, 1]$); il est trivial qu'un espace probabilisé filtré isomorphe à un espace probabilisé filtré standard est lui-même standard, et facile de vérifier qu'un espace probabilisé filtré immersible dans un espace probabilisé filtré standard est lui aussi standard (car l'immersibilité est transitive).

Si $(U_n)_{n \in \mathbb{N}}$ est une suite indépendante de v.a. (de lois non nécessairement diffuses) et si \mathcal{F}_n est la tribu engendrée par les U_m pour $m \leq n$, l'espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ est standard. Pour établir que l'espace probabilisé filtré de l'exemple vu plus haut n'est pas engendré par un jeu de pile ou face, il suffira donc à Vershik de montrer qu'il n'est pas standard; nous verrons plus loin comment.

PROPOSITION 1. — Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé filtré standard, $\nu : -\mathbb{N} \rightarrow -\mathbb{N}$ une application croissante non bornée, et \mathcal{E} la filtration définie par $\mathcal{E}_n = \mathcal{F}_{\nu(n)}$. L'espace probabilisé filtré $(\Omega, \mathcal{E}, \mathbb{P})$ est standard.

DÉMONSTRATION. — Pour n tel que $d = \nu(n) - \nu(n-1) \geq 2$, utiliser une bijection bimesurable et préservant la mesure de Lebesgue entre $[0, 1]$ et $[0, 1]^d$. ■

Le théorème d'isomorphisme lacunaire (théorème 3 plus bas) montrera combien la réciproque à la proposition 1 est loin d'être vraie.

La proposition suivante exprime que le fait d'être standard est une propriété asymptotique au voisinage de $n = -\infty$.

PROPOSITION 2. — Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé filtré et m un instant fixé dans $-\mathbb{N}$; on pose $\mathcal{F}_n^m = \mathcal{F}_{m \wedge n}$. Si l'espace probabilisé filtré $(\Omega, \mathcal{F}_n^m, \mathbb{P})$ est standard, $(\Omega, \mathcal{F}, \mathbb{P})$ l'est aussi.

DÉMONSTRATION. — Sans restreindre la généralité, on peut supposer l'existence d'un vecteur aléatoire (U_{m+1}, \dots, U_0) indépendant de \mathcal{F}_m et de loi uniforme sur $[0, 1]^{|m|}$; il suffit alors de poser $\mathcal{G}_n = \mathcal{F}_n$ pour $n \leq m$ et $\mathcal{G}_n = \sigma(\mathcal{F}_m, U_{m+1}, \dots, U_n)$ pour $m < n \leq 0$, et d'observer que $(\Omega, \mathcal{F}, \mathbb{P})$ est immersible dans $(\Omega, \mathcal{G}, \mathbb{P})$, qui est lui-même standard. ■

REMARQUE. — La proposition 2 reste vraie si l'on y remplace l'instant fixe m par un temps d'arrêt, c'est-à-dire une variable aléatoire N à valeurs dans $-\mathbb{N}$ (la valeur $-\infty$ est interdite!) telle que, pour chaque $n \in -\mathbb{N}$, l'événement $\{N = n\}$ soit dans \mathcal{F}_n .

2. LE CRITÈRE DE VERSHIK

Avant de pouvoir énoncer la condition nécessaire et suffisante donnée par Vershik pour qu'un espace probabilisé filtré soit standard, il nous faut introduire quelques notions.

Si (K, ρ) est un espace métrique compact, l'ensemble K' de toutes les mesures de probabilités sur K est métrisable et compact pour la topologie faible (K' est inclus dans le dual de $C(K)$); cette topologie peut être définie à l'aide de la distance ρ' donnée par

$$\rho'(\mu, \nu) = \inf_{\lambda \text{ de marges } \mu \text{ et } \nu} \int_{K \times K} \rho(k_1, k_2) \lambda(dk_1, dk_2) = \sup_{f \text{ contraction}} \left[\int f d\mu - \int f d\nu \right],$$

où l'infimum porte sur toutes les probabilités λ sur le produit $K \times K$ ayant μ et ν comme marges, et le supremum sur toutes les fonctions $f : K \rightarrow \mathbb{R}$ telles que $|f(k_1) - f(k_2)| \leq \rho(k_1, k_2)$ pour tous k_1 et k_2 dans K .

Si $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé et (K, ρ) un espace métrique compact, nous noterons $L(\mathcal{A}, K)$ l'ensemble de toutes les classes d'équivalence pour \mathbb{P} de v.a. sur (Ω, \mathcal{A}) à valeurs dans K . Si \mathcal{B} est une sous-tribu de \mathcal{A} , la *loi conditionnelle sachant \mathcal{B}* d'une v.a. $R \in L(\mathcal{A}, K)$ est définie comme la v.a. $R' \in L(\mathcal{B}, K')$ telle que, pour chaque borélien G de K , on ait $[R'(\omega)](G) = \mathbb{P}[R \in G | \mathcal{B}](\omega)$ pour presque tout ω ; nous la noterons $\mathcal{L}[R | \mathcal{B}]$.

Si maintenant $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé filtré et (K, ρ) un espace métrique compact, pour tout $R \in L(\mathcal{F}_0, K)$ on peut définir de proche en proche

$$\begin{aligned} \pi_0 R &= R \in L(\mathcal{F}_0, K), \\ \pi_{-1} R &= \mathcal{L}[\pi_0 R | \mathcal{F}_{-1}] \in L(\mathcal{F}_{-1}, K'), \\ \pi_{-2} R &= \mathcal{L}[\pi_{-1} R | \mathcal{F}_{-2}] \in L(\mathcal{F}_{-2}, K''), \quad \text{etc.} \end{aligned}$$

En appelant (K_n, ρ_n) le $|n|$ -ième itéré de l'espace (K, ρ) par la transformation qui en fait (K', ρ') , on a ainsi, pour chaque $n \leq 0$, une variable aléatoire $\pi_n R \in L(\mathcal{F}_n, K_n)$, que l'on peut appeler la *prédiction itérée* de R à l'instant n .

Pour $n < -1$, cette prédiction itérée $\pi_n R$ recèle en général bien plus d'information que la simple loi conditionnelle $\mathcal{L}[R | \mathcal{F}_n]$: cette dernière indique seulement l'état à l'instant n des prévisions concernant la valeur prise par R , alors que $\pi_n R$ décrit aussi les prévisions quant à la façon dont ces prévisions elles-mêmes vont pouvoir évoluer au cours du temps, et ainsi de suite jusqu'au $|n|$ -ième degré. Pour illustrer ceci, prenons par exemple trois v.a. indépendantes (non constantes) N, X_{-1} et X_0 , telles que N prenne les valeurs -1 et 0 , et que X_{-1} et X_0 aient la même loi; supposons que $\mathcal{F}_{-2} = \sigma(N)$, $\mathcal{F}_{-1} = \sigma(N, X_{-1})$ et $\mathcal{F}_0 = \sigma(N, X_{-1}, X_0)$, et considérons la v.a. $R = X_N$. La v.a. $\mathcal{L}[R | \mathcal{F}_{-2}]$, qui est constante, dit seulement que R est indépendante de \mathcal{F}_{-2} et a même loi que X_{-1} et X_0 ; alors que la v.a. $\pi_{-2} R$, plus informative, dit que si $N = -1$, R sera entièrement dévoilée à l'instant -1 , et que si $N = 0$, R sera encore complètement indéterminée à l'instant -1 .

THÉORÈME 1 (Critère de Vershik). — *Pour qu'un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ soit standard, il faut et il suffit que soit satisfait le critère suivant, où (K, ρ) est le compact $[0, 1]$ muni de la distance usuelle : Pour toute v.a. $R \in L(\mathcal{F}_0, K)$, il existe une suite $(\mu_n)_{n \in \mathbb{Z}_0}$ telle que μ_n soit dans K_n pour chaque n , et que $E[\rho_n(\pi_n R, \mu_n)]$ tende vers zéro quand $n \rightarrow -\infty$.*

Une condition nécessaire et suffisante pour que $\mathcal{F}_{-\infty}$ soit grossière est que les lois conditionnelles $\mathcal{L}[R | \mathcal{F}_n]$ deviennent déterministes quand $n \rightarrow -\infty$; le critère de Vershik, condition bien plus forte, demande que les prédictions itérées $\pi_n R$ deviennent déterministes quand $n \rightarrow -\infty$.

REMARQUES. — 1) Le critère peut s'énoncer sans faire intervenir explicitement la suite déterministe des μ_n . En effet, si X est une v.a. à valeurs dans (K, ρ) , appelant ξ la loi de X , ϕ la fonction $k \mapsto E[\rho(X, k)]$ sur K , et dispersion de X le nombre

$$\text{disp } X = E[\phi \circ X] = \iint_{K \times K} \rho(k_1, k_2) \xi(dk_1) \xi(dk_2),$$

on a l'encadrement

$$\inf_{k \in K} E[\rho(X, k)] \leq \text{disp } X \leq 2 \inf_{k \in K} E[\rho(X, k)] ;$$

un énoncé équivalent au critère de Vershik est donc : *pour toute v.a. $R \in L(\mathcal{F}_0, K)$, $\text{disp } \pi_n R \rightarrow 0$ quand $n \rightarrow -\infty$.*

2) Le critère reste vrai si l'on y remplace $[0, 1]$ par n'importe quel espace métrique compact infini, ou encore si l'on y remplace « pour toute R à valeurs dans $[0, 1]$ » par « pour tout (K, ρ) , où K est un ensemble fini muni de la distance $\rho(k_1, k_2) = \mathbb{1}_{\{k_1 \neq k_2\}}$, et toute R à valeurs dans K ». Mais j'ignore si l'on obtient encore une condition équivalente en se restreignant aux K n'ayant que deux points, c'est-à-dire en remplaçant les variables aléatoires R par les événements de \mathcal{F}_0 . (C'est bien équivalent lorsqu'il existe un temps d'arrêt $N > -\infty$ et une v.a. mesurable pour \mathcal{F}_0 , indépendante de \mathcal{F}_N et de loi diffuse; mais le cas dyadique, par exemple, échappe à cette condition.)

La nécessité du critère de Vershik (utilisée pour montrer que l'exemple vu plus haut n'est pas standard) se vérifie sans peine, au moyen des trois arguments suivants :

a) *Stabilité par immersions.* — Si \mathcal{F} est immergée dans $(\Omega, \mathcal{G}, \mathbb{P})$ et si ce dernier satisfait au critère de Vershik, il en va de même de $(\Omega, \mathcal{F}, \mathbb{P})$. Cela résulte immédiatement de ce que, par immersion, les prédictions itérées $\pi_n R$ d'une v.a. R mesurable pour \mathcal{F}_0 sont les mêmes, qu'elles soient calculées dans la filtration \mathcal{F} ou dans la filtration \mathcal{G} . Pour établir que tout espace standard satisfait au critère, il suffit donc de montrer que tout espace engendré par un processus $(U_n)_{n \in \mathbb{Z}}$ fait de v.a. indépendantes y satisfait. On est ainsi ramené au cas où $\mathcal{F}_n = \sigma(U_m, m \leq n)$.

b) *Robustesse de la prédiction.* — Si R et S sont deux v.a. mesurables pour \mathcal{F}_0 et à valeurs dans $K = [0, 1]$, on vérifie facilement, à l'aide de la définition de ρ' , l'inégalité $E[\rho_n(\pi_n R, \pi_n S)] \leq E[\rho(R, S)]$ (plus précisément, mais nous n'en aurons pas besoin, le processus $(\rho_n(\pi_n R, \pi_n S))_{n \in \mathbb{Z}}$ est une sous-martingale). Pour montrer que $\pi_n R$ est proche d'une constante lorsque n est assez petit, cette inégalité permet de se restreindre à des R pris dans une partie dense de $L(\mathcal{F}_0, K)$.

c) *Prédiction d'une fonctionnelle des innovations.* — Comme partie dense de $L(\mathcal{F}_0, K)$, nous prendrons l'ensemble des v.a. de la forme $f(U_{n+1}, \dots, U_0)$ pour un $n \leq 0$ et une fonction borélienne $f : \mathbb{R}^{|n|} \rightarrow K$; la densité est assurée par le classique théorème de Doob sur les martingales directes ($\mathcal{T} = \mathbb{N}$). Lorsque R est de cette forme, il est immédiat, par récurrence descendante, que, pour chaque $m \in \{n, \dots, 0\}$, la prédiction itérée $\pi_m R$ est de la forme $g_m(U_{n+1}, \dots, U_m)$ pour une fonction borélienne $g_m : \mathbb{R}^{m-n} \rightarrow K_m$ (c'est ici qu'intervient l'indépendance des U_ℓ). En particulier, pour $m = n$, on voit que la v.a. $\pi_n R$ est constante; ceci reste vrai a fortiori lorsque l'on remplace n par un instant antérieur. La nécessité du critère est établie.

Il est plus ardu de montrer que le critère est suffisant; nous allons seulement esquisser ci-dessous les grandes lignes de l'argument. Pour plus de détails, voir Vershik [24], ou [11] pour un langage plus probabiliste, ou encore Feldman-Smorodinsky [13] pour une méthode légèrement différente.

Étant données, dans un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, deux sous-tribus \mathcal{B} et \mathcal{C} telles que $\mathcal{C} \subset \mathcal{B}$, rappelons que, pour qu'il existe une v.a. X mesurable pour \mathcal{B} , de loi diffuse, indépendante de \mathcal{C} et telle que $\mathcal{C} \vee \sigma(X) = \mathcal{B}$ (une telle X sera appelée un *complément à \mathcal{C} dans \mathcal{B}*), il faut et il suffit que toute v.a. V vérifiant $\mathcal{C} \vee \sigma(V) = \mathcal{B}$ ait une loi diffuse. Nous dirons en ce cas que \mathcal{B} est *conditionnellement diffuse par rapport à \mathcal{C}* ; et un espace probabilisé filtré tel que chaque \mathcal{F}_n soit conditionnellement diffuse par rapport à \mathcal{F}_{n-1} sera dit *de type diffus*. Pour ce type d'espaces, on a une sorte de réciproque à l'inégalité du b) précédent :

LEMME 1. — *L'espace probabilisé filtré étant de type diffus, soient $R \in L(\mathcal{F}_n, K)$ et $L \in L(\mathcal{F}_{n-1}, K')$ deux v.a. étagées, et telles que les valeurs (dans K') prises par L soient des probabilités (sur K) à supports finis. Il existe une v.a. $S \in L(\mathcal{F}_n, K)$ vérifiant les deux égalités $\mathcal{L}[S|\mathcal{F}_{n-1}] = L$ et $\mathbb{E}[\rho(R, S)] = \mathbb{E}[\rho'(\mathcal{L}[R|\mathcal{F}_{n-1}], L)]$.*

DÉMONSTRATION. — Remarquons d'abord que, par compacité faible de l'ensemble des probabilités sur $K \times K$, l'infimum dans la définition de $\rho'(\mu, \nu)$ est atteint pour un certain λ ; si de plus K est fini, ce λ optimal peut être choisi fonction mesurable du couple (μ, ν) .

Par hypothèse, il existe un ensemble fini $H \subset K$ tel que les valeurs prises par R soient des points de H et que les valeurs (en nombre fini) prises par L soient des probabilités à supports dans H . Choisissons $\mu = \mathcal{L}[R|\mathcal{F}_{n-1}](\omega)$ et $\nu = L(\omega)$, on obtient pour presque tout ω l'existence d'une probabilité $\lambda(\omega)$ sur $H \times H$, dont la dépendance en ω est mesurable pour \mathcal{F}_{n-1} , et qui vérifie

$$\sum_{k \in H} \lambda(\omega, \cdot, k) = \mathcal{L}[R|\mathcal{F}_{n-1}](\omega) \quad \sum_{h \in H} \lambda(\omega, h, \cdot) = L(\omega)$$

et

$$\sum_{(h,k) \in H \times H} \rho(h, k) \lambda(\omega, h, k) = \rho'(\mathcal{L}[R|\mathcal{F}_{n-1}](\omega), L(\omega)).$$

Posons $\Sigma(\omega) = \sum_{k \in H} \lambda(\omega, R(\omega), k) = \sum_{h \in H} \mathbb{1}_{\{R=h\}} \mathbb{P}[R=h|\mathcal{F}_{n-1}]$. De

$$\mathbb{P}[\Sigma=0 \text{ et } R=h|\mathcal{F}_{n-1}] = \mathbb{P}[R=h|\mathcal{F}_{n-1}] \mathbb{1}_{\{\mathbb{P}[R=h|F_{n-1}]=0\}} = 0,$$

on déduit $\mathbb{P}[\Sigma=0] = 0$ et $\Sigma > 0$ p.s. Appelons k_1, \dots, k_p les points de H et pour $0 \leq j \leq p$ posons

$$Q_j(\omega) = \frac{\lambda(\omega, R(\omega), k_1) + \dots + \lambda(\omega, R(\omega), k_j)}{\Sigma(\omega)};$$

ces v.a. sont mesurables pour $\mathcal{F}_{n-1} \vee \sigma(R)$ et vérifient $0 = Q_0 \leq Q_1 \leq \dots \leq Q_p = 1$. Mais, R étant étagée, \mathcal{F}_n est conditionnellement diffuse par rapport à $\mathcal{F}_{n-1} \vee \sigma(R)$; il existe donc un complément X à $\mathcal{F}_{n-1} \vee \sigma(R)$ dans \mathcal{F}_n , de loi uniforme sur $[0, 1]$. Ceci permet de définir une v.a. S , mesurable pour \mathcal{F}_n et à valeurs dans H , par la formule

$$S(\omega) = k_j \iff Q_{j-1}(\omega) < X(\omega) \leq Q_j(\omega),$$

et l'on a

$$\mathbb{P}[S = k|\mathcal{F}_{n-1} \vee \sigma(R)] = \frac{\lambda(\omega, R(\omega), k)}{\Sigma(\omega)};$$

$$\mathbb{P}[R = h \text{ et } S = k|\mathcal{F}_{n-1} \vee \sigma(R)] = \mathbb{1}_{\{R=h\}} \frac{\lambda(\omega, h, k)}{\mathbb{P}[R=h|\mathcal{F}_{n-1}]};$$

$$\mathbb{P}[R = h \text{ et } S = k|\mathcal{F}_{n-1}] = \mathbb{P}[R = h|\mathcal{F}_{n-1}] \frac{\lambda(\omega, h, k)}{\mathbb{P}[R=h|\mathcal{F}_{n-1}]} = \lambda(\omega, h, k)$$

(où tout s'annule quand le dénominateur est nul). En conséquence, la loi conditionnelle de (R, S) par rapport à \mathcal{F}_{n-1} vaut λ . Ceci implique d'une part $\mathcal{L}[S|\mathcal{F}_{n-1}] = L$ et d'autre part

$$\mathbb{E}[\rho(R, S)|\mathcal{F}_{n-1}] = \sum_{(h,k) \in H \times H} \rho(h, k) \lambda(h, k) = \rho'(\mathcal{L}[R|\mathcal{F}_{n-1}], L),$$

d'où $\mathbb{E}[\rho(R, S)] = \mathbb{E}[\rho'(\mathcal{L}[R|\mathcal{F}_{n-1}], L)]$. ■

Posons $K_0^\circ = K$ et, pour tout $n < 0$, appelons K_n° l'ensemble, inclus dans K_n , de toutes les probabilités portées par un nombre fini de points de K_{n+1}° . Il est facile de vérifier que, étant donné un $R \in L(\mathcal{F}_0, K)$, toutes les $|n| + 1$ v.a. $\pi_n R, \pi_{n+1} R, \dots, \pi_0 R$ sont étagées si et seulement si $\pi_n R$ est étagée et à valeurs dans K_n° . Les v.a. R ayant cette propriété seront dites *étagées jusqu'à n* . Avec ces notations, en itérant $|n|$ fois le lemme 1, on obtient sans peine :

LEMME 2. — *Toujours en supposant que l'espace est de type diffus, soient $n \in -\mathbb{N}$, $R \in L(\mathcal{F}_0, K)$ et $L \in L(\mathcal{F}_n, K_n^\circ)$; on suppose en outre L étagée. Il existe une v.a. $S \in L(\mathcal{F}_0, K)$, étagée jusqu'à n , telle que $\pi_n S = L$ et que $\mathbb{E}[\rho(R, S)] = \mathbb{E}[\rho_n(\pi_n R, L)]$.*

Ceci a lieu dans tout espace probabilisé filtré de type diffus. Pour montrer que le critère de Vershik est suffisant, supposons-le satisfait, et cherchons à établir que l'espace est standard. Quitte à immerger $(\Omega, \mathcal{F}, \mathbb{P})$ dans un espace plus gros, nous devons construire un processus engendrant la filtration et formé de v.a. indépendantes. En faisant un produit, c'est-à-dire un grossissement indépendant, par « l' » espace probabilisé filtré standard diffus, et en vérifiant que le critère est stable par de tels produits, on se ramène au cas où $(\Omega, \mathcal{F}, \mathbb{P})$ est de type diffus.

Pour $R \in L(\mathcal{F}_0, K)$ et $\varepsilon > 0$, nous savons par hypothèse qu'il existe un $n \leq 0$ et un $\mu \in K_n$ tels que $\mathbb{E}[\rho_n(\pi_n R, \mu)] < \varepsilon$. Il est immédiat de vérifier que K_n° est dense dans K_n ; on peut donc choisir μ dans K_n° . En prenant la v.a. constante $L = \mu$ dans le lemme 2, on obtient une $S \in L(\mathcal{F}_0, K)$ telle que $\pi_n S$ soit déterministe, que $\pi_{n+1} S, \dots, \pi_0 S$ soient étagées, et que $\mathbb{E}[|S - R|] < \varepsilon$. Puisque chaque $\pi_\ell S$ est étagée, il existe pour tout $\ell \in \{n+1, \dots, 0\}$ un complément X_ℓ à $\mathcal{F}_{\ell-1} \vee \sigma(\pi_\ell S)$ dans \mathcal{F}_ℓ . On peut définir une filtration \mathcal{E} en prenant \mathcal{E}_ℓ grossière pour $\ell \leq n$, et

$$\mathcal{E}_\ell = \sigma(\pi_{n+1} S, X_{n+1}, \pi_{n+2} S, X_{n+2}, \dots, \pi_\ell S, X_\ell) \quad \text{pour } \ell > n;$$

on montre que $\mathcal{F}_n \vee \mathcal{E}_\ell = \mathcal{F}_\ell$ pour tout $\ell \geq n$, que \mathcal{E}_0 et \mathcal{F}_n sont indépendantes, et donc que \mathcal{E} est immergée dans $(\Omega, \mathcal{F}, \mathbb{P})$. Pour $\ell \in \{n+1, \dots, 0\}$ il existe un complément U_ℓ à $\mathcal{E}_{\ell-1}$ dans \mathcal{E}_ℓ ; nécessairement, U_ℓ est aussi un complément à $\mathcal{F}_{\ell-1}$ dans \mathcal{F}_ℓ . Puisque S est mesurable pour \mathcal{E}_0 , elle est fonction borélienne de U_{n+1}, \dots, U_0 ; il est donc établi que les v.a. de la forme $f(U_{n+1}, \dots, U_0)$, où f est borélienne, n dans $-\mathbb{N}$, et chaque U_ℓ un complément à $\mathcal{F}_{\ell-1}$ dans \mathcal{F}_ℓ , sont denses dans L^1 .

Si en outre R est étagée, à valeurs dans une partie finie F de K , en remplaçant f par $g \circ f$, où $g : K \rightarrow F$ est l'identité sur F et est continue en tout point de F , on obtient l'existence d'une S de la forme ci-dessus, et proche de R au sens fort où $\mathbb{P}[S \neq R] < \varepsilon$.

Supposons maintenant donnés $n \in -\mathbb{N}$, $\varepsilon > 0$, des compléments U_{n+1}, \dots, U_0 comme ci-dessus, et une v.a. $R \in L^1(\mathcal{F}_0)$. En raison de la séparabilité essentielle, il existe une suite croissante $(\mathcal{F}_n^k)_{k>0}$ de sous-tribus essentiellement finies de \mathcal{F}_n telle que

$\bigvee_k \mathcal{F}_n^k = \mathcal{F}_n$; par un théorème de Doob, R est donc approchable dans L^1 par des v.a. de la forme $\phi(T, U_{n+1}, \dots, U_0)$ où T est étagée et mesurable pour \mathcal{F}_n . En appliquant la remarque précédente à T au lieu de R et à la filtration décalée $(\mathcal{F}_m)_{m \leq n}$, on obtient l'existence de T' , mesurable pour \mathcal{F}_n , telle que $P[T' \neq T] < \varepsilon$ et qui soit de la forme $T' = \psi(V_{m+1}, \dots, V_n)$, où $m < n$ et où les V_ℓ sont respectivement des compléments à $\mathcal{F}_{\ell-1}$ dans \mathcal{F}_ℓ . Il s'ensuit que R est approchable par $\phi(\psi(V_{m+1}, \dots, V_n), U_{n+1}, \dots, U_0)$. Résumons : U_{n+1}, \dots, U_0 étant donnés, on obtient une partie dense de $L^1(\mathcal{F}_0)$ en considérant les v.a. de L^1 qui sont de la forme $h(V_{m+1}, \dots, V_n, U_{n+1}, \dots, U_0)$, où h est borélienne, $m < n$ et, où V_ℓ est un complément à $\mathcal{F}_{\ell-1}$ dans \mathcal{F}_ℓ pour $m < \ell \leq n$.

Pour terminer la démonstration, considérons une v.a. intégrable R qui engendre la tribu \mathcal{F}_0 . En partant par exemple de $n_0 = -1$ et d'un complément U_0 à \mathcal{F}_{-1} dans \mathcal{F}_0 , l'étape précédente permet de construire de proche en proche une suite strictement décroissante $(n_k)_{k > 0}$ et des compléments U_ℓ à $\mathcal{F}_{\ell-1}$ dans \mathcal{F}_ℓ tels que, pour chaque k , R soit à distance au plus $1/k$ d'un élément de $L^1(\sigma(U_\ell, n_k < \ell \leq 0))$. Il en résulte que R est mesurable pour la tribu limite $\sigma(U_\ell, \ell \in -\mathbb{N})$; et \mathcal{F}_0 , engendrée par R , l'est aussi par les U_ℓ . Enfin, la filtration \mathcal{F} engendrée par les U_ℓ est immergée dans (Ω, \mathcal{F}, P) et vérifie $\mathcal{F}'_0 = \mathcal{F}_0$; ces deux propriétés suffisent à entraîner que $\mathcal{F} = \mathcal{F}$, et (Ω, \mathcal{F}, P) est donc standard. ■

Comme nous l'avons dit plus haut, ce critère est utilisable en pratique, puisqu'il permet de vérifier que l'exemple donné dans la section 1 n'est pas standard. Une autre situation, non fabriquée ad hoc, où le critère fait la preuve de son efficacité, est le problème « T, T^{-1} », résolu par D. Heicklen et C. Hoffman [15] : Si T est un automorphisme d'un espace probabilisé (E, \mathcal{E}, μ) , et si l'entropie de T n'est pas nulle, le processus de Markov à valeurs dans E , de loi stationnaire μ et de probabilités de transition $p(x, \cdot) = \frac{1}{2}(\delta_{Tx} + \delta_{T^{-1}x})$ engendre un espace probabilisé filtré qui n'est pas standard.

Dans la même veine que le critère, voici un autre résultat également emprunté à Vershik [24] :

THEOREME 2. — Soit (Ω, \mathcal{F}, P) un espace probabilisé filtré; pour chaque $n \leq 0$, supposons \mathcal{F}_n engendrée par \mathcal{F}_{n-1} et par une v.a. indépendante de \mathcal{F}_{n-1} , de loi soit diffuse, soit uniforme sur un ensemble fini. Alors (Ω, \mathcal{F}, P) est standard (si et) seulement s'il est engendré par un processus formé de v.a. indépendantes.

Pour la démonstration, nous renvoyons à Vershik [24]. Il ne la donne complètement que dans le cas où aucune des lois n'est diffuse; mais dans le cas général les instants n pour lesquels \mathcal{F}_n admet une v.a. diffuse indépendante de \mathcal{F}_{n-1} se traitent facilement par les mêmes arguments que plus haut. Une variante de sa méthode est proposée par J. Feldman et M. Smorodinsky ([13], théorème 3.2).

Dans le théorème 2, on ne peut se passer de l'hypothèse d'uniformité des lois. Un contre-exemple très simple est fourni par un processus de Markov stationnaire $(X_n)_{n \leq 0}$ ne prenant que deux valeurs, tel que $P[X_n = X_{n-1}] = p$, où $p \in]0, 1[$ est fixé et différent de $\frac{1}{2}$. Par un argument de couplage, le critère de confort-I, dont nous parlerons en fin d'exposé, entraîne facilement que l'espace probabilisé filtré engendré par X est

standard. Mais cet espace ne peut pas être engendré par un processus formé de v.a. indépendantes, car les deux seuls événements de \mathcal{F}_n non dégénérés et indépendants de \mathcal{F}_{n-1} sont $\{X_n = X_{n-1}\}$ et $\{X_n = X_{n-1}\}$ (c'est ici qu'intervient la restriction $p \neq \frac{1}{2}$), et la tribu engendrée par tous les événements de ce type est indépendante de la v.a. X_0 (échanger les deux points de l'espace d'états), donc strictement incluse dans \mathcal{F}_0 .

Au premier abord, la théorie de Vershik est surprenante, et l'existence d'espaces filtrés non standard inattendue. Une fois habitué, et quelques exemples assimilés, on fait face à un deuxième choc, le théorème d'isomorphisme lacunaire, selon lequel un espace probabilisé filtré non standard peut toujours être rendu standard par un changement de temps adéquat.

THEORÈME 3 (d'isomorphisme lacunaire). — *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé filtré tel que la tribu initiale $\mathcal{F}_{-\infty}$ soit grossière. Il existe une application strictement croissante $\nu : -\mathbb{N} \rightarrow -\mathbb{N}$ telle que, si l'on pose $\mathcal{G}_n = \mathcal{F}_{\nu(n)}$, l'espace probabilisé filtré $(\Omega, \mathcal{G}, \mathbb{P})$ soit standard.*

La démonstration de cet étonnant résultat est assez semblable à celle du théorème 1 plus haut ; on ne dispose plus de la puissante hypothèse qu'est le critère, mais l'on sait que toutes les martingales convergent à $-\infty$, et l'on peut à volonté accélérer cette convergence en regroupant l'innovation par paquets, qui correspondront à des intervalles $]\nu(n-1), \nu(n)]$. Pour les détails, voyez Vershik [24], ou [11].

3. CHANGEMENTS NON STANDARD DE PROBABILITÉS

À partir d'ici, nous abandonnons la convention selon laquelle l'axe des temps \mathcal{T} est toujours $-\mathbb{N}$: nous rencontrerons aussi le cas où $\mathcal{T} = \mathbb{R}_+$; nous précisons donc explicitement, lorsque le contexte sera ambigu, l'axe des temps utilisé, en parlant de *temps continu* quand $\mathcal{T} = \mathbb{R}_+$ et de *temps discret* quand $\mathcal{T} = -\mathbb{N}$; les instants seront notés m ou n en temps discret et s ou t en temps continu.

Nous appellerons *espace de Wiener* l'espace probabilisé filtré (à temps continu) engendré par un mouvement brownien.

En 1995, L. Dubins, Feldman, Smorodinsky et Tsirelson [8] et [14] ont utilisé le critère de Vershik pour construire, sur l'espace de Wiener, des probabilités aux propriétés étranges. Leur construction commence avec un jeu de pile ou face infini $(X_n)_{n \in \mathbb{Z}}$ et l'espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ qu'il engendre. Au moyen d'une martingale de densités $(D_n)_{n \in \mathbb{Z}}$ markovienne par blocs (avec des blocs de plus en plus gros au voisinage de $-\infty$), ils parviennent à définir une probabilité \mathbb{Q} équivalente à \mathbb{P} , vérifiant même $1 - \varepsilon < d\mathbb{Q}/d\mathbb{P} = D_0 < 1 + \varepsilon$, et cependant telle que l'espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{Q})$ ne soit pas standard ! Très compliquée, leur construction n'est pas explicite (« tel ensemble ayant une mesure non nulle, on peut y choisir un point... ») ; malgré deux tentatives [18] et [4] pour la simplifier un peu, l'argument reste assez indigeste. Par contre, transférer le résultat du jeu de pile ou face vers le mouvement brownien ne pose pas de difficultés : Étant donné un mouvement brownien $(B_t)_{t \in \mathbb{R}_+}$ pour une probabilité $\bar{\mathbb{P}}$, on choisit dans \mathbb{R}_+ une suite strictement croissante $(t_n)_{n \in \mathbb{Z}}$, et il suffit de définir la densité $d\bar{\mathbb{Q}}/d\bar{\mathbb{P}}$ comme $\phi(\text{sgn}(B_{t_n} - B_{t_{n-1}}), n \leq 0)$, où la fonction ϕ est celle qui définit $d\mathbb{Q}/d\mathbb{P}$ à partir des X_n pour le jeu de pile ou face : $D_0 = \phi(X_n, n \leq 0)$.

Historiquement, cette construction de Dubins, Feldman, Smorodinsky et Tsirelson a été le premier exemple fournissant la réponse (négative) à une question posée par M. Yor [27] en 1979 : Appelons *faiblement brownien* un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ tel que la filtration \mathcal{F} possède la propriété de représentation prévisible par rapport à un mouvement brownien ; ceci signifie que \mathcal{F}_0 est grossière et que toutes les martingales sont des intégrales stochastiques par rapport à ce mouvement brownien ; c'est l'analogie en temps continu d'un espace probabilisé filtré à temps discret dyadique et asymptotiquement grossier. Tout espace probabilisé filtré faiblement brownien est-il engendré par un mouvement brownien ?

En fait, une fois que l'on a à sa disposition la théorie de Vershik, et si l'on ne s'intéresse pas aux changements de probabilités mais seulement à la question de Yor, l'argument de transfert du jeu de pile ou face vers le mouvement brownien fournit très facilement des exemples bien plus simples, de la façon suivante.

On se donne, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, deux objets indépendants : un mouvement brownien $(B_t)_{t>0}$, et une filtration à temps discret $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{Z}}$ telle que $(\Omega, \mathcal{F}, \mathbb{P})$ soit non standard, mais dyadique et asymptotiquement grossier (l'exemple de Vershik donné plus haut, avant la définition des espaces standard, fait parfaitement l'affaire). Il existe des v.a. $(X_n)_{n \in \mathbb{Z}}$ telles que X_n soit indépendante de \mathcal{F}_{n-1} , prenne les valeurs 1 et -1 avec même probabilité $\frac{1}{2}$, et que $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(X_n)$. On se donne aussi, dans \mathbb{R}_+ , une suite strictement croissante $(t_n)_{n \in \mathbb{Z}}$ qui tend vers 0 quand n tend vers $-\infty$; en remarquant que l'ensemble des v.a. $X'_n = X_n \operatorname{sgn}(B_{t_n} - B_{t_{n-1}})$ est indépendant de B , on définit un nouveau mouvement brownien $(B'_t)_{t>0}$ par

$$B'_t = B'_{t_{n-1}} + X'_n(B_t - B_{t_{n-1}})$$

si t est dans l'intervalle $I_n =]t_{n-1}, t_n]$ pour un $n \leq 0$, et par $B'_t = B'_{t_0} + (B_t - B_{t_0})$ si $t \in]t_0, +\infty[$; remarquer que le signe de l'accroissement $B'_{t_n} - B'_{t_{n-1}}$ est X_n . Soient \mathcal{B} et \mathcal{B}' les filtrations (à temps continu) engendrées par B et B' . On peut définir une filtration à temps continu \mathcal{G} par $\mathcal{G}_t = \mathcal{F}_{n(t)} \vee \mathcal{B}'_t$, où $n(t)$ est le plus grand n tel que $t_n \leq t$; observer que l'on a aussi $\mathcal{G}_t = \mathcal{F}_n \vee \mathcal{B}'_t$ pour chaque n tel que $t_n \leq t$. Comme les accroissements de B' dans l'intervalle I_n sont fonction de X_n et des accroissements de B dans I_n , ils sont indépendants de $\mathcal{F}_{n-1} \vee \mathcal{B}_{t_{n-1}}$ et a fortiori de $\mathcal{G}_{t_{n-1}}$. Comme, par définition de \mathcal{G} , on a

$$\mathcal{G}_t = \mathcal{G}_{t_{n-1}} \vee \sigma(B'_s - B'_{t_{n-1}}, s \in]t_{n-1}, t]) \quad \text{pour } t \in]t_{n-1}, t_n],$$

B' est un mouvement brownien pour \mathcal{G} ; comme en outre $\mathcal{G}_{-\infty} \subset \mathcal{F}_{-\infty} \vee \mathcal{B}_0$ est grossière, $(\Omega, \mathcal{G}, \mathbb{P})$ jouit de la propriété de représentation prévisible par rapport à B' , et est donc faiblement brownien. Et cependant \mathcal{G} ne saurait être engendrée par aucun mouvement brownien (ni même aucun processus dont les accroissements sur les I_n sont mutuellement indépendants), puisque la filtration à temps discret \mathcal{F} , qui n'est pas standard, est immergée dans l'espace probabilisé filtré à temps discret $(\Omega, (\mathcal{G}_{t_n})_{n \in \mathbb{Z}}, \mathbb{P})$, qui n'est donc pas standard non plus.

Avant de quitter cette section, signalons encore un autre contre-exemple, qui touche à une question voisine : Partant de l'espace de Wiener $(\Omega, \mathcal{F}, \mathbb{P})$ et effectuant des changements de temps suffisamment réguliers, obtient-on un espace isomorphe à l'espace initial? Inspirée des idées de Dubins, Feldman, Smorodinsky et Tsirelson, mais beaucoup plus simple que leur changement de probabilité, une construction faite dans [9] donne une famille $(T_t)_{t>0}$ de temps d'arrêt de $(\Omega, \mathcal{F}, \mathbb{P})$, vérifiant les propriétés suivantes: chaque fonction $t \mapsto T_t(\omega)$ est C^∞ et de dérivée dT_t/dt vérifiant $1-\varepsilon < dT_t/dt < 1+\varepsilon$; et cependant la filtration changée de temps $(\mathcal{F}_{T_t})_{t>0}$ n'est engendrée par aucun mouvement brownien. Ici encore, la construction s'effectue d'abord en temps discret, et est seulement ensuite transférée au cadre brownien.

4. CONFORT ET POINTS TRIPLES

Revenons à la question de Yor mentionnée dans la section précédente. Depuis 1988, M. Barlow, J. Pitman et Yor [3] soupçonnaient que les processus de Walsh y fournissent un contre-exemple. Ce sont des processus dont la filtration est faiblement brownienne, et ces auteurs conjecturaient qu'elle n'est pas engendrée par un mouvement brownien.

Le processus de Walsh le plus simple est une sorte de mouvement brownien à valeurs dans une (é)toile à trois rayons $T = \{z \in \mathbb{C} : z^3 \in \mathbb{R}_+\}$; sa distance à l'origine est (en loi) la même que pour un mouvement brownien usuel, mais les excursions hors de l'origine, au lieu d'être affectées comme d'habitude de signes ± 1 indépendants, choisissent leur direction indépendamment, et uniformément, parmi les trois rayons de la toile. L'idée de Barlow, Pitman et Yor était que le choix entre trois possibilités (les trois rayons), effectué par le processus aux instants aléatoires où il quitte l'origine, est impossible dans une filtration brownienne. Plus précisément, on sait définir une classe d'instants aléatoires, les *temps honnêtes*,⁵ tels que les débuts d'excursions aient lieu à de tels instants; et si L est un temps honnête, il y a deux tribus, \mathcal{F}_L et \mathcal{F}_{L+} , qui représentent respectivement les événements connus à l'instant L et ceux connus juste après L , et qui vérifient toujours $\mathcal{F}_L \subset \mathcal{F}_{L+}$. Dans le cas où la filtration \mathcal{F} abriterait un processus de Walsh W , en notant L le temps honnête égal au dernier instant avant 2001 où W passe à l'origine, \mathcal{F}_{L+} devrait contenir une v.a. à *trois* valeurs (le rayon choisi pour quitter l'origine, c'est-à-dire celui où se trouve le point W_{2001}) indépendante de \mathcal{F}_L . La stratégie que proposaient Barlow, Pitman et Yor consiste à montrer que si L est n'importe quel temps honnête dans une filtration \mathcal{F} engendrée par un mouvement brownien, \mathcal{F}_{L+} est engendrée par \mathcal{F}_L et par une v.a. ne prenant que *deux* valeurs; cette propriété, appelée « conjecture de Barlow » par Yor [28], rendrait impossible l'existence dans \mathcal{F}_{L+} d'une v.a. trivaluée indépendante de \mathcal{F}_L , et, par là même, d'un processus de Walsh dans \mathcal{F} .

Une bonne partie de ce programme a été menée à bien par Tsirelson en 1996 : en introduisant une idée radicalement nouvelle, le *confort*, il a démontré dans [20] que les processus de Walsh ne peuvent être immergés dans la filtration engendrée par un mouvement brownien, fût-il de dimension plus grande que 1, ou même infinie. Une des nouveautés essentielles de son approche est qu'il ne travaille pas sur *un* processus de Walsh supposé vivre dans *une* filtration brownienne, mais qu'il considère simultanément

5. Ce sont exactement les fins des ensembles optionnels.

deux copies de cette situation, qui soient très proches (condition (iv) de la définition 6 ci-dessous), mais en même temps un tout petit peu indépendantes (condition (iii)).

DÉFINITION 4. — *En temps discret ou continu, on dit que deux filtrations \mathcal{F} et \mathcal{G} , définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, sont co-immersées si elles sont toutes deux immergées dans une même troisième.*

Il est facile de voir qu'en ce cas, \mathcal{F} et \mathcal{G} sont toutes deux immergées dans la plus petite filtration qui les contient toutes deux.

DÉFINITION 5. — *En temps continu, on dit que deux filtrations \mathcal{F} et \mathcal{G} , définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et co-immersées, sont séparées s'il existe une constante $c < 1$ telle que, pour toute martingale M pour \mathcal{F} et toute martingale N pour \mathcal{G} , on ait*

$$[M, N]^2 \leq c [M, M] [N, N] .$$

Sans la constante c , ce serait l'inégalité de Kunita-Watanabe usuelle; la constante vient renforcer l'inégalité, de façon uniforme pour toutes les martingales de \mathcal{F} et \mathcal{G} . Noter que la condition de co-immersion donne un sens à la formule : le calcul de la covariation $[M, N]$ se fait dans une filtration commune, qui est celle dans laquelle \mathcal{F} et \mathcal{G} sont immergées. Nous n'utiliserons cette définition que dans le cas où toutes les martingales de \mathcal{F} et \mathcal{G} sont continues — c'est la situation que considérait Tsirelson lorsqu'il a inventé cette notion —; elle s'interprète alors comme une sorte d'indépendance partielle, l'indépendance étant obtenue pour $c = 0$.

DÉFINITION 6. — *En temps continu, on dit qu'un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$ est confortable si, pour toute v.a. $R \in L^\infty(\Omega, \mathcal{F}_\infty, \mathbb{P})$ et tout $\varepsilon > 0$, il existe un espace probabilisé $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ pourvu de deux filtrations $\bar{\mathcal{F}}$ et $\bar{\mathcal{F}}'$ telles que*

- (i) *les trois espaces probabilisés filtrés $(\Omega, \mathcal{F}, \mathbb{P})$, $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ et $(\bar{\Omega}, \bar{\mathcal{F}}', \bar{\mathbb{P}})$ sont isomorphes;*
- (ii) *les deux filtrations $\bar{\mathcal{F}}$ et $\bar{\mathcal{F}}'$ sont co-immersées;*
- (iii) *les deux filtrations $\bar{\mathcal{F}}$ et $\bar{\mathcal{F}}'$ sont séparées (la constante de séparation c pouvant bien entendu dépendre de R et ε);*
- (iv) *les deux v.a. $R' \in L^\infty(\bar{\mathcal{F}}_\infty)$ et $R'' \in L^\infty(\bar{\mathcal{F}}'_\infty)$, respectivement obtenues à partir de R par les isomorphismes de la condition (i), vérifient $\bar{\mathbb{E}} [|R' - R''|] < \varepsilon$.*

Ceci signifie que l'on peut faire vivre ensemble (ii) deux copies (i) de l'espace, tout à la fois proches (iv) et un peu indépendantes (iii). La démarche de Tsirelson consiste à établir que les espaces probabilisés filtrés engendrés par des mouvements browniens (pouvant être multidimensionnels) sont toujours confortables, alors que l'espace probabilisé filtré engendré par un processus de Walsh n'est jamais confortable. Il en résultera qu'un processus de Walsh ne peut engendrer le même espace qu'un mouvement brownien. Plus précisément, comme tout espace probabilisé filtré immergé dans un espace probabilisé filtré confortable l'est aussi (cela résulte facilement de la définition 6), l'espace engendré par un processus de Walsh n'est immersible dans aucun espace engendré par un mouvement brownien.

La première étape de Tsirelson est donc celle-ci :

THÉORÈME 4. — *Un espace probabilisé filtré engendré par un mouvement brownien est toujours confortable.*

DÉMONSTRATION. — Prendre deux copies indépendantes B^0 et $B^{\pi/2}$ du brownien générateur, et considérer l'espace probabilisé filtré $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ engendré par le couple $(B^0, B^{\pi/2})$. Pour tout réel α , le processus $B^\alpha = B^0 \cos \alpha + B^{\pi/2} \sin \alpha$ est un mouvement brownien dans $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$; la filtration \mathcal{F}^α qu'il engendre est donc immergée dans cet espace. Si $\alpha - \beta$ n'est pas multiple de π , \mathcal{F}^α et \mathcal{F}^β sont séparées, avec constante $c = \cos^2(\alpha - \beta) < 1$, car l'inégalité de Kunita-Watanabe renforcée, vraie pour les mouvements browniens B^α et B^β , s'étend aux autres martingales par représentation prévisible.

Pour conclure, il ne reste qu'à établir que, étant donnée une fonctionnelle bornée R^0 de B^0 , la copie R^α de R^0 dans la filtration \mathcal{F}^α vérifie $\bar{\mathbb{E}}[|R^\alpha - R^0|] < \varepsilon$ pour un α assez voisin de 0 (mais non nul, pour que \mathcal{F}^α et \mathcal{F}^0 soient séparées). Cela résulte facilement du joli théorème suivant, qui semble classique chez les probabilistes russes sous le nom de « lemme de Slutsky » :

Si une suite $(Z_p)_{p \in \mathbb{N}}$ de v.a. ayant même loi converge en probabilité vers une limite Z_∞ , et si h est une fonction borélienne, la suite $(h \circ Z_p)_{p \in \mathbb{N}}$ converge en probabilité vers $h \circ Z_\infty$. En effet, si l'on n'exige pas que les Z_p aient la même loi, c'est vrai pour toute h continue; et lorsque les Z_p ont même loi, l'ensemble des h pour lesquelles c'est vrai est stable par limites simples de suites. ■

La deuxième étape est tout aussi simple à énoncer que la première; mais sa réalisation demandera bien plus d'efforts :

THÉORÈME 5. — *Si, dans un espace probabilisé filtré $(\Omega, \mathcal{F}, \mathbb{P})$, il existe un processus de Walsh (pour la filtration \mathcal{F}), cet espace n'est pas confortable.*

ÉLÉMENTS DE DÉMONSTRATION. — Considérant deux copies de \mathcal{F} co-immérgées et séparées, on a deux processus de Walsh W' et W'' dans une même filtration. Tsirelson montre tout d'abord que les instants où W' et W'' sont tous deux à l'origine sont rares : mesuré en termes du temps local de l'un, le temps passé par l'autre à l'origine est nul. Ceci est établi par une méthode de comparaison de solutions d'inéquations aux dérivées partielles; pour une autre démonstration (par le calcul stochastique), voir [12]. C'est ici qu'est utilisée l'hypothèse de séparation.

Tsirelson s'intéresse ensuite à la distance $D_t = d(W'_t, W''_t)$, mesurée sur la toile T . Le processus D est une sous-martingale, dont le compensateur ne croît que lorsque W' et W'' sont égaux, ou que l'un au moins d'entre eux est à l'origine. Le résultat du paragraphe précédent permet de minorer ce compensateur par une somme de termes de temps locaux de W' et W'' à l'origine, qui se calculent séparément; on obtient ainsi $\bar{\mathbb{E}}[D_t] \geq \frac{1}{3} \sqrt{2t/\pi}$ (si l'on remplaçait T par une toile à r rayons au lieu de 3, le coefficient $\frac{1}{3}$ deviendrait $\frac{r-2}{r}$, et le théorème subsisterait pourvu que $r > 2$).

Enfin, la minoration ci-dessus, universelle et indépendante en particulier de la constante de séparation c , montre que les v.a. complexes W'_t et W''_t ne sont pas proches, et empêche que ne soit satisfaite la condition (iv) de la définition 6 : l'espace ne peut être confortable. ■

En appliquant les mêmes méthodes à des processus à valeurs dans T un peu plus généraux (les « semimartingales-araignées », qui sont les semimartingales continues, de compensateur porté par l'ensemble des instants où le processus est à l'origine), Tsirelson établit, sous l'hypothèse de confort, la formule $dL_1 \wedge dL_2 \wedge dL_3 = 0$, où L_i est le temps local à l'origine du processus projeté sur le i -ième rayon. Soient maintenant U_1, U_2 et U_3 trois ouverts bornés de \mathbb{R}^d , deux-à-deux disjoints. En choisissant comme semimartingale-araignée le processus $(G_1 \circ B, G_2 \circ B, G_3 \circ B)$, où G_i est une fonction de Green dans U_i et B un mouvement brownien dans \mathbb{R}^d , il obtient $\mu_1 \wedge \mu_2 \wedge \mu_3 = 0$, où μ_i est « la » mesure harmonique sur la frontière de U_i (seule la classe d'équivalence de μ_i est intrinsèquement définie). Ce résultat selon lequel, en dimension d quelconque, les frontières harmoniques ont fort peu de points triples, avait été conjecturé par C. Bishop [5] en 1991 (en dimension $d = 3$ on savait seulement que la multiplicité est ≤ 4 , et, en grandes dimensions, ≤ 10). En donner une démonstration non probabiliste est, selon Tsirelson [21], « a challenge for classical analysis! ».

En utilisant les outils introduits par Tsirelson, mais avec une définition du confort légèrement modifiée, la conjecture de Barlow a été complètement établie dans [2] (voir aussi [6] pour une simplification). L'idée est ici, partant d'un temps honnête L tel que \mathcal{F}_{L+} ne soit pas engendrée par \mathcal{F}_L et par une v.a. à deux valeurs, de construire une martingale à valeurs dans T (une « martingale-araignée ») quittant l'origine à l'instant L , et de lui faire ensuite subir les traitements infligés par Tsirelson aux processus de Walsh, pour montrer que l'espace ne peut être confortable.

Le confort, ou plutôt son absence, a également été utilisé par S. Watanabe [26] pour montrer qu'une certaine diffusion dans le plan, symétrique pour la mesure de Lebesgue et de probabilités de transition continues, introduite par Ikeda et lui-même en 1971, ne peut pas non plus s'immerger dans un espace de Wiener. Ce processus se déplace de façon brownienne dans le plan; lorsqu'il rencontre les rayons de la toile T , il s'agit aussi de façon brownienne le long du rayon, à une vitesse beaucoup plus rapide (réglée par le temps local de la composante normale au rayon); lorsqu'il est à l'origine, ce qui se produit de façon récurrente, son comportement est régi par une loi d'excursions explicite. Watanabe montre qu'aux instants où le processus quitte l'origine, la conjecture de Barlow est violée; d'où le résultat.

Dans le même article, il en déduit que le « mouvement brownien gluant » (mouvement brownien réel ralenti lorsqu'il passe en 0 de façon à faire coïncider son temps local avec le temps ordinaire passé en 0) n'est pas confortable non plus. L'argument consiste simplement à remarquer qu'à changement de temps près, deux mouvements browniens gluants indépendants forment une diffusion du type vu précédemment, la toile (à quatre rayons) étant formée des deux axes.

Ce dernier résultat est aussi obtenu, indépendamment, par Warren [25], en étudiant directement deux copies co-immérgées de ce processus. (Mais ma fille soutient que sans étudier les mathématiques, tout le monde sait que rien de gluant ne peut être confortable.)

Nous avons vu que, pour prouver la conjecture de Barlow, la définition du confort a dû être modifiée; voici, pour terminer, une autre variation sur cette notion, qui fournit un lien entre critère de Vershik et confort. Revenons en temps discret, les filtrations étant indexées par $-\mathbb{N}$, et reprenons la définition 6, en y remplaçant bien sûr \mathcal{F}_∞ par \mathcal{F}_0 , mais aussi la condition (iii) par

(iii') *il existe un $n \in -\mathbb{N}$ tel que les tribus \mathcal{F}_n et \mathcal{F}'_n soient indépendantes sous $\bar{\mathbb{P}}$.*

La propriété obtenue est appelée *confort-I* (le I signifie indépendance); elle a été utilisée, sans être explicitement définie, par Smorodinsky [19], comme condition nécessaire pour que l'espace probabilisé filtré soit standard; et cette condition nécessaire se trouve être aussi suffisante [11]. Elle fournit ainsi un substitut au critère de Vershik (qui est d'ailleurs mis à contribution pour montrer qu'elle est suffisante).

Un intérêt du confort-I est d'être parfois plus simple à mettre en œuvre que le critère de Vershik, parce que l'on n'a pas à calculer dans un espace métrique dépendant de n ; l'exemple donné juste après le théorème 2 en est une illustration. Mais le confort-I donne aussi un nouvel éclairage au concept d'espace standard, par le lien qu'il établit avec la notion de *couplage*. Le couplage consiste à faire évoluer simultanément deux versions d'un même processus, de conditions initiales différentes, en cherchant à maximiser leur probabilité de rencontre; c'est une méthode familière aux probabilistes pour obtenir des estimations en loi (voir le traité de Lindvall [16] ou l'exposé de Diaconis [7]). L'analogie avec le confort-I saute aux yeux : si la filtration \mathcal{F} est engendrée par un processus X , la condition (iii') dit que jusqu'à un instant passé n on disposait de deux copies indépendantes X' et X'' de X , la condition (ii) dit qu'après n les deux processus évoluent conjointement, et (iv), qu'à la fin, les deux copies donnent à une certaine fonctionnelle de X des valeurs proches. Pour un exemple éclairant sur ce que le couplage peut apporter à l'étude des filtrations, voir la démonstration par Arnaudon [1] du fait que le mouvement brownien sur une variété compacte à d dimensions en temps éternel (l'axe des temps est \mathbb{R}) a une filtration brownienne à d dimensions, ou [10] pour le cas, plus simple, où la variété est une sphère.

Une question stimulante pour les travaux à venir me paraît être la recherche d'un analogue, en temps continu, de la théorie de Vershik : comprendre ce qui fait qu'un espace probabilisé filtré est, ou n'est pas, isomorphe à l'espace de Wiener (qui devrait être un analogue, en temps continu, de l'espace probabilisé filtré standard dyadique).

Les progrès actuels à cet égard sont de deux types : d'une part, de plus en plus de bruits stochastiques continus, autres que les bruits blancs, sont définis et explorés par Tsirelson [22] et [23]. D'autre part, les démonstrations du fait qu'une filtration donnée est engendrée par un mouvement brownien sont de moins en moins explicitement constructives (voyez M. Malric [17]).

Mais, pour impressionnantes que soient les avancées de ces dernières années, bien du chemin reste à parcourir. La question posée par Yor il y a vingt ans reste d'actualité, une fois précisée et reformulée à la lumière des connaissances actuelles :

Si un espace probabilisé filtré faiblement brownien est immersible dans l'espace de Wiener, est-il nécessairement isomorphe à l'espace de Wiener (c'est-à-dire engendré par un mouvement brownien) ?

RÉFÉRENCES

- [1] M. ARNAUDON – *La filtration du mouvement brownien indexé par \mathbb{R} dans une variété compacte*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 304–314, 1999.
- [2] M. BARLOW, M. ÉMERY, F. KNIGHT, S. SONG & M. YOR – *Autour d'un théorème de Tsirelson sur des filtrations browniennes et non browniennes*. Séminaire de Probabilités XXXII, Springer Lecture Notes in Math. 1686, 264–305, 1998.
- [3] M. BARLOW, J. PITMAN & M. YOR – *On Walsh's Brownian motions*. Séminaire de Probabilités XXIII, Springer Lecture Notes in Math. 1372, 275–293, 1989.
- [4] S. BEGHDADI-SAKRANI & M. ÉMERY – *On certain probabilities equivalent to coin-tossing, d'après Schachermayer*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 240–256, 1999.
- [5] C. BISHOP – *A characterization of Poissonian domains*. Ark. Mat. 29, 1–24, 1991.
- [6] B. DE MEYER – *Une simplification de l'argument de Tsirelson sur le caractère non-brownien des processus de Walsh*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 217–220, 1999.
- [7] P. DIACONIS – *From shuffling cards to walking around the building: an introduction to modern Markov chain theory*. Proceedings of the International Congress of Mathematicians, Doc. Math., Extra Vol. I, 187–204, 1998.
- [8] L. DUBINS, J. FELDMAN, M. SMORODINSKY & B. TSIRELSON – *Decreasing sequences of σ -fields and a measure change for Brownian motion*. Ann. Prob. 24, 882–904, 1996.
- [9] M. ÉMERY & W. SCHACHERMAYER – *Brownian filtrations are not stable under equivalent time-changes*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 267–276, 1999.
- [10] M. ÉMERY & W. SCHACHERMAYER – *A remark on Tsirelson's stochastic differential equation*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 291–303, 1999.
- [11] M. ÉMERY & W. SCHACHERMAYER – *On Vershik's standardness criterion and Tsirelson's notion of cosiness*. Séminaire de Probabilités XXXV, à paraître.
- [12] M. ÉMERY & M. YOR – *Sur un théorème de Tsirelson relatif à des mouvements browniens corrélés et à la nullité de certains temps locaux*. Séminaire de Probabilités XXXII, Springer Lecture Notes in Math. 1686, 306–312, 1998.
- [13] J. FELDMAN & M. SMORODINSKY – *Decreasing sequences of measurable partitions: product type, standard, and prestandard*. Ergodic Theory and Dynamical Systems 20, 1079–1090, 2000.

- [14] J. FELDMAN & B. TSIRELSON – *Decreasing sequences of σ -fields and a measure change for Brownian motion II*. Ann. Prob. 24, 905–911, 1996.
- [15] D. HEICKLEN & C. HOFFMAN. *T, T^{-1} is not standard*. Ergodic Theory and Dynamical Systems 18, 875–878, 1998.
- [16] T. LINDVALL – *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, 1992.
- [17] M. MALRIC – *Filtrations quotients de la filtration brownienne*. Séminaire de Probabilités XXXV, à paraître.
- [18] W. SCHACHERMAYER – *On certain probabilities equivalent to Wiener measures, d'après Dubins, Feldman, Smorodinsky and Tsirelson*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 221–239, 1999.
- [19] M. SMORODINSKY – *Processes with no standard extension*. Israel J. Math. 107, 327–331, 1998.
- [20] B. TSIRELSON – *Triple points: From non-Brownian filtrations to harmonic measures*. GAFA, Geom. funct. anal. 7, 1096–1142, 1997.
- [21] B. TSIRELSON – *Within and beyond the reach of Brownian innovation*. Proceedings of the International Congress of Mathematicians, Doc. Math., Extra Vol. III, 311–320, 1998.
- [22] B. TSIRELSON – *Toward stochastic analysis beyond the white noise*. Prépublication. <http://www.math.tau.ac.il/~tsirel/Research/Recent/toward.html>
- [23] B. TSIRELSON – *Noise sensitivity on continuous products: an answer to an old question of J. Feldman*. Prépublication. <http://www.math.tau.ac.il/~tsirel/Research/Recent/products.html>
- [24] A. M. VERSHIK – *Approximation in measure theory*. Thèse de doctorat, Leningrad 1973. Version anglaise augmentée et mise à jour : *The theory of decreasing sequences of measurable partitions*. St. Petersburg Math. J. 6, 705–761, 1995.
- [25] J. WARREN – *On the joining of sticky Brownian motion*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 257–266, 1999.
- [26] S. WATANABE – *The existence of a multiple spider martingale in the natural filtration of a certain diffusion in the plane*. Séminaire de Probabilités XXXIII, Springer Lecture Notes in Math. 1709, 277–290, 1999.
- [27] M. YOR – *Sur l'étude des martingales continues extrémales*. Stochastics 2, 191–196, 1979.

- [28] M. YOR – *Some Aspects of Brownian Motion. Part II: Some Recent Martingale Problems*. Birkhäuser, 1997.

Michel ÉMERY

C.N.R.S. et Université Louis Pasteur
I.R.M.A.

(URA 7501 du C.N.R.S.)

7 rue René Descartes

F-67084 STRASBOURG Cedex

E-mail : `emery@math.u-strasbg.fr`

L'INÉGALITÉ DE PENROSE
[d'après H. Bray, G. Huisken et T. Ilmanen, ...]

par Marc HERZLICH

L'inégalité de Penrose est un énoncé purement géométrique. Les objets qui y interviennent trouvent cependant leur source et une grande part de leurs motivations dans la théorie de la relativité générale. Il apparaît donc opportun (et c'est le choix fait par l'auteur de ces lignes) de consacrer une première partie de cet exposé à une description rapide du contexte physique, avant d'entrer plus avant dans les détails proprement mathématiques.

1. UN RAPIDE SURVOL DE RELATIVITÉ GÉNÉRALE.

La théorie physique de la relativité générale entretient des liens étroits et féconds avec la géométrie [8, 22, 30, 33]. Elle modélise l'espace-temps comme une variété différentielle (lisse) N de dimension 4 munie d'une métrique lorentzienne γ , de signature $(-+++)$, représentant le champ de gravitation. Les *équations d'Einstein* relient le champ de gravitation aux autres interactions en présence. La métrique y apparaît par l'intermédiaire de l'un des avatars de sa courbure, le *tenseur d'Einstein*, combinaison de la courbure de Ricci Ric^γ et de sa trace la courbure scalaire Scal^γ ; les autres champs sont combinés pour former le tenseur d'énergie-impulsion \mathbf{T} , dont l'expression varie suivant la nature des interactions en présence. Les équations s'écrivent alors :

$$(1.1) \quad \text{Ric}^\gamma - \frac{1}{2} \text{Scal}^\gamma \gamma = 8\pi \mathbf{T}.$$

Le tenseur d'Einstein est à divergence nulle, ce qui fournit les lois de conservation usuelles de la théorie des champs ($\text{div } \mathbf{T} = 0$).

La situation à laquelle nous allons particulièrement nous intéresser est celle des *systèmes isolés* dans laquelle, de manière très schématique, on impose à l'ensemble des interactions de décroître au fur et à mesure que l'on s'éloigne de leur source et même de s'annuler à l'infini. Cette exigence est difficile à formuler géométriquement dans un cadre lorentzien, mais on peut donner une famille d'exemples archétypiques d'un tel comportement : les métriques de Schwarzschild ont été introduites par K. Schwarzschild dès 1916 [47] pour

modéliser l'extérieur d'une étoile fixe et statique de masse m , représenté par la variété $N = \mathbf{R} \times (\mathbf{R}^3 \setminus B_0(\frac{m}{2}))$ munie de la métrique :

$$(1.2) \quad \gamma_{\text{schw.}} = -\frac{(1 - \frac{m}{2r})^2}{(1 + \frac{m}{2r})^2} d\tau^2 + \left(1 + \frac{m}{2r}\right)^4 \text{eucl}$$

où eucl désigne la métrique euclidienne de \mathbf{R}^3 (cette écriture, qui n'est pas la plus commune pour une métrique de Schwarzschild, se révélera la plus commode pour la suite). Cette famille de métriques est à symétrie sphérique et elles sont toutes "asymptotes" à l'infini à la métrique de Minkowski $-d\tau^2 + dx^2 + dy^2 + dz^2$ sur \mathbf{R}^4 (la situation où les interactions gravitationnelles sont absentes). La métrique semble devenir singulière lorsque l'on approche l'hypersurface $\{r = \frac{m}{2}\}$. Il ne s'agit en réalité que de singularités du système de coordonnées. Un changement de coordonnées, découvert par M. Kruskal [31], permet un prolongement lisse des métriques de Schwarzschild au-delà de l'hypersurface. Une fois les espaces de Schwarzschild ainsi "complétés", ils peuvent alors être interprétés comme le champ de gravitation d'un *trou noir* de masse m dont l'ouvert difféomorphe à $N = \mathbf{R} \times (\mathbf{R}^3 \setminus B_0(\frac{m}{2}))$ constitue la région "extérieure". L'hypersurface de séparation est l'*horizon* du trou noir : aucune courbe de type temps ou lumière (aucune particule) ne peut s'échapper de la région intérieure. Les sphères $\{\tau = \tau_0, r = \frac{m}{2}\}$ contenues dans l'horizon sont des surfaces minimales (et même totalement géodésiques) et elles entourent une (vraie) singularité de la métrique au centre du trou noir. La relation entre rayon de l'horizon et masse n'est pas fortuite ; elle est même au cœur de cet exposé.

1.1. Variétés asymptotiquement plates

Instruits par les exemples de Schwarzschild, nous imaginons désormais que toutes nos variétés sont le produit d'un intervalle (de type temps) par une variété *riemannienne* M de dimension 3. Via les équations de Gauss et Codazzi, tous les objets géométriques "de dimension 4" peuvent s'écrire en fonction des métriques et secondes formes fondamentales des différentes *tranches d'espace* $\{\tau\} \times M$. Cette traduction a l'avantage de transformer de très nombreuses interrogations sur la géométrie lorentzienne de l'espace-temps en des questions de *pure géométrie riemannienne*. Dans le cas qui nous préoccupe, se donner un système isolé revient à imposer à chaque tranche d'être *asymptotiquement plate* (d'autres conditions géométriques, analogues mais non nécessairement équivalentes, sont parfois utilisées, voir [23, 42, 51]) :

DÉFINITION 1.1. — *Une variété riemannienne (M^n, g) non compacte, à un seul bout, est dite asymptotiquement plate d'ordre $\rho > 0$ s'il existe un compact $K \subset M$ et un difféomorphisme (dit carte à l'infini) $z : M \setminus K \rightarrow \mathbf{R}^n \setminus B_0(r)$ (où $r > 0$) tels que les coefficients de la métrique vérifient dans cette carte $\{z^1, \dots, z^n\}$*

$$(1.3) \quad g_{ij} - \delta_{ij} = O(|z|^{-\rho}), \quad \frac{\partial}{\partial z^k} g_{ij} = O(|z|^{-\rho-1}), \quad \frac{\partial^2}{\partial z^k \partial z^\ell} g_{ij} = O(|z|^{-\rho-2}).$$

Cette définition reposant sur des cartes particulières et des expressions en coordonnées peut paraître surprenante. Elle définit cependant un bon objet géométrique ; nous avons néanmoins préféré l'exposer en premier lieu dans la mesure où elle fournit une vision très concrète des variétés asymptotiquement plates. Elle est justifiée par le travail dû à S. Bando, A. Kasue et H. Nakajima [2] qui montre l'équivalence entre platitude asymptotique et des conditions bien précises de décroissance de la courbure et de comportement du volume des grandes sphères géodésiques à l'infini.

1.2. La masse

Dans l'expression (1.2) des métriques de Schwarzschild, le paramètre m est la masse de l'étoile. Cette identification se trouve confortée par l'étude de l'équation différentielle régissant la trajectoire d'une géodésique radiale qui est de la forme $r'' = -mr^{-2} + O(r^{-3})$. Contrairement à l'intuition, la théorie des systèmes isolés ne dispose pourtant pas d'une notion universelle (et locale) de masse et estimer la quantité de matière présente dans une région bornée de l'espace(-temps) reste à l'heure actuelle une question mal comprise (voir la fin de ce texte).

La manière dont a été réalisée l'identification entre le paramètre m et la masse dans la métrique de Schwarzschild amène alors à l'idée que la masse totale d'un système isolé doit être calculée de façon *globale*, à l'aide d'un procédé de *passage à la limite*. La définition la plus couramment utilisée aujourd'hui est celle donnée en 1962 par R. Arnowitt, S. Deser et C. Misner [1], qui attribue une masse à toute *hypersurface asymptotiquement plate* d'un espace-temps. Cette définition, comme la précédente, est purement riemannienne et s'étend à toutes les dimensions.

DÉFINITION 1.2. — Soient (M^n, g) une variété riemannienne asymptotiquement plate, S_r une sphère de rayon r dans une carte à l'infini et ν sa normale sortante. La masse est

$$(1.4) \quad m(M, g) = \frac{1}{16\pi} \lim_{r \rightarrow \infty} \int_{S_r} (\partial_i g_{ij} - \partial_j g_{ii}) \nu^j \, d\text{vol}_{S_r}$$

si cette intégrale existe.

Le coefficient 16π est choisi ici pour que le calcul de la masse de toute tranche d'espace $\{\tau\} \times M$ de la métrique de Schwarzschild (1.2) soit exactement m .

Cet objet, calculé d'une manière apparemment non-tensorielle, – et à partir des *dérivées premières* de la métrique de surcroît ! – est pourtant bien défini. La confirmation de ce fait, connu de manière heuristique par les physiciens dès les années 60, se fera attendre jusqu'en 1985, année où est démontré le théorème crucial qui permet d'attribuer une signification non ambiguë à la masse, dû à R. Bartnik [4] et indépendamment à P. Chruściel [14].

THÉORÈME 1.3. — Si la variété riemannienne (M, g) de dimension n est asymptotiquement plate d'ordre $\rho > \frac{n-2}{2}$ et si sa courbure scalaire est dans l'espace de Lebesgue L^1 ,

alors sa masse est bien définie. C'est un invariant riemannien, indépendant du choix de la carte.

V. Denisov et O. Solovév ont construit dans [16] un contre-exemple en dimension 3, d'ordre de décroissance à l'infini exactement égal à $1/2$ et dont deux cartes à l'infini fournissent deux masses différentes.

Si les équations d'Einstein du vide ($\mathbf{T} = 0$) sont vérifiées dans un espace-temps asymptotiquement plat $N = \mathbf{R} \times M$, la masse est indépendante de la tranche choisie pour le calcul [15] et on peut donc parler de la masse de l'espace-temps. Dans le cas général, il est naturel d'espérer que la masse soit positive si les interactions en présence sont physiquement réalistes et qu'elle ne puisse être nulle que pour l'espace-temps trivial de Minkowski. Les critères utilisés par les physiciens pour décider du caractère raisonnable des interactions sont des inégalités sur leur tenseur d'énergie-impulsion \mathbf{T} , dites *conditions d'énergie* [22], qui se traduisent donc, *via* les équations d'Einstein, en inégalités sur la courbure.

Un cas particulier très important (et très étudié) est celui où l'hypersurface asymptotiquement plate (M, g) dont on calcule la masse est de courbure moyenne nulle (par exemple si elle est totalement géodésique, *i.e.* si sa seconde forme est nulle). Les conditions d'énergie impliquent alors que *la courbure scalaire de (M, g) est positive*. Tout ce qui précède apporte donc une motivation profonde au résultat fondamental suivant, dû à R. Schoen et S.T. Yau [44] en 1978. Une preuve alternative a été donnée en 1981 par E. Witten [52] (voir aussi [4, 13, 30, 32]). Cet énoncé est purement riemannien et ne fait plus référence à la relativité.

THÉORÈME 1.4 (Théorème de la masse positive). — *Si (M, g) est une variété riemannienne asymptotiquement plate de dimension 3, complète, sans bord et à courbure scalaire positive, alors sa masse est positive, et elle est nulle si et seulement si (M, g) est isométrique à l'espace euclidien $(\mathbf{R}^3, \text{eucl})$.*

Cette inégalité et la rigidité associée au cas d'égalité constituent donc un remarquable énoncé de géométrie riemannienne globale, dont certains corollaires sont frappants. On peut citer par exemple celui choisi par R. Geroch dans un exposé de 1974 au cours d'un symposium de l'A.M.S. pour tenter d'attirer l'attention des mathématiciens sur les problèmes issus de la relativité [19] : *la seule métrique riemannienne à courbure scalaire positive et plate en-dehors d'un compact de \mathbf{R}^3 est la métrique euclidienne*.

Ce théorème fait partie du paysage depuis 20 ans. Outre son importance en relativité, il joue un rôle crucial en géométrie conforme en relation avec la courbure scalaire, comme en témoigne le célèbre *problème de Yamabe* [32]. Mais on ne connaît pas encore ses conditions de validité les plus générales : à l'heure actuelle, il est vérifié en toutes dimensions si la variété est spinorielle (E. Witten) ou si la dimension est inférieure ou égale à 7 dans le cas général (preuve due à R. Schoen et S.-T. Yau, qui utilise des surfaces minimales, d'où

la restriction ; ils ont affirmé à plusieurs reprises posséder une preuve valable en toutes dimensions, mais, à la connaissance de l'auteur, elle semble ne jamais avoir été publiée).

1.3. Trous noirs et inégalité de Penrose

En 1973, R. Penrose a conjecturé une inégalité frappante pour la masse d'un espace-temps asymptotiquement plat contenant un trou noir [43].

Qu'est-ce qu'un trou noir ? En étant volontairement flou, un espace-temps asymptotiquement plat (défini comme précédemment, c'est-à-dire feuilleté par des hypersurfaces riemanniennes $\{\tau\} \times M$ asymptotiquement plates) en contient un s'il existe une région ("bornée" en espace) dont aucune courbe de type temps ou lumière ne peut s'échapper et où toutes les géodésiques de ce type sont incomplètes. Si l'on se limite au cas où les tranches d'espace $\{\tau\} \times M$ sont totalement géodésiques, la présence d'une surface minimale dans $\{\tau\} \times M$ est un bon indice de l'existence d'un trou noir [22, chap. 9]. Il existe alors toujours une collection de sphères minimales, dites "les plus extérieures", c'est-à-dire délimitant une région extérieure qui ne contient aucune autre sphère minimale que les composantes de son bord (ce fait est non trivial, voir la section 3.6 pour plus de détails). Ces surfaces forment l'*horizon* du trou noir [22, chap. 9] : l'intérieur est caché aux yeux de tout observateur extérieur.

Dans ce cadre désormais riemannien, la conjecture de Penrose s'énonce : *si (M, g) est une variété riemannienne asymptotiquement plate de dimension 3 qui contient une surface minimale compacte et qui satisfait aux hypothèses du théorème de la masse positive, alors*

$$(1.5) \quad m(M, g) \geq \frac{1}{4} \sqrt{\frac{\text{Aire}(S)}{\pi}}$$

où S est l'ensemble des sphères minimales "les plus extérieures" contenues dans M .

L'inégalité (1.5) est inspirée à la fois de la thermodynamique des trous noirs (qui impose des relations entre masse et entropie, c'est-à-dire aire) et d'une autre conjecture fameuse de relativité, la *censure cosmique*. Celle-ci prédit que la situation rencontrée dans les métriques de Schwarzschild (une singularité centrale entourée par un horizon) est en fait générique : tout espace-temps singulier admet une perturbation dans laquelle la singularité doit être entourée d'un horizon. Cet énoncé est probablement le problème ouvert le plus important pour les spécialistes de relativité ; sa vérification aurait des conséquences concrètes en astrophysique. Or R. Penrose présente dans [43] un argument, heuristique mais convaincant, qui va à l'encontre de la censure cosmique si l'inégalité (1.5) est violée !

On comprend donc aisément l'importance de la conjecture de Penrose et son attrait. Pris du côté de la physique, son énoncé est très naturel : la masse d'un espace-temps statique qui contient un trou noir est minorée par la masse de la métrique de Schwarzschild dont le trou noir a un horizon de même aire (c'est-à-dire, en termes physiques, de même

entropie). De plus, il n'est pas surprenant que l'énoncé ne concerne que la surface minimale la plus extérieure : la masse étant calculée à l'infini, ne peut y contribuer que la partie de l'espace-temps située au-delà de l'horizon puisque celui-ci masque le contenu intérieur aux yeux de l'infini... Mathématiquement parlant, il s'agit d'un résultat d'isolation qui impose à toute variété à courbure scalaire positive suffisamment différente de l'espace euclidien (qui ne contient pas de surface minimale compacte) d'avoir une masse non seulement strictement positive mais aussi notablement éloignée de zéro. Et les deux points de vue se rejoignent pour prédire que le cas d'égalité est caractéristique d'une tranche de la métrique de Schwarzschild (1.2), c'est-à-dire de la situation la plus symétrique.

La conjecture a résisté pendant près de 25 ans aux différentes tentatives de démonstrations (voir cependant [6, 39] pour des cas particuliers). Ce n'est que récemment (1997) que G. Huisken et T. Ilmanen [26, 27] ont pu apporter la preuve de sa véracité lorsque l'horizon est connexe. En 1999, H. Bray [9, 11] en a fourni une autre preuve.

THÉORÈME 1.5 (Bray, Huisken–Ilmanen). — *Soit (M, g) une variété asymptotiquement plate de dimension 3 à courbure scalaire positive ou nulle, ayant pour bord (intérieur) une sphère minimale S et ne contenant pas d'autres surfaces minimales. Alors*

$$(1.6) \quad m(M, g) \geq \frac{1}{4} \sqrt{\frac{\text{Aire}(S)}{\pi}}.$$

Le cas d'égalité n'est vérifié que si M est isométrique à $(\mathbf{R}^3 \setminus B_0(\frac{m}{2}), (1 + \frac{m}{2r})^4 \text{eucl})$.

2. APPROCHES POSSIBLES

Il existe plusieurs stratégies possibles pour étudier la masse, inventées au cours des années 70 pour démontrer le théorème de la masse positive. L'approche qui a conduit pour la première fois au succès, c'est-à-dire celle de R. Schoen et S.T. Yau, était une démonstration par l'absurde. Il était donc difficile de l'utiliser pour en déduire des estimations effectives sur la masse. En revanche, la simplicité de la méthode mise à profit par E. Witten dans sa propre preuve de la masse positive faisait d'elle une candidate naturelle pour l'attaque de l'inégalité de Penrose.

2.1. L'approche spinorielle

L'argument de E. Witten se fonde sur la géométrie spinorielle : sur toute variété riemannienne spinorielle, la formule de Lichnerowicz [34] relie l'opérateur de Dirac \mathcal{D} au laplacien brut D^*D de la connexion de Levi-Civita :

$$(2.1) \quad \mathcal{D}^*\mathcal{D} = D^*D + \frac{1}{4} \text{Scal}^g.$$

E. Witten a alors remarqué que, lorsque cette formule, évaluée sur un champ de spineurs ψ asymptotiquement constant (cette notion étant définie avec précision), est intégrée sur

des domaines de plus en plus grands, le terme de bord tend vers un multiple de la masse à l'infini, *i.e.*

$$(2.2) \quad \left(4\pi \lim_{\infty} |\psi|^2\right) m(M, g) = \int_M |D\psi|^2 + \frac{1}{4} \int_M \text{Scal}^g |\psi|^2 - \int_M |\mathcal{D}\psi|^2.$$

La positivité de la masse est alors immédiate si l'on sait exhiber un spineur harmonique (c'est-à-dire annulé par \mathcal{D}) asymptotiquement constant et de dérivée de carré intégrable.

L'existence d'un spineur solution est obtenue grâce à la théorie des espaces de Sobolev ou de Hölder à poids développée pour les variétés asymptotiquement plates [4, 36, 37]. Le cas d'égalité s'obtient aisément en remarquant que la nullité de la masse entraîne que le spineur est parallèle, donc la courbure de Ricci est nulle. Le cas limite de l'estimation classique du volume due à Bishop [7] et le caractère asymptotiquement plat font le reste.

Le principal intérêt de cette stratégie est la simplicité avec laquelle elle peut être mise en œuvre. Malheureusement, elle ne conduit pas à une preuve de l'inégalité de Penrose. La principale difficulté est d'interpréter dans ce contexte spinoriel l'exigence de ne considérer que la surface minimale *la plus extérieure*. Faire abstraction de cette restriction conduit néanmoins à un résultat, moins fort mais valable pour toute surface minimale. La borne inférieure obtenue fait intervenir la constante géométrique

$$(2.3) \quad \sigma(S) = \sqrt{\frac{\text{Aire}(S)}{\pi}} \left(\inf_{f \in C_c^\infty} \frac{\int_{M_S} |df|^2}{\int_S f^2} \right).$$

Après normalisation près, il s'agit de l'inverse de la norme de l'application trace qui envoie le complété (pour la norme de type Sobolev $\|df\|_{L^2(M_S)}$) de l'espace C_c^∞ des fonctions lisses à support compact sur l'extérieur M_S de la surface minimale S dans l'espace $L^2(S)$ des fonctions de carré intégrable sur S . Elle est donc strictement positive par nature et invariante par changement d'échelle, la valeur précise de l'aire de S n'y apparaissant qu'à des fins de normalisation. Le résultat, obtenu dans [24] à une époque où aucune preuve de l'inégalité de Penrose n'était disponible, peut alors s'énoncer :

THÉORÈME 2.1. — *Si (M, g) est une variété asymptotiquement plate de dimension 3 à courbure scalaire positive, ayant pour bord une sphère minimale S , alors*

$$(2.4) \quad m(M, g) \geq \frac{\sigma(S)}{2 + 2\sigma(S)} \sqrt{\frac{\text{Aire}(S)}{\pi}}.$$

Le cas d'égalité n'est vérifié que si $M = (\mathbf{R}^3 \setminus B_0(\frac{m}{2}), (1 + \frac{m}{2r})^4 \text{eucl})$.

La preuve repose avant tout sur une étude précise du cas d'égalité : celui-ci fait ressortir le rôle joué par la métrique euclidienne, dont la masse est nulle, qui est la seule métrique conforme à la (tranche spatiale de la) métrique de Schwarzschild de courbure scalaire nulle et dans laquelle la sphère S est à courbure moyenne constante égale à $4\sqrt{\frac{\pi}{\text{Aire}(S)}}$. La preuve suit alors le cours suivant : on déforme notre métrique d'origine en une métrique conforme $\bar{g} = \varphi^4 g$ qui est à courbure scalaire nulle et dont la courbure moyenne de S

vérifie exactement (dans la métrique \bar{g}) l'égalité précédente. Un calcul direct livre alors que la différence entre la masse de g et celle de \bar{g} est minorée par la constante annoncée. Il reste à exhiber pour \bar{g} un théorème de masse positive, obtenu par la méthode de Witten, pour la partie non bornée M_S . La subtilité provient de l'apparition, lors de l'intégration de la formule de Lichnerowicz (2.2), de termes de bord sur S contenant l'opérateur de Dirac de celle-ci. Il est alors possible de contrôler son signe en faisant appel aux travaux de C. Bär et O. Hijazi sur la première valeur propre de l'opérateur de Dirac [3, 25].

2.2. Le flot de la courbure moyenne inverse

De nombreuses tentatives ont été faites pour donner une expression plus intrinsèque de la masse. L'idée la plus intéressante est due à S.W. Hawking [21] : sur une variété asymptotiquement plate de dimension 3,

$$(2.5) \quad m(M, g) = \lim_{t \rightarrow \infty} \frac{\sqrt{\text{Aire}(S_t)}}{64\pi^{3/2}} \left(16\pi - \int_{S_t} H_t^2 \right)$$

si les surfaces S_t (de courbure moyenne H_t) convergent très vite (en un sens précis) vers des sphères euclidiennes à l'infini. L'expression à l'intérieur de la limite dans (2.5) est appelée *masse de Hawking* $m_H(S_t)$ d'une surface S_t et vaut zéro pour toute sphère ronde dans l'espace euclidien (c'est le théorème de Gauss-Bonnet !)

Cette remarque conduit à une nouvelle approche de la masse positive, fondée sur une *formule de monotonie*. R. Geroch [18] et P. S. Jang et R. Wald [29] ont en effet imaginé que toute variété asymptotiquement plate (de dimension 3) "raisonnable" puisse être recouverte par des sphères (topologiques) évoluant sous le flot de la *courbure moyenne inverse*. Plus précisément, ils espèrent disposer d'une application lisse et *régulière* (sans points critiques) $F : \mathbf{R} \times \mathbf{S}^2 \rightarrow M$, $(t, x) \mapsto F_t(x)$, obéissant à l'équation

$$(2.6) \quad \frac{\partial F_t}{\partial t}(x) = \frac{1}{H_t(x)} \nu_t(x)$$

où ν_t est le champ de normales unitaires sortant et H_t la courbure moyenne de $S_t = F_t(\mathbf{S}^2)$. Par exemple, si S_0 est une sphère ronde de rayon r_0 dans l'espace euclidien \mathbf{R}^3 , la solution du flot (2.6) est fournie par les sphères concentriques S_t de rayons $r(t) = e^{t/2} r_0$.

LEMME 2.2. — Si $\text{Scal}^g \geq 0$, la masse de Hawking $m_H(S_t)$ est une fonction croissante.

Preuve. — Tout d'abord on a, pour tout t , $\text{Aire}(S_t) = e^t \text{Aire}(S_0)$ puisque la courbure moyenne est précisément la dérivée du volume infinitésimal. Notons K_t la courbure de Gauss et A_t^0 la partie à trace nulle de la seconde forme fondamentale A_t de S_t , on peut calculer la dérivée de H_t par rapport à t à partir de la formule (2.6). Cela conduit à :

$$(2.7) \quad \frac{d}{dt} \int_{S_t} H_t^2 = - \int_{S_t} 2 H_t^{-2} |dH_t|^2 + 2 |A_t^0|^2 + 2 \text{Ric}(\nu_t, \nu_t).$$

L'équation de Gauss fournit par ailleurs

$$(2.8) \quad -2 \operatorname{Ric}(\nu_t, \nu_t) + \operatorname{Scal}^g = 2 K_t + |A_t^0|^2 - \frac{1}{2} H_t^2$$

(une remarque déjà utilisée par R. Schoen et S.T. Yau dans le contexte de la courbure scalaire [45]), d'où

$$(2.9) \quad \frac{d}{dt} \left(16\pi - \int_{S_t} H_t^2 \right) \geq \int_{S_t} -2 K_t + \operatorname{Scal}^g + |A_t^0|^2 + \frac{1}{2} H_t^2 \geq -\frac{1}{2} \left(16\pi - \int_{S_t} H_t^2 \right)$$

(en utilisant Gauss-Bonnet). Compte tenu du comportement de l'aire décrit plus haut, on obtient finalement

$$(2.10) \quad \frac{d}{dt} m_H(S_t) = \frac{\sqrt{\operatorname{Aire}(S_0)}}{64\pi^{3/2}} \frac{d}{dt} \left(e^{\frac{t}{2}} \left(16\pi - \int_{S_t} H_t^2 \right) \right) \geq 0,$$

ce qui conclut la preuve. \square

Si l'on fait l'hypothèse que le flot peut évoluer de cette façon *régulière* à partir d'une très petite sphère géodésique à $t = \varepsilon$ (dont la masse de Hawking est très proche de zéro puisque la géométrie riemannienne est asymptote à l'ordre 2 à la géométrie euclidienne au voisinage d'un point) et se comporte lorsque t tend vers l'infini comme les grandes sphères géodésiques, ce procédé fournit donc une nouvelle preuve de la masse positive.

C'est cette méthode que mettent en œuvre G. Huisken et T. Ilmanen dans leur preuve de l'inégalité de Penrose, au prix de formidables difficultés techniques. Même dans les cas les plus simples, il est en effet inévitable de voir se créer en temps fini des singularités du flot. Les contrôler de manière efficace et développer les outils permettant d'y faire face constitue l'essentiel de leur travail.

2.3. Trous noirs et inégalité isopérimétrique

L'idée de H. Bray pour démontrer l'inégalité est de considérer le quotient de la masse par la racine carrée de l'aire de l'horizon $\mathcal{P}(g) = \operatorname{Aire}(S(g))^{-\frac{1}{2}} m(g)$, qu'il dénomme *quotient de Penrose* de la variété et qu'il interprète comme une constante isopérimétrique pour l'horizon $S(g)$, ou encore comme une fonctionnelle sur l'espace des métriques asymptotiquement plates qui possèdent un horizon.

Plutôt que de tenter de rechercher directement un minimum de cette fonctionnelle, il définit un flot de métriques (et d'horizons associés) qui fait décroître le quotient de Penrose. Pour ce faire, il construit une famille de métriques conformes $g_t = \varphi_t^4 g$ sur M , de quotient décroissant en t et munies d'horizons $S(g_t)$ qui, lorsque t tend vers l'infini, ont progressivement tendance à "enfler" (lorsqu'on les étudie avec la métrique d'origine) et donc à tendre vers l'infini. Le point crucial est le suivant : la restriction d'une métrique asymptotiquement plate arbitraire à l'extérieur d'un compact de plus en plus grand approxime de mieux en mieux une métrique de Schwarzschild qui est en quelque sorte sa *meilleure approximation à l'ordre 1 à l'infini*. Puisque $S(g_t)$ "tend vers l'infini", il est

donc permis d'espérer que la métrique g_t tende vers une métrique de Schwarzschild de quotient de Penrose $(16\pi)^{-1/2}$. Cette approche est entièrement nouvelle et originale dans l'étude de la masse et des variétés asymptotiquement plates. Contrairement à celle de G. Huisken et T. Ilmanen, elle débouche sur une inégalité de Penrose générale où l'horizon n'est plus nécessairement connexe (voir le Théorème 4.7).

3. ESQUISSE DE LA PREUVE DE G. HUISKEN ET T. ILMANEN

Comme dit précédemment, leur preuve consiste à rendre rigoureuse l'approche du flot de la courbure moyenne inverse introduite par R. Geroch, P.S. Jang et R. Wald. Leur travail se décompose en deux parties distinctes : une étude générale du flot, valable en toute généralité et en toutes dimensions (ou au moins en dimension $n < 8$ en ce qui concerne les résultats de régularité), et une application au cas particulier de l'inégalité de Penrose sur une variété asymptotiquement plate à courbure scalaire positive.

3.1. Difficultés

Lorsque l'on examine du point de vue des mathématiques la stratégie des physiciens, on se heurte immédiatement à un obstacle de taille : *il est a priori impossible d'exclure l'apparition de singularités lors du flot*. Un exemple simple (sur lequel nous reviendrons plus tard) permet déjà de s'en convaincre : le flot de la courbure moyenne inverse appliqué à une sphère ronde S_0 dans \mathbf{R}^3 produit une famille de sphères rondes concentriques S_t , paramétrées de telle sorte que $\text{Aire}(S_t) = e^t \text{Aire}(S_0)$. Prenons maintenant comme donnée initiale deux sphères disjointes : chacune évolue indépendamment tant qu'elles restent disjointes et une singularité apparaît au premier point de contact !

L'accident qui survient ici montre qu'il est crucial de ne pas restreindre le type topologique des surfaces qui apparaissent. Dès lors, la formulation lisse du flot (2.6) de la section 2.2 doit être remplacée par une formulation *faible*, exprimée dans le langage de la théorie géométrique de la mesure. G. Huisken et T. Ilmanen procèdent alors en trois étapes :

(a) la réécriture du flot lisse comme une équation aux dérivées partielles portant sur une fonction u dont les surfaces S_t sont les *surfaces de niveau*. Cette idée, inspirée par les travaux de L.C. Evans et J. Spruck [17], conduit ici à l'équation équivalente

$$(3.1) \quad \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) = |\nabla u|.$$

(b) il serait alors agréable de voir l'EDP (3.1) précédente comme l'équation d'Euler-Lagrange d'une fonctionnelle portant sur u et sur ∇u . Malheureusement, ce n'est pas le cas. Pour contourner cette difficulté, G. Huisken et T. Ilmanen s'inspirent de travaux de R. Hardt et X. Zhou [20], S. Luckhaus [38], A. Lichnerowsky et R. Temam [35] et A. Visintin

[50]. Si Ω est un ouvert de M , K un compact de Ω et u une fonction lipschitzienne fixée de Ω , ils introduisent la fonctionnelle

$$(3.2) \quad J_u^K(v) = \int_K |\nabla v| + v |\nabla u|$$

portant sur toutes les fonctions v de $C_{\text{loc}}^{0,1}(\Omega)$ qui diffèrent de u sur K . Une fonction u lisse et sans points critiques vérifie l'équation (3.1) dans Ω si et seulement si $J_u^K(u) \leq J_u^K(v)$ pour tout K et tout v . L'introduction des compacts K est destinée à compenser le manque de compacité des niveaux de u , une difficulté inhérente à la théorie.

(c) On peut alors remplacer la formulation variationnelle (3.2) par une nouvelle fonctionnelle portant sur les parties de \mathbf{R}^n . Pour une partie F et un compact K de \mathbf{R}^n ,

$$(3.3) \quad \tilde{J}_u^K(F) = \text{Aire}(\partial F \cap K) - \int_{F \cap K} |\nabla u|,$$

où le bord ∂F doit être pris ici au sens de la théorie géométrique de la mesure (on prendra donc F parmi les courants de *périmètre fini*, voir par exemple [48, sections 14, 26, 27]). On dit alors qu'une partie E minimise \tilde{J}_u si et seulement si $\tilde{J}_u^K(E) \leq \tilde{J}_u^K(F)$ pour toute partie F de M telle que la différence symétrique de E et F soit relativement compacte et pour tout choix de compact K qui contienne cette différence symétrique. Suivant le contexte, il sera commode d'utiliser l'une ou l'autre des fonctionnelles J_u ou \tilde{J}_u . Ces aller-retours (incessants dans [26]) sont justifiés par le résultat suivant :

PROPOSITION 3.1. — Soit Ω un ouvert de M et $u \in C_{\text{loc}}^{0,1}(\Omega)$. Alors u minimise la fonctionnelle J_u sur Ω si et seulement si, pour tout t , l'ensemble de niveau $E_t(u) = \{u < t\}$ minimise \tilde{J}_u sur Ω .

Esquisse de preuve. – Un sens est immédiat en remarquant que, pour tout v différent de u sur un compact K , la formule de la coaire [48, §12.6] fournit, si l'on note $E_t(v) = \{v < t\}$:

$$(3.4) \quad \begin{aligned} J_u^K(v) &= \int_{\inf v}^{\sup v} \text{Aire}(\partial E_t(v)) - \int_K \int_{\inf v}^{\sup v} \chi_{\{v(x) < t\}} |\nabla u| \, dt \, d\text{vol}_g + (\sup v) \int_K |\nabla u| \\ &= \int_{\inf v}^{\sup v} \tilde{J}_u^K(E_t(v)) + (\sup v) \int_K |\nabla u|. \end{aligned}$$

L'autre sens se démontre par l'absurde. Soit F , qu'on suppose d'abord inclus dans un E_{t_0} , tel que $\tilde{J}_u^K(F) < \tilde{J}_u^K(E_{t_0})$ (les théorèmes généraux de théorie géométrique de la mesure permettent de le choisir minimal pour cette propriété). On définit alors v telle que $E_t(v) = \{v < t\} = E_t \cap F$ et on compare $J_u^K(u)$ à $J_u^K(v)$ à l'aide de la formule précédente (3.4) et de l'inégalité très générale

$$(3.5) \quad \tilde{J}_u(A \cap B) + \tilde{J}_u(A \cup B) \leq \tilde{J}_u(A) + \tilde{J}_u(B).$$

Il s'ensuit que $\tilde{J}_u^K(E_t) = \tilde{J}_u^K(E_t \cap F)$ pour presque tout t et finalement, pour une suite $\{t_i\}$ croissante vers t_0 , $\tilde{J}_u^K(E_{t_0}) \leq \underline{\lim} \tilde{J}_u^K(E_{t_i} \cup F) \leq \tilde{J}_u^K(F)$, en utilisant une deuxième

fois l'inégalité (3.5). On procède alors de manière analogue pour un F contenant E_{t_0} et le cas général (F quelconque) est alors une conséquence de (3.5). \square

Nous dénotons maintenant par E_0 la condition initiale, qui est un ouvert arbitraire de M (dans le cas de l'inégalité de Penrose, il s'agit de la partie compacte de M délimitée par la surface minimale la plus extérieure). Le problème de minimisation (avec obstacle) de G. Huisken et T. Ilmanen peut donc être récrit de la manière suivante :

PROBLÈME 3.2. — *Trouver une famille $(E_t)_{t>0}$ de parties de \mathbf{R}^3 , croissante pour l'inclusion, telle que*

- 1) E_0 est contenu dans tous les E_t ;
- 2) si l'on définit une fonction u , qui est lipschitzienne, par $E_t = \{u < t\}$, alors E_t minimise la fonctionnelle \tilde{J}_u dans $M - E_0$, pour tout $t > 0$.

Il n'y a aucune raison pour que l'équation soit *a priori* satisfaite en $t = 0$: E_0 est choisi de manière arbitraire et n'a donc aucune raison de minimiser la fonctionnelle \tilde{J}_u . En effet, les conditions précédentes impliquent au mieux que $F_0 = \{u \leq 0\}$ minimise \tilde{J}_u dans $M - E_0$, mais rien n'empêche $\{u = 0\}$ de contribuer à la fonctionnelle (Figure 1). De manière générale, il faut donc imposer des conditions supplémentaires pour relier la condition initiale à l'évolution future du flot.

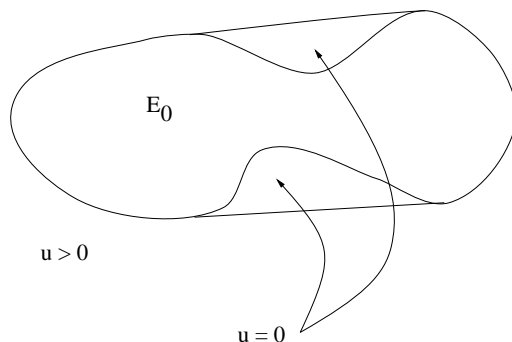


Fig. 1 : Une mauvaise condition initiale.

3.2. Régularité

La régularité des solutions éventuelles au flot reformulé est la conséquence de théorèmes classiques de théorie géométrique de la mesure, issus des travaux de W. Allard, F. Almgren et H. Federer. Leur application à la situation étudiée ici se trouve essentiellement contenue dans les travaux de U. Massari et M. Miranda [40] et P. Sternberg, G. Williams et W. Ziemer [49].

THÉORÈME 3.3. — *On suppose que la dimension est inférieure ou égale à 7. Soit f une fonction de $L^\infty(M)$ et U une partie de M de bord de classe C^1 (resp. $C^{1,\alpha}$ pour $0 < \alpha < 1$, resp. C^2 et $f = 0$). Si E minimise la fonctionnelle $\text{Aire}(\partial E) + \int_E f$ (parmi toutes les parties de M qui contiennent l'obstacle U), alors ∂E est une sous-variété de classe C^1 (resp. $C^{1,\alpha}$, resp. $C^{1,1}$ et C^∞ en-dehors des points de contact avec l'obstacle).*

3.3. Comportement lors des sauts

Le principal intérêt de la fonctionnelle \tilde{J}_u introduite au paragraphe précédent est de gérer au mieux le comportement des surfaces de niveau lorsque se produit un changement de topologie ou *saut* (ce qui est inévitable, comme nous l'avons déjà vu, si l'on souhaite que le flot soit défini pour tout temps).

Soit u une solution du Problème 3.2. Les ensembles de niveau $E_t = \{u < t\}$ minimisent donc la fonctionnelle \tilde{J}_u . Comme le second terme dans la définition (3.2) de \tilde{J}_u est toujours négatif et décroissant pour l'inclusion, leur bord est de périmètre minimum parmi toutes les parties de M qui les contiennent. Les parties $F_t = \{u \leq t\}$ sont la limite des E_s pour $s \rightarrow t^+$ et il est aisé de montrer qu'ils vérifient la même propriété de minimisation. On a ainsi $\text{Aire}(\partial E_t) = \text{Aire}(\partial F_t)$ dès qu'ils sont compacts. Nous dirons alors qu'une partie de M est une enveloppe minimisante (stricte) si elle est de périmètre minimum (strict) parmi toutes les parties qui la contiennent.

Comme F_t minimise non seulement le périmètre (du bord) mais aussi \tilde{J}_u et que ∇u doit être non nul au moins quelque part sur toute partie contenant strictement F_t (et qui en diffère d'un ensemble de mesure non-nulle), on en déduit que les F_t sont toujours des enveloppes minimisantes strictes. Le comportement d'une solution donnée du flot faible est maintenant intuitivement clair : tant que E_t reste une enveloppe minimisante stricte, le flot évolue de manière classique ; dès que E_t n'est plus strictement minimisant, le flot saute à F_t , qui est la plus petite enveloppe minimisante stricte contenant E_t .

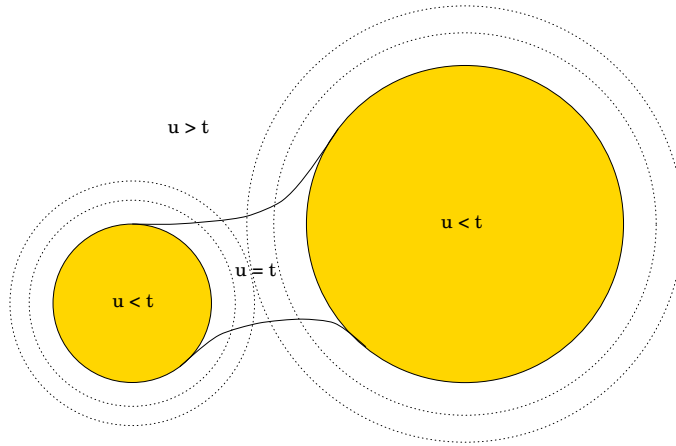


Fig. 2 : Il est temps de sauter...

La Figure 2 donne une idée du phénomène pour l'évolution de deux sphères disjointes évoquée dans la section 3.1. Les sphères en pointillés représentent l'évolution ultérieure du flot classique, qui aboutit à une singularité au premier point de contact. Le flot faible "saute" à la région formée des deux sphères reliées par un collier caténoïdal dès que la réunion des deux sphères peut être entourée par une autre partie de même périmètre. Cette situation est réminiscente d'un autre phénomène, bien connu, concernant les surfaces minimales : la solution du problème de Plateau (c'est-à-dire de la recherche de la surface

d'aire minimale de bord prescrit dans \mathbf{R}^3) pour deux cercles parallèles est un tronç de caténoïde si la distance des deux cercles est faible et deux disques sinon.

Une propriété étrange de ce comportement est qu'il est impossible d'exclure la possibilité d'un saut jusqu'à l'infini. En effet, si u est une solution du problème, alors $u_t = \min(u, t)$ est également une solution, *et ce pour tout t !* On voit ici que, prises sans précautions, les solutions du flot faible souffrent d'une forme particulièrement brutale de non-unicité (Figure 3).

Il peut même arriver que le saut jusqu'à l'infini soit le seul comportement possible. Si M est par exemple formée d'un bout cylindrique recollé à une demi-sphère et si l'on prend l'équateur comme condition initiale, la seule solution du flot faible est la fonction nulle sur la partie cylindrique puisqu'il est impossible de trouver une enveloppe minimisante stricte qui contienne l'équateur !

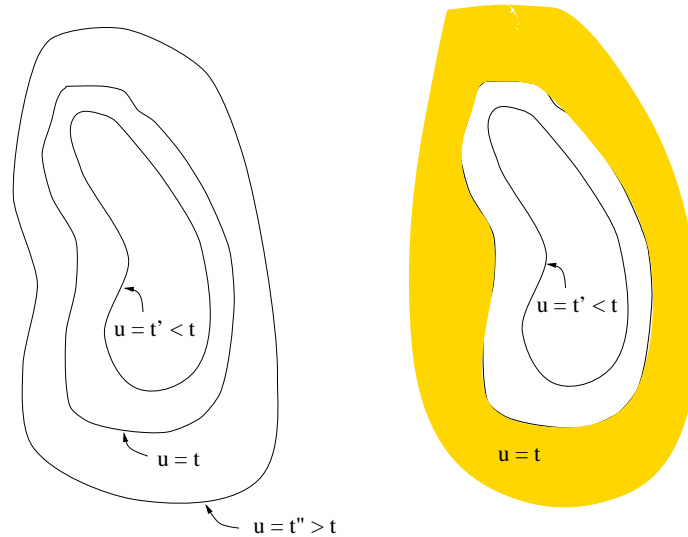


Fig. 3 : Un saut jusqu'à l'infini...

L'étude du comportement lors des sauts permet néanmoins de répondre à une question laissée auparavant en suspens : celle du comportement à $t = 0$. Puisque la fonctionnelle \tilde{J}_u sur $M - E_0$ se réduit ici au périmètre du bord, on obtient

PROPOSITION 3.4. — *Soit $(E_t)_{t>0}$ une solution du Problème 3.2 de donnée initiale E_0 . Alors E_0 vérifie l'équation à $t = 0$ si et seulement si E_0 est une enveloppe minimisante.*

Rappelons que tous les E_t minimisent \tilde{J}_u , qui doit donc être constante en t . Dès lors, la formule de la coaire et les considérations précédentes assurent :

PROPOSITION 3.5. — *Soit $(E_t)_{t>0}$ une solution du Problème 3.2. Si tous les E_t sont relativement compacts, alors $e^{-t} \text{Aire}(\partial E_t)$ est une fonction constante en t , égale à $\text{Aire}(\partial F_0)$. Si E_0 est une enveloppe minimisante, alors $\text{Aire}(\partial E_0) = \text{Aire}(\partial F_0)$.*

3.4. Propriétés d'unicité

Comme on vient de le voir, elles sont nécessairement subtiles. L'énoncé suivant, où aucune hypothèse n'est innocente, donne une idée de ce que l'on peut quand même obtenir.

PROPOSITION 3.6. — *Soit Ω un ouvert de M dont aucune composante connexe n'est compacte. Alors :*

(i) *si u et v minimisent les fonctionnelles associées J_u et J_v dans Ω et si $\{v > u\}$ est relativement compact dans Ω , alors $u \geq v$ dans l'ouvert ;*

(ii) *si $(E_t)_{t>0}$ et $(F_t)_{t>0}$ sont deux solutions du Problème 3.2 dans Ω de données initiales $E_0 \subset F_0$ et si E_t est relativement compact dans Ω pour tout t , alors $E_t \subset F_t$ pour tout t .*

Compte tenu de ce qui précède et de l'équivalence entre le flot classique (2.6) de la section 2.2 et sa formulation faible (étapes **a**)–**c**) de la section 3.1), on a alors

PROPOSITION 3.7. — *Soit $(E_t)_{t>0}$ une solution du Problème 3.2. Si à t_0 donné ∂E_{t_0} est une sous-variété lisse, compacte, à courbure moyenne strictement positive et est une enveloppe minimisante, alors le flot faible et le flot classique coïncident sur un petit intervalle de temps après t_0 .*

Toutes les hypothèses sont nécessaires. La non-annulation de la courbure moyenne donne un sens à l'équation classique, son signe assure son caractère parabolique, il existe donc une solution au flot classique sur un voisinage tubulaire de la donnée initiale. On utilise ensuite le résultat de comparaison 3.6(ii) dans ce voisinage, à la condition expresse que l'intersection de chaque ensemble de niveau $E_t = \{u < t\}$ du flot faible avec ce voisinage y soit relativement compacte (l'exemple des deux sphères montre la nécessité de cette hypothèse). D'après la discussion sur les sauts de la section précédente, la condition d'enveloppe assure que $E_{t_0} = \{u < t_0\}$ est l'intérieur de $F_{t_0} = \{u \leq t_0\}$, qui est la limite pour $t \rightarrow t_0^+$ (pour la distance de Hausdorff) des E_t ($t > t_0$).

3.5. Existence de solutions convenables

G. Huisken et T. Ilmanen obtiennent une solution u , lipschitzienne et surtout *propre*, du flot faible formulé à l'aide de la fonctionnelle J_u , par un procédé de *régularisation elliptique* (voir [17, 28]), c'est-à-dire comme limite d'une suite de fonctions lisses solutions d'équations régularisées.

Il est illusoire de prétendre donner une idée précise des arguments tout en conservant une taille raisonnable à ce rapport. Nous nous contenterons donc de décrire rapidement la méthode. Pour tout $\varepsilon > 0$, G. Huisken et T. Ilmanen montrent d'abord l'existence de solutions particulières u^ε aux *équations régularisées*

$$(3.6) \quad \operatorname{div} \left(\frac{\nabla u^\varepsilon}{\sqrt{|\nabla u^\varepsilon|^2 + \varepsilon^2}} \right) = \sqrt{|\nabla u^\varepsilon|^2 + \varepsilon^2}$$

sur des domaines Ω_ε , compris entre la donnée initiale E_0 et des hypersurfaces A_ε qui s'enfuient à l'infini si ε tend vers 0. Cette procédure peut être interprétée géométriquement si l'on remarque que la fonction $(z, x) \mapsto u^\varepsilon(x) - \varepsilon z$ fournit une solution au flot classique dans $\mathbf{R} \times \Omega_\varepsilon$.

Les solutions souhaitées des équations régularisées sont obtenues par application répétée de la méthode de continuité. Il reste alors un long travail pour mettre à jour les estimées vérifiées par la suite des solutions et qui assurent sa convergence dans $C_{\text{loc}}^{0,1}$, modulo extraction d'une sous-suite, vers la solution du flot faible souhaitée.

La mise en œuvre de ce schéma de preuve requiert un contrôle fin de la géométrie à l'infini. C'est en effet celle-ci qui assure qu'il existe des solutions propres, qui donc résistent à la tentation de "sauter jusqu'à l'infini" comme dans la Figure 3. Plutôt que d'imposer des contraintes particulières au comportement asymptotique de la variété, G. Huisken et T. Ilmanen préfèrent faire une hypothèse équivalente mais de nature analytique : l'existence d'une fonction, lipschitzienne et propre, sous-solution (en un sens qui est précisé dans [26]) au flot faible en-dehors d'un compact (arbitraire, éventuellement très grand). L'existence de cette sous-solution est l'ingrédient essentiel de toutes les estimées portant sur la suite des solutions des équations régularisées : elles sont toutes obtenues par comparaison avec la sous-solution et application de formes variées et parfois raffinées du principe du maximum. Le résultat de ces efforts est le

THÉORÈME 3.8 (Existence de solutions). — *Soit M une variété connexe, complète et E_0 une condition initiale au sens du Problème 3.2. S'il existe une sous-solution v dans $C_{\text{loc}}^{0,1}$ du flot faible à l'extérieur d'un compact et qui est propre, alors il existe une unique solution u propre dans $C_{\text{loc}}^{0,1}(M - E_0)$, de donnée initiale E_0 .*

3.6. Région extérieure, formule de monotonie et inégalité de Penrose

Plaçons-nous dans la situation de la conjecture de Penrose : une variété asymptotiquement plate de dimension 3 contenant une surface minimale. Nous commençons, comme promis, par quelques préliminaires sur les régions extérieures de M , c'est-à-dire les composantes de ce qui reste lorsque l'on ôte de M toutes les surfaces minimales compactes plongées ainsi que toutes les parties compactes de M délimitées par ces mêmes surfaces. Des arguments relativement élémentaires de géométrie et topologie de dimension 3 montrent alors [26] :

LEMME 3.9. — *Chaque région extérieure M_e a la topologie de \mathbf{R}^3 privé d'un nombre fini de sphères, minimales et stables. Le bout unique de M_e est asymptotiquement plat et ∂M_e est un nombre fini de surfaces minimales stables.*

Nous supposerons désormais que M_e est diffeomorphe à $\mathbf{R}^3 - B_0$. L'unique composante du bord, notée S_0 , est donc la sphère minimale *la plus extérieure* de l'inégalité de Penrose, telle qu'énoncée au Théorème 1.5. On procède alors comme suit.

(a) la géométrie à l'infini étant bien maîtrisée, le Théorème 3.8 fournit une fonction propre et lipschitzienne u dont les ensembles de niveau $E_t = \{u < t\}$ sont solution du Problème 3.2 avec comme condition initiale la partie de M délimitée par S_0 , qui est une sphère minimale et stable. Selon le principe exposé à la Proposition 3.4, l'équation est vérifiée en $t = 0$.

(b) *monotonie de la masse de Hawking* : Comme les solutions u^ε des équations régularisées de la section 3.5 sont des solutions du flot classique sur le cylindre $\mathbf{R} \times M$ (associées aux $N_t^\varepsilon = \{(u^\varepsilon(x) - \varepsilon z) = t\}$), la preuve de la monotonie (Lemme 2.2) fournit, pour $r \leq s$,

$$(3.7) \quad \int_{N_r^\varepsilon} \varphi H^2 - \int_{N_s^\varepsilon} \varphi H^2 \leq \int_r^s \int_{N_t^\varepsilon} \varphi (H^2 - 2|A|^2 - 2\text{Ric}(\nu, \nu)) + \langle d\varphi, H\nu - 2\frac{dH}{H} \rangle,$$

où φ est une fonction plateau en la variable z (son introduction est nécessaire car les N_t^ε convergent vers les cylindres $\mathbf{R} \times \partial E_t$ qui ne sont pas compacts !) Faire converger les inégalités 3.7 lorsque ε tend vers 0 est une des parties la plus technique du travail. Les termes les plus délicats sont ceux en la courbure moyenne H^2 (le lemme de Fatou est insuffisant car incompatible avec l'inégalité voulue) et en la seconde forme $|A|^2$. Une version faible de ces deux notions doit être utilisée (voir [48, §16]) et leur comportement lors des convergences bien contrôlé (au passage, G. Huisken et T. Ilmanen montrent et utilisent un théorème de Gauss-Bonnet pour ces notions faibles). On conclut que la masse de Hawking (faible) des niveaux de la solution fournie par la section 3.5 est croissante.

(c) *comportement asymptotique de $E_t = \{u < t\}$* : il est gouverné par les propriétés d'unicité du flot faible. Considérons la solution précédente du flot de la courbure moyenne inverse restreinte au complémentaire d'une grande boule B fixée de \mathbf{R}^3 , munie de la famille des métriques rééchelonnées $g_R = R^{-2}g$. Elles convergent vers la métrique euclidienne sur $\mathbf{R}^3 - \{0\}$ lorsque R tend vers l'infini ; de plus, il est possible d'exhiber une suite de solutions qui converge vers une solution du flot dans $\mathbf{R}^3 - \{0\}$. Il reste à s'assurer que la seule solution du flot sur $\mathbf{R}^3 - \{0\}$ dont tous les ensembles de niveau sont relativement compacts est donnée par les sphères concentriques en l'origine. Ce point est vraisemblable si l'on relit la Proposition 3.6 mais nécessite une preuve *ad hoc*. Les bords ∂E_t sont donc, asymptotiquement, de plus en plus sphériques.

Nous avons alors en mains tous les instruments nécessaires : l'existence globale du flot faible à partir de l'horizon et la monotonie de la masse de Hawking. Sa valeur à l'origine est la racine carrée de l'aire de l'horizon (convenablement renormalisée) et elle tend à l'infini vers la masse. L'inégalité de Penrose est donc démontrée.

REMARQUE 3.10. — L'étape (b), en particulier par l'usage du théorème de Gauss-Bonnet, ne vaut que si les surfaces ∂E_t , et donc l'horizon, sont connexes. La preuve s'étend au cas général en démarrant le flot sur la composante d'aire maximale de l'horizon, l'interrompant dès que les surfaces de niveau approchent de trop près d'autres composantes ; on construit alors "à la main" un saut qui englobe les nouvelles composantes et

qui préserve la monotonie de la masse de Hawking, avant de reprendre l'évolution normale du flot. On minore ainsi la masse en fonction de l'aire de la composante la plus grande, ce qui est moins précis que la conjecture initiale de Penrose, qui est le résultat de H. Bray décrit dans les paragraphes suivants.

4. LA PREUVE DE H. BRAY

La démonstration de H. Bray est complètement originale et emploie une stratégie inédite dans l'étude de la masse. Son point de départ s'inspire néanmoins des premiers travaux de R. Schoen et S.T. Yau [44].

LEMME 4.1. — Pour toute variété asymptotiquement plate (M, g) avec un horizon Σ , il existe une suite de métriques g_σ à courbure scalaire nulle, asymptotiquement plates et *globalement conformes* à la métrique euclidienne en-dehors de la boule $B(\sigma)$, qui tend (uniformément) vers g lorsque σ tend vers l'infini. De plus $m(g_\sigma) - m(g) \leq C\sigma^{-1}$ (où $C > 0$) et chaque métrique g_σ possède un horizon Σ_σ extérieur à Σ et dont l'aire tend vers $\text{Aire}_g(\Sigma)$ lorsque σ tend vers l'infini.

4.1. Le flot conforme

Le lemme 4.1 montre qu'il suffit de démontrer l'inégalité de Penrose dans la classe des métriques asymptotiquement plates et *harmoniquement plates*, c'est-à-dire à courbure scalaire nulle et globalement conforme à l'espace euclidien en-dehors d'un compact.

Nous nous plaçons désormais dans ce cadre et, pour fixer les notations, nous notons $g(x) = \mathcal{U}^4(x)$ eucl pour tout x de $M_e - K$, où M_e est toujours la région extérieure au sens de la section 3.6, de bord Σ (non nécessairement connexe) et où la fonction \mathcal{U} est harmonique ($\text{Scal}^g = 0$) et tend vers 1 à l'infini. On se trouve donc ramené à une situation semblable aux métriques de Schwarzschild (qui peuvent donc guider l'intuition plus efficacement que dans le cas général) et où l'essentiel de l'information (au moins au voisinage de l'infini) est contenu dans une fonction. Partant d'une métrique harmoniquement plate, il est alors naturel de chercher à diminuer le quotient de Penrose en restant à l'intérieur de cette classe de métriques.

PROPOSITION 4.2. — *Il existe une fonction $(t, x) \mapsto v_t(x)$ sur $\mathbf{R} \times M_e$ et des hypersurfaces lisses (non nécessairement connexes) $\Sigma(t)$ telles que*

$$(4.1) \quad \Delta_g v_t = 0 \text{ à l'extérieur de } \Sigma(t), \quad v_t = 0 \text{ à l'intérieur de } \Sigma(t), \quad \lim_{t \rightarrow \infty} v_t = -e^{-t}$$

et de plus $\Sigma(t)$ est la surface minimale la plus extérieure pour

$$(4.2) \quad g_t = u_t^4 g = \left(1 + \int_0^t v_t dt\right)^4 g.$$

La fonction v est lipschitzienne en t et lisse en x sauf sur $\Sigma(t)$ où elle est C^1 .

Le cas d'égalité est assez éclairant et mérite donc d'être détaillé. Partant de la métrique de Schwarzschild $g^S = (1 + \frac{m}{2r})^4$ eucl de masse m , le flot conforme défini par H. Bray produit la famille $g_t^S = \mathcal{U}_t^4$ eucl avec

$$(4.3) \quad \mathcal{U}_t(x) = \begin{cases} e^{-t} + \frac{me^{2t}}{2r} & \text{si } r > \frac{me^{2t}}{2}, \\ \sqrt{\frac{2m}{r}} & \text{si } r \leq \frac{me^{2t}}{2}. \end{cases}$$

Il s'agit donc d'une famille de métriques de Schwarzschild de masse constante m , donc toutes isométriques, mais définies à l'extérieur d'horizons $S(g_t) = \{r = \frac{me^{2t}}{2}\}$ qui "filent à l'infini". La formulation générale du flot est obtenue en recherchant des facteurs conformes qui vérifient dans la métrique initiale g les mêmes équations que celles vérifiées par les fonctions (4.3) dans le modèle de Schwarzschild g^S .

Les conditions portant sur g_t doivent être comprises comme analogues à une *équation différentielle* : la fonction v_t est la dérivée de u_t (le facteur conforme apparaissant dans la définition de g_t) et est déterminée par les équations (4.1). L'existence du flot et ses propriétés de régularité sont d'ailleurs obtenues par un procédé de discrétisation fort semblable à la méthode d'Euler pour les équations différentielles ordinaires.

LEMME 4.3. — L'aire (totale) de $\Sigma(t)$ est constante, égale à Aire(Σ).

Esquisse (très rapide) de preuve. – L'horizon initial Σ est de courbure moyenne strictement négative pour la normale extérieure dans g_t dès que $t > 0$. Dès lors, $\Sigma(t)$, surface minimale la plus extérieure pour g_t , est à l'extérieur de Σ et Aire($\Sigma(t)$) \leq Aire(Σ).

L'inégalité dans l'autre sens est nettement plus délicate et repose sur une étude fine du comportement des approximations qui sont utilisées dans la méthode d'Euler. Pour un pas $\varepsilon > 0$ petit, les approximations g_t^ε sont en effet suffisamment explicites pour que l'on puisse contrôler l'aire de leurs surfaces minimales les plus extérieures et surtout leur taux de variation. On en déduit alors que la surface minimale la plus extérieure pour g_t^ε est d'aire égale à Aire(Σ) + $o(1)$ lorsque le pas ε tend vers 0. \square

4.2. Masse et fonctions de Green

La partie la plus originale du travail de H. Bray est celle consacrée au comportement de la masse lors de l'évolution. Elle fait intervenir, de manière assez inattendue, le théorème de la masse positive.

Il met en évidence une relation (très générale) entre la masse de la variété et un objet classique de théorie du potentiel, la capacité de l'horizon, définie par

$$(4.4) \quad \mathcal{E}(\Sigma, g) = \frac{1}{2\pi} \inf_{\mathcal{D}} \int_{M_e} |d\psi|^2 \quad \text{où } \mathcal{D} = \{\psi \in C_c^\infty(M), \psi = 1 \text{ sur } \Sigma\}.$$

La borne inférieure est bien sûre atteinte pour une fonction ψ_0 harmonique hors de Σ , nulle à l'infini et valant 1 sur l'horizon. Si la métrique est asymptotiquement plate, une

telle fonction s'écrit toujours $\psi_0 = \frac{c}{2|x|} + O(|x|^{-2})$ (voir [4]) et la formule de Green assure que $\mathcal{E}(\Sigma, g) = c$.

THÉORÈME 4.4. — *Dans les conditions précédentes, $m(g) \geq \frac{1}{2} \mathcal{E}(\Sigma, g)$.*

Esquisse de preuve. — Soit ψ_0 une fonction minimisante pour la capacité. On considère alors le double $\bar{M} = M_e \cup (-M_e)$ (obtenu par réflexion le long de Σ) muni de la fonction harmonique φ_0 définie par $\varphi_0(x) = 1 - \frac{\psi_0(x)}{2}$ si x est dans M_e et $\varphi_0(x) = \frac{\psi_0(-x)}{2}$ si x est dans $(-M_e)$, et de la métrique $\bar{g} = \varphi_0^4 g$. La variété \bar{M} semble avoir deux bouts mais, en réalité, il n'en est rien : compte-tenu du comportement asymptotique de la fonction ψ_0 , l'infini de $(-M_e)$ se comporte comme un voisinage épointé de l'origine dans \mathbf{R}^3 ; la métrique s'y étend de manière C^∞ car \bar{g} est de la forme \mathcal{V}^4 eucl avec \mathcal{V} harmonique (théorème d'élimination des singularités). Un calcul direct montre de plus que la masse de (\bar{M}, \bar{g}) (en l'infini de M_e , qui est le seul vrai bout de \bar{M}) vaut $m(g) - \frac{1}{2} \mathcal{E}(\Sigma, g)$. Si \bar{g} était lisse sur Σ , elle serait redevable du théorème de la masse positive puisqu'elle est à courbure scalaire positive. Dans le cas présent, il faut régulariser astucieusement la métrique au voisinage de Σ et appliquer une version faible de la masse positive démontrée par H. Bray et F. Finster [10]. \square

LEMME 4.5. — La masse $m(g_t)$ est une fonction décroissante en t .

Preuve. — Le calcul de la masse est effectué à l'infini : l'hypothèse de platitude harmonique assure donc qu'à l'extérieur d'un compact $g_t = u_t^4 g_0 = (u_t \mathcal{U})^4$ eucl. De plus la fonction $u_t = 1 + \int_0^t v_t dt$ est dérivable presque partout car monotone.

Imaginons (pour simplifier) que ce soit le cas à $t = 0$. Comme $1 + e^t v_t$ est harmonique, tend vers 0 à l'infini et vaut 1 sur l'horizon $\Sigma(t)$, $v_t = e^{-t} \left(-1 + \frac{\mathcal{E}(\Sigma(t), g_t)}{2|x|} \right) + O(|x|^{-2})$ et, de manière analogue, $\mathcal{U}(x) = \left(1 + \frac{m(g)}{2|x|} \right) + O(|x|^{-2})$. La masse $m(t)$ se déduit directement du produit $u_t \mathcal{U}$ et on calcule $m'(0) = \mathcal{E}(\Sigma, g) - 2m(0)$, qui est négatif d'après le Théorème 4.4. La situation à $t > 0$ se traite exactement de la même façon. \square

La relation entre masse et capacité n'est pas réellement nouvelle mais elle n'apparaissait jusque-là qu'en filigrane dans plusieurs travaux antérieurs [12]. En l'isolant du contexte, le travail de H. Bray souligne le rôle central du Théorème 4.4 et fournit par là des preuves beaucoup plus conceptuelles à des résultats plus anciens.

4.3. Comportement asymptotique du flot et conclusion

Il reste à comprendre le comportement des métriques g_t lorsque t devient grand. Pris un à un, les arguments employés par H. Bray ne sont pas tous difficiles, mais impliquent des estimées souvent longues, techniques, calculatoires, et l'ensemble demande une grande virtuosité. Tout le travail est néanmoins facilité par l'hypothèse de platitude harmonique : à partir d'un certain moment, tout se passe à l'extérieur d'une boule de \mathbf{R}^3 et l'étude

des métriques g_t s'y résume à l'étude des fonctions harmoniques $\mathcal{U}_t = u_t \mathcal{U}$. Nous nous contentons de décrire succinctement les différentes étapes.

(a) *les surfaces $\Sigma(t)$ sont comprises entre les deux sphères $S(ae^{2t})$ et $S(be^{2t})$.* Ce point est la partie la plus difficile du travail de H. Bray. Très schématiquement, il est acquis en montrant que toute autre situation conduirait à ce que l'aire de $\Sigma(t)$ ne soit pas égale à celle de Σ , en contradiction avec le Lemme 4.3. Interviennent dans la preuve de longues estimations explicites de l'aire de surfaces dans l'espace euclidien, en fonction de l'intégrale de Willmore $\int_{\Sigma(t)} H^2$. Tout ceci utilise profondément l'invariance conforme de l'intégrale de Willmore et donc l'hypothèse de platitude harmonique. En procédant avec un peu de soin, on obtient un contrôle explicite des constantes a et b .

(b) *contraction des coordonnées* : compte-tenu de **(a)**, il est commode de rééchelonner toute la situation d'un facteur e^{2t} . On dispose donc de deux familles de métriques :

$$(4.5) \quad \begin{aligned} G_t &= (e^t \mathcal{U}_t(xe^{2t}))^4 \text{ eucl} = U_t^4 \text{ eucl} \quad \text{sur } \mathbf{R}^3 - B(e^{-2t}) \\ \tilde{G}_t &= W_t^4 \text{ eucl} = (\varphi_0 U_t)^4 \text{ eucl} \quad \text{sur } \mathbf{R}^3, \end{aligned}$$

où φ_0 est définie comme dans la preuve du Théorème 4.4. Elles sont asymptotiquement plates, et la première est bien sûr isométrique au bout harmoniquement plat de (M_e, g_t) .

D'après le Lemme 4.5, leurs masses (avec des notations évidentes) sont donc reliées par $m'(t) = -2\tilde{m}(t)$. Compte-tenu du théorème de la masse positive et de sa décroissance, $m(t)$ converge, lorsque t tend vers l'infini, vers un $M \geq 0$ et $\int^\infty \tilde{m}(t) < \infty$.

(c) La fin de la preuve repose sur une *étude précise de l'espace \mathcal{H} des métriques harmoniquement plates* en dehors d'une boule fixée de \mathbf{R}^3 . Cette étude a un intérêt qui dépasse largement le cadre strict de l'inégalité de Penrose. Il n'est pas impossible qu'elle permette de préciser la structure de *l'espace des modules des métriques asymptotiquement plates*. Une clef de la démonstration est fournie par la

PROPOSITION 4.6. — *La masse est continue sur \mathcal{H} . De plus, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, si $m(U^4 \text{ eucl}) \leq \delta$, alors $|U - 1|$ est uniformément bornée par ε .*

On en déduit (presque) immédiatement que \tilde{G} tend uniformément vers la métrique euclidienne. Les fonctions U_t et W_t n'étant pas indépendantes (leurs valeurs sont gouvernées, à l'origine, par le choix du flot conforme), la suite de métriques $\{G_t\}$ tend vers une métrique de Schwarzschild : la suite de fonctions U_t tend uniformément vers $1 + \frac{m}{2r}$.

(d) La preuve est achevée si l'on sait montrer que *les surfaces $\Sigma(t)$ convergent vers l'horizon de Schwarzschild* (au sens, bien sûr, de la théorie géométrique de la mesure). Ce point découle en fait du **(c)**, car une connaissance suffisamment fine des fonctions U_t suffit, avec un peu de travail, à contrôler le comportement des horizons $\Sigma(t)$.

Contrairement à la preuve de G. Huisken et T. Ilmanen, celle de H. Bray ne suppose à aucun moment que l'horizon d'origine Σ est connexe. L'énoncé vaut donc pour toutes

les régions extérieures décrites à la section 3.6, ce qui constitue la version la plus générale connue à ce jour de l'inégalité de Penrose :

THÉORÈME 4.7. — *Soit (M, g) une variété asymptotiquement plate de dimension 3 à courbure scalaire positive, ayant pour bord (intérieur) une famille finie Σ de sphères minimales et ne contenant pas d'autres surfaces minimales. Alors*

$$(4.6) \quad m(M, g) \geq \frac{1}{4} \sqrt{\frac{\text{Aire}(\Sigma)}{\pi}}.$$

En cas d'égalité, Σ est connexe et $M = (\mathbb{R}^3 \setminus B_0(\frac{m}{2}), (1 + \frac{m}{2r})^4 \text{eucl})$.

5. QUELQUES IDÉES POUR ALLER PLUS LOIN...

Pour les spécialistes des aspects mathématiques de la relativité générale, la preuve de l'inégalité de Penrose ne constitue qu'une première étape dans un programme beaucoup plus vaste. Il reste en effet à comprendre le cas où la variété M est une hypersurface asymptotiquement plate arbitraire (non nécessairement totalement géodésique) d'un espace-temps lorentzien. Tel est en effet le cadre originel du travail de R. Penrose, passage obligé lorsque l'on s'intéresse à la conjecture de la censure cosmique déjà évoquée. La masse doit alors être complétée par trois autres quantités $\{p_1, p_2, p_3\}$ dépendant de la géométrie extrinsèque de l'hypersurface et l'horizon n'est plus une collection de surfaces minimales mais de surfaces *piégées* [41, page 435]. Sous les conditions d'énergie les plus générales, l'ensemble (m, p_1, p_2, p_3) , qui est habituellement vu comme un vecteur de \mathbb{R}^4 muni de la métrique de Minkowski et porte le nom de *vecteur de moment-énergie*, vérifie une généralisation de la masse positive [46, 52] :

$$m^2 - (p_1)^2 - (p_2)^2 - (p_3)^2 \geq 0.$$

Il est difficile de citer ici une conjecture précise et la forme mathématique exacte de ce que devrait être l'inégalité de Penrose dans ce cadre général est encore un sujet de débat. Avant de passer à la situation lorentzienne, une étape intermédiaire possible est d'étudier la situation *asymptotiquement hyperbolique* où l'espace euclidien modèle est remplacé par l'espace hyperbolique de dimension 3, qui est une autre hypersurface naturelle de l'espace de Minkowski. Cet objectif ne paraît pas déraisonnable à l'heure actuelle et pourrait être atteint avec les techniques de G. Huisken et T. Ilmanen.

Les résultats et les techniques de H. Bray, G. Huisken et T. Ilmanen permettent aussi d'ouvrir quelques perspectives pour un autre problème de grande ampleur, tant en relativité qu'en géométrie des variétés asymptotiquement plates. Comme nous l'avons signalé, il n'existe pas de notion locale de masse en relativité. Une question majeure de la recherche actuelle est donc de définir une quantité "quasi-locale" de masse. La proposition la plus géométrique est celle de R. Bartnik [5] qui suggère de définir la masse d'un ouvert Ω de

M comme l'infimum des masses des variétés asymptotiquement plates à courbure scalaire positive qui contiennent Ω isométriquement. Cette définition séduisante manquait jusqu'à présent de bases solides : ainsi, il n'était nullement assuré qu'un ouvert non localement isométrique à \mathbf{R}^3 avait une masse de Bartnik strictement positive. Les outils développés par G. Huisken et T. Ilmanen permettent de répondre affirmativement à cette question. Ils précisent également que les masses de Bartnik d'une exhaustion Ω_i d'une variété M asymptotiquement plate à courbure scalaire positive convergent vers la masse (usuelle) de M [26]. De leur côté, les techniques de H. Bray laissent entrevoir des liens intéressants entre la masse de Bartnik (ou d'autres masses quasi-locales) et la théorie du potentiel [9], dans l'esprit du Théorème 4.4.

Tous ces résultats permettent d'envisager maintenant une étude plus approfondie de la masse de Bartnik, ou d'autres masses quasi-locales, qui pourraient être des outils puissants pour une meilleure compréhension de la géométrie des variétés asymptotiquement plates.

Remerciements. Ce texte n'aurait pu exister sans de nombreuses conversations avec Tom Ilmanen et Hubert Bray, ni sans la lecture assidue et ô combien précieuse qu'en a faite Gilles Carron. J'ai aussi apprécié les remarques utiles de Jean-Pierre Bourguignon, Piotr Chruściel, Hélène Davaux, Zindine Djadli, Paul Gauduchon et Jacques Lafontaine. Qu'ils en soient tous chaleureusement remerciés.

BIBLIOGRAPHIE

- [1] R. ARNOWITT, S. DESER et C. W. MISNER – “Coordinate invariance and energy expressions in General Relativity”, *Phys. Rev.* **122** (1961), p. 997–1006.
- [2] S. BANDO, A. KASUE et H. NAKAJIMA – “On a construction of coordinates at infinity on manifolds with fast curvature decay and maximal volume growth”, *Invent. math.* **97** (1989), p. 313–349.
- [3] C. BÄR – “Lower eigenvalues estimates for Dirac operators”, *Math. Ann.* **293** (1992), p. 39–46.
- [4] R. BARTNIK – “The mass of an asymptotically flat manifold”, *Commun. Pure. Appl. Math.* **39** (1986), p. 661–693.
- [5] ———, “New definition of quasi-local mass”, *Phys. Rev. Lett.* **62** (1989), p. 2346–2348.
- [6] ———, “Quasi-spherical metrics and prescribed scalar curvature”, *J. Diff. Geom.* **37** (1993), p. 31–71.
- [7] R. BISHOP et R. CRITTENDEN – *Geometry of manifolds*, Pure Appl. Math., vol. XV, Acad. Press, New York, 1964.
- [8] J.-P. BOURGUIGNON – “Stabilité par déformation non-linéaire de la métrique de Minkowski”, Sémin. Bourbaki n° 740, Astérisque, vol. 201-203, Soc. math. France, 1991.
- [9] H. BRAY – “Proof of the Riemannian Penrose conjecture using the positive mass theorem”, *J. Diff. Geom.*, p. à paraître.
- [10] H. BRAY et F. FINSTER – “Curvature estimates and the positive mass theorem”, *preprint* (1998).
- [11] H. BRAY et R. SCHOEN – “Recent proofs of the Riemannian Penrose inequality”, *Current Developments in Mathematics*, Harvard Univ., 1999.
- [12] G. L. BUNTING et A. K. MASOOD-UL-ALAM – “Non-existence of multiple black holes in asymptotically euclidean vacuum spacetimes”, *Gen. Rel. Grav.* **19** (19), p. 147–154.

- [13] Y. CHOQUET-BRUHAT – “Positive energy theorems”, *Relativity, groups and topology II*, Les Houches XL, 1983 (B. De Witt et R. Stora, éd.), Elsevier, Amsterdam, 1984, p. 740–785.
- [14] P. T. CHRUSCIEL – “Boundary conditions at spatial infinity from a Hamiltonian point of view”, *Topological properties and global structure of spacetime*, Erice 1985 (P. Bergmann et V. de Sabbata, éd.), Plenum, New York, 1986, p. 49–59.
- [15] ———, “On the invariant mass conjecture in General Relativity”, *Commun. Math. Phys.* **120** (1988), p. 233–248.
- [16] V. DENISOV et O. SOLOVEV – “The energy determined in General Relativity on the basis of the traditional Hamiltonian approach does not have physical meaning”, *Theor. Math. Phys.* **56** (1983), p. 832–838, english translation.
- [17] L. C. EVANS et J. SPRUCK – “Motion of level sets by mean curvature”, *J. Diff. Geom.* **33** (1991), p. 635–681.
- [18] R. GEROCH – “Energy extraction”, *Ann. N. Y. Acad. Sci.* **224** (1973), p. 108–117.
- [19] ———, “General Relativity”, in *Differential geometry*, Proc. Symp. Pure Math., vol. 27, Amer. Math. Soc., Providence, 1975.
- [20] R. HARDT et X.ZHOU – “An evolution problem for linear growth functionals”, *Comm. Partial Differential Equations* **19** (1994), p. 1879–1907.
- [21] S. HAWKING – “Gravitational radiation in an expanding universe”, *J. Math. Phys.* **9** (1968), p. 598–604.
- [22] S. W. HAWKING et G. ELLIS – *The large-scale structure of space-time*, Cambridge Univ. Press, Cambridge, 1973.
- [23] M. HERZLICH – “Compactification conforme des variétés asymptotiquement plates”, *Bull. Soc. math. France* **125** (1997), p. 55–92.
- [24] ———, “A Penrose-like inequality for the mass of Riemannian asymptotically flat manifolds”, *Commun. Math. Phys.* **188** (1997), p. 121–133.
- [25] O. HIJAZI – “Première valeur propre de l’opérateur de Dirac et nombre de Yamabe”, *C. R. Acad. Sci. Paris* **313** (1991), p. 865–868.
- [26] G. HUISKEN et T. ILMANEN – “The inverse mean curvature flow and the Riemannian Penrose inequality”, *J. Diff. Geom.*, à paraître.
- [27] ———, “Proof of the Penrose inequality”, *Int. Math. Res. Not.* **20** (1997), p. 1045–1058, annonce.
- [28] T. ILMANEN – *Elliptic regularization and partial regularity for motion by mean curvature*, *Memoirs Amer. Math. Soc.*, vol. 520, Amer. Math. Soc., Providence, RI, 1994.
- [29] P. S. JANG et R. WALD – “The positive energy conjecture and the cosmic censor hypothesis”, *J. Math. Phys.* **18** (1977), p. 41–44.
- [30] J. KAZDAN – “Positive energy in General Relativity”, *Sém. Bourbaki n° 593*, *Astérisque*, vol. 92-93, Soc. math. France, 1982.
- [31] M. KRUSKAL – “Maximal extension of the Schwarzschild metric”, *Phys. Rev.* **119** (1960), p. 1743–1745.
- [32] J. LEE et T. H. PARKER – “The Yamabe problem”, *Bull. Amer. Math. Soc.* **17** (1987), p. 37–91.
- [33] A. LICHNEROWICZ – *Théories relativistes de la gravitation et de l’électromagnétisme*, Masson, Paris, 1955.
- [34] ———, “Spineurs harmoniques”, *C. R. Acad. Sci. Paris* **257** (1963), p. 7–9.
- [35] A. LICHNEWSKY et R. TEMAM – “Pseudosolutions of the time-dependant minimal surface problem”, *J. Diff. Eq.* **30** (1978), p. 340–363.

- [36] R. B. LOCKHART – “Fredholm properties of a class of elliptic operators on non-compact manifolds”, *Duke Math. J.* **48** (1983), p. 289–312.
- [37] R. B. LOCKHART et R. MCOWEN – “Elliptic differential operators on non-compact manifolds”, *Ann. Scuola. Norm. Sup. Pisa* **12** (1985), p. 409–447.
- [38] S. LUCKHAUS – “Solutions for the two-phase Stefan problem with the Gibbs-Thompson law for the melting temperature”, *Eur. J. Appl. Math.* **1** (1990), p. 101–111.
- [39] E. MALEC et N. O’MURCHADHA – “Trapped surfaces and the Penrose inequality in spherically symmetric geometries”, *Phys. Rev. D* **49** (1994), p. 6931–6934.
- [40] U. MASSARI et M. MIRANDA – *Minimal surfaces of codimension one*, North-Holland Math. Stud., vol. 91, Elsevier, 1984.
- [41] B. O’NEILL – *Semi-Riemannian geometry*, Acad. Press, San Diego, 1983.
- [42] R. PENROSE – “Conformal treatment of infinity”, *Relativity, groups and topology* (C. de Witt et B. de Witt, édés.), Ecole d’été de Physique Théorique, Les Houches 1963, Gordon and Breach, 1963, p. 563–584.
- [43] ———, “Naked singularities”, *Ann. N. Y. Acad. Sci.* **224** (1973), p. 125–134.
- [44] R. SCHOEN et S. T. YAU – “On the proof of the positive mass conjecture in General Relativity”, *Commun. Math. Phys.* **65** (1979), p. 45–76.
- [45] ———, “On the structure of manifolds with positive scalar curvature”, *Manuscripta math.* **28** (1979), p. 159–183.
- [46] ———, “The energy and linear-momentum of spacetimes in general relativity”, *Commun. Math. Phys.* **79** (1981), p. 47–51.
- [47] K. SCHWARZSCHILD – “Über das Gravitationsfeld eines Massen nach der Einsteinschen Theorie”, *Sitz. König. Preuss. Akad. Wiss.* (1916), p. 189–196.
- [48] L. SIMON – *Lectures on geometric measure theory*, Proc. Centre Math. Anal., vol. 3, Austr. Nat. Univ., 1983.
- [49] P. STERNBERG, G. WILLIAMS et W. P. ZIEMER – “Regularity of constrained area minimizing hypersurfaces”, *J. Diff. Eq.* **94** (1991), p. 83–94.
- [50] A. VISINTIN – “Nucleation and mean curvature flow”, *Comm. Partial Differential Equations* **23** (1998), p. 17–53.
- [51] R. WALD – *General Relativity*, Univ. Chicago Press, Chicago, 1984.
- [52] E. WITTEN – “A new proof of the positive energy theorem”, *Commun. Math. Phys.* **80** (1981), p. 381–402.

Marc HERZLICH

Département de Mathématiques

UMR 5030 du CNRS

Université Montpellier II

F-34095 MONTPELLIER Cedex 5

E-mail : herzlich@math.univ-montp2.fr

FINITE GROUP ACTIONS ON ACYCLIC 2-COMPLEXES

by Alejandro ADEM

1. A BRIEF HISTORY AND MOTIVATION

A simple consequence of the Brouwer fixed point theorem is that any cyclic group acting on a closed disk \mathbb{D}^n must have a *fixed point*. The classical work of P.A. Smith [18] shows that if P is a finite p -group, then any action of P on \mathbb{D}^n must have a fixed point. From this there arises a very evident question: is there a group of *composite order* which can act on some \mathbb{D}^n *without* any fixed points? This was settled in the affirmative by Floyd and Richardson in 1959 (see [7]), when they constructed fixed point free actions of the alternating group A_5 on disks.

These examples stood out as special exceptions for several years – indeed no other such actions were known to exist until Oliver (see [13]) obtained a complete characterization of those finite groups which can act on disks without stationary points. To explain it we first need to introduce some group-theoretic concepts.

DEFINITION 1.1. — For p and q primes, let \mathcal{G}_p^q be the class of finite groups G with normal subgroups $P \triangleleft H \triangleleft G$, such that P is of p -power order, G/H is of q -power order, and H/P is cyclic; and let $\mathcal{G}_p = \cup_q \mathcal{G}_p^q$, $\mathcal{G} = \cup_p \mathcal{G}_p$.

THEOREM 1.2. — *A finite group G has a fixed point free action on a disk if and only if $G \notin \mathcal{G}$. In particular, any non-solvable group has a fixed point free action on a disk, and an abelian group has such an action if and only if it has three or more non-cyclic Sylow subgroups.*

The smallest group with a fixed point free action on a disk is in fact the alternating group A_5 ; the smallest *abelian* group with such an action is $C_{30} \times C_{30}$. Oliver also proved that a group G will have a fixed point free action on a finite \mathbb{Z}_p -acyclic¹ complex if and

¹Recall that a complex X is said to be \mathbb{Z}_p -acyclic if its reduced mod p homology is identically zero; if its reduced integral homology vanishes it is said to be *acyclic*.

only if $G \notin \mathcal{G}_p$. Note that a group G will act without fixed points on a contractible complex if and only if it acts without fixed points on an *acyclic* complex.

Taking into account Oliver's result, an obvious problem is that of constructing fixed point free actions on contractible or acyclic complexes of small dimension. A well-known theorem by J.-P. Serre states that any finite group acting on a tree must have a fixed point (see [16]). However, the situation for contractible 2-dimensional complexes is much more complicated – in fact it is an open question whether or not it is possible for a finite group to act on such a complex without fixed points. We will restrict our attention from now on to the case of acyclic 2-dimensional complexes.

Our starting point is the classical example of an A_5 -action on an acyclic 2-dimensional complex without fixed points, which we now briefly recall. In fact it is an essential ingredient in the construction due to Floyd and Richardson which we discussed above. This example is constructed by considering the left A_5 action on the Poincaré sphere $\Sigma^3 = SO(3)/A_5$; as the action has a single fixed point (corresponding to the fact that A_5 is self-normalizing in $SO(3)$) we may remove an open 3-disk U around it to obtain an acyclic compact 3-manifold $\Sigma^3 - U$ with a fixed point free action of A_5 . This in turn can be collapsed to a 2-dimensional subcomplex $X \simeq \Sigma^3 - U$ upon which A_5 still acts without fixed points. Equivalently we could identify Σ^3 with the space obtained by identifying opposite faces of the solid dodecahedron in an appropriate way² and consider the A_5 action induced by the usual action on the dodecahedron. The fixed point is the center of D and by collapsing to its boundary we obtain an explicit 2-dimensional complex $X = \partial D / \simeq$ with a fixed point free action of A_5 which has 6 pentagonal 2-cells, 10 edges and 5 vertices. Note that if we take the join $A = A_5 * X$ with the induced diagonal action of A_5 , then we obtain a simply connected and acyclic complex, hence a contractible complex with a fixed point free action. From this we can obtain a fixed point free A_5 action on a disk via regular neighborhoods (as explained in [4], p. 57). This is the basic step in the construction of the Floyd–Richardson examples.

Now an obvious question arises from all of this: can we characterize those finite groups which can act without fixed points on acyclic 2-dimensional complexes? Indeed, are there even other examples of such actions? Remarkably it turns out that these actions are only possible for a small class of *simple* groups, and their precise determination and description will require using the classification of finite simple groups.

²To be precise: identify opposite faces of the dodecahedron by the map which pushes each face through the dodecahedron and twists it by $\frac{2\pi}{10}$ about the axis of the push in the direction of a right hand screw (see [12]).

2. STATEMENT OF RESULTS

In this note we will report on recent work of Oliver and Segev (see [15]) where they provide a complete description of the finite groups which can act on a 2-dimensional acyclic complex without fixed points. Their work builds on previous contributions by Oliver [13], [14], Segev [17] and Aschbacher–Segev [2]. To state their main result we need to introduce a useful technical condition for G –CW complexes. From now on we will use the term G –complex to refer to a G –CW–complex, however these results also hold for simplicial complexes with an admissible G –action³.

DEFINITION 2.1. — A G –complex X is said to be *essential* if there is no normal subgroup $1 \neq N \triangleleft G$ with the property that for each $H \subseteq G$, the inclusion $X^{HN} \rightarrow X^H$ induces an isomorphism on integral homology.

If there were such a normal subgroup N , then the G –action on X is ‘essentially’ the same as the G –action on X^N , which factors through a G/N –action. For 2–dimensional complexes we have:

THEOREM 2.2. — *Let G be any finite group and let X be any 2–dimensional acyclic G –complex. Let N denote the subgroup generated by all normal subgroups $N' \triangleleft G$ such that $X^{N'} \neq \emptyset$. Then X^N is acyclic, X is essential if and only if $N = 1$, and if $N \neq 1$ then the action of G/N on X^N is essential.*

Based on this we restrict our attention to essential complexes, and we can now state the main result in [15]:

THEOREM 2.3. — *Given a finite group G , there is an essential fixed point free 2–dimensional acyclic G –complex if and only if G is isomorphic to one of the simple groups $PSL_2(2^k)$ for $k \geq 2$, $PSL_2(q)$ for $q \equiv \pm 3 \pmod{8}$ and $q \geq 5$ or $Sz(2^k)$ for odd $k \geq 3$. Furthermore the isotropy subgroups of any such G –complex are all solvable.*

Among the groups listed above, only the *Suzuki groups* $Sz(q)$ are not commonly known; we will provide a precise definition for them as subgroups of $GL_4(\mathbb{F}_q)$ in §5. Note that the theorem is stated for arbitrary acyclic 2–dimensional complexes; there is no need to require that the complexes be finite.

Our main goal will be to explain the proof of this result. This naturally breaks up into a number of different steps. We begin in §3 by explaining how the theorem can be reduced to simple groups, based mostly on a theorem due to Segev [17]. Next in §4 we describe techniques for constructing the desired actions, using methods derived from Oliver’s original work on group actions on acyclic complexes as well as a more detailed analysis of the associated subgroup lattices. This is then applied in §5 to provide explicit

³A simplicial complex X with a G action is called admissible if the action permutes the simplices linearly and sends a simplex to itself only via the identity.

descriptions of fixed point free actions on an acyclic 2–complex for the simple groups listed in the main theorem. In §6 we sketch conditions which imply the *non–existence* of fixed point free actions on acyclic 2–complexes for most simple groups; this part requires detailed information about the intricate subgroup structure for the finite simple groups. Finally in §7 we use the classification of finite simple groups and the previous results to outline the proof of the main theorem, which has been previously reduced to verification for simple groups. We also make a few concluding remarks.

REMARK 2.4. — The background required to understand these results and their proofs includes: (1) very basic equivariant algebraic topology; (2) familiarity with subgroup complexes and related constructions; and (3) a very detailed knowledge of the subgroup structure of the finite simple groups. As many of the arguments in the proofs depend on the particular properties of these groups, our synopsis cannot hope to contain complete details. However the original paper by Oliver and Segev [15] is written in a clear style accessible to a broad range of mathematicians and hence those interested in a deeper understanding of the results presented here should consult it directly.

3. REDUCTION TO SIMPLE GROUPS

The goal of this section will be to explain how we can restrict our attention to finite simple groups. This is based on the following key result due to Segev [17] :

THEOREM 3.1. — *Let X be any 2–dimensional acyclic G –complex. Then the subcomplex of fixed points X^G is either acyclic or empty. If G is solvable then X^G is acyclic.*

Proof. Although Segev’s original proof uses the Odd Order Theorem, it can be proved more directly. One can show that if X is an acyclic G –complex, then $H_1(X^G, \mathbb{Z}) = H_2(X^G, \mathbb{Z}) = 0$. Hence we are reduced to establishing that there is only one connected component (provided X^G is non–empty). For solvable groups this can be proved directly using induction and Smith Theory. Otherwise we consider a minimal group G for which a counterexample exists. If X^G has k components then in fact it can be shown that X looks roughly like the join of an acyclic fixed point free G –complex Y with a set of k points. However as X is 2–dimensional, Y would have to be 1–dimensional, in other words a tree, and this cannot hold. \square

REMARK 3.2. — The reader should keep in mind that Theorem 3.1 is a basic tool in many of our subsequent arguments and it will be used explicitly and implicitly on several occasions.

COROLLARY 3.3. — *Let X be any 2–dimensional acyclic G –complex. Assume that $A, B \subset X$ are G –invariant acyclic subcomplexes such that $X^G \subset A \cup B$; then $A \cap B \neq \emptyset$.*

Proof. Assume that $A \cap B = \emptyset$ and let Z denote the G -complex obtained by identifying A and B each to a point. Then Z is acyclic since A and B both are, and Z^G consists of two points, thus contradicting Theorem 3.1. \square

As an immediate consequence of Corollary 3.3 we obtain

LEMMA 3.4. — *Let X be a 2-dimensional acyclic G -complex. Then if $H, K \subset G$ are such that $H \subset N_G(K)$ and X^H, X^K are both non-empty, then $X^{HK} \neq \emptyset$. Moreover, if $H \subset G$ is such that $X^H = \emptyset$, then $X^{C_G(H)} \neq \emptyset$.*

Proof. Since H normalizes K , both X^H and X^K are H -invariant acyclic subcomplexes of X . Hence we conclude from Corollary 3.3 that $\emptyset \neq X^H \cap X^K = X^{HK}$. For the second part, it suffices to prove it when H is minimal among subgroups without fixed points. Fix a pair $M, M' \subset H$ of distinct maximal subgroups (note that by Theorem 3.1, H is non-solvable). Then X^M and $X^{M'}$ are non-empty, but $X^M \cap X^{M'} = X^{\langle M, M' \rangle} = X^H = \emptyset$. Hence X^M and $X^{M'}$ are disjoint $C_G(H)$ -invariant acyclic subcomplexes of X , meaning (by Corollary 3.3) that their union cannot contain $X^{C_G(H)}$, whence it must be non-empty. \square

We can now prove one of the main reduction results, which allows us to restrict our attention to essential complexes.

THEOREM 3.5. — *Let G be any finite group, and let X be any 2-dimensional acyclic G -complex. Let N be the subgroup generated by all normal subgroups $N' \triangleleft G$ such that $X^{N'} \neq \emptyset$. Then X^N is acyclic; X is essential if and only if $N = 1$ and if $N \neq 1$ then the action of G/N on X^N is essential.*

Proof. If $X^{N_1} \neq \emptyset$ and $X^{N_2} \neq \emptyset$ for $N_1, N_2 \triangleleft G$, then $X^{(N_1, N_2)} \neq \emptyset$ by Lemma 3.4. So we infer that X^N is non-empty, hence acyclic (by Theorem 3.1). Note that the action of any non-trivial normal subgroup of G/N on X^N has empty fixed point set, hence the action of G/N on X^N is always essential. Finally, assume that $N \neq 1$; by Theorem 3.1 we have that for all $H \subset G$, X^H and X^{NH} are acyclic or empty; and $X^{NH} \neq \emptyset$ if $X^H \neq \emptyset$, by Lemma 3.4. Hence the inclusion $X^{NH} \rightarrow X^H$ is always an equivalence of integral homology, and hence X is not essential. \square

This result will allow us to focus our attention on actions of simple groups.

THEOREM 3.6. — *If G is a non-trivial finite group for which there exists an essential 2-dimensional acyclic G -complex X , then G is almost simple. In fact there is a normal simple subgroup $L \triangleleft G$ such that $X^L = \emptyset$ and such that $C_G(L) = 1$.*

Proof. We know from Theorem 3.5 that $X^N = \emptyset$ for all normal subgroups $1 \neq N \triangleleft G$, including the case $N = G$. Now fix a minimal normal subgroup $1 \neq L \triangleleft G$; we know from Theorem 3.1 that L is not solvable, as $X^L = \emptyset$. Hence L is a direct product of isomorphic non-abelian simple groups (see [8], Thm 2.1.5). Assume that L is not simple; by Lemma 3.4, $X^H \neq \emptyset$ for some simple factor $H \triangleleft L$. Also, note that $L = \langle gHg^{-1} \mid g \in G \rangle$

since it is a minimal normal subgroup. Now we have that $X^{gHg^{-1}} = gX^H \neq \emptyset$ for all $g \in G$, hence applying the same lemma once again, but now to the L -action on X , we infer that $X^L \neq \emptyset$, a contradiction.

So L is simple; now set $H = C_G(L)$. Then we have that $H \triangleleft G$ (this follows from the fact that $L \triangleleft G$) and so $X^H \neq \emptyset$, by Lemma 3.4. However we have assumed that the action is essential, whence $H = 1$. \square

The condition $C_G(L) = 1$ is equivalent to $G \subseteq \text{Aut}(L)$. Using this proposition we can decide which groups admit essential fixed point free actions on acyclic 2-dimensional complexes by first determining the *simple groups* with such actions and then looking at automorphism groups only for that restricted collection.

As the proof of the main theorem will require explicit knowledge about the finite simple groups, it seems appropriate to briefly recall their classification, we refer to [9] for a detailed explanation. We should point out that it is by now common knowledge that complete details of the proof of the Classification Theorem were not available when it was announced in 1981; crucial work involving the so-called *quasithin groups* was never published and is known to contain gaps. Fortunately this has been resolved thanks to more recent work by Aschbacher and Smith and although a full account has not yet been published, a draft of their manuscript (over 1200 pages long!) is now available on the world wide web (see [3]).

The following theorem encapsulates our understanding of finite simple groups, and its proof requires literally thousands of pages of mathematical arguments by many authors.

THEOREM 3.7. — *Let L denote a non-abelian finite simple group, then it must be isomorphic to one of the following groups:*

- an alternating group A_n for $n \geq 5$
- a finite group of Lie type, i.e. a finite Chevalley group or a twisted analogue⁴
- one of the 26 sporadic simple groups.

4. TECHNIQUES FOR CONSTRUCTING ACTIONS

One of the main results in [15] is an explicit listing of conditions which imply the existence of fixed point free actions on acyclic complexes. We first introduce

DEFINITION 4.1. — A non-empty family⁵ \mathcal{F} of subgroups of a group G is said to be *separating* if it has the following three properties: (a) $G \notin \mathcal{F}$; (b) any subgroup of an element in \mathcal{F} is in \mathcal{F} ; and (c) for any $H \triangleleft K \subseteq G$ with K/H solvable, $K \in \mathcal{F}$ if $H \in \mathcal{F}$.

⁴We should mention that the Tits group ${}^2F_4(2)'$ is actually of index 2 in the full Lie type group ${}^2F_4(2)$.

⁵A family is a collection of subgroups of a group G which is closed under conjugation.

It is not hard to see that any maximal subgroup in a separating family of subgroups of G is self-normalizing. If G is solvable, then it has no separating family of subgroups. For G not solvable we let \mathcal{SLV} denote the family of solvable subgroups, which is the minimal separating family for G .

DEFINITION 4.2. — Given G and a family of subgroups \mathcal{F} , a (G, \mathcal{F}) -complex is a G -complex such that all of its isotropy subgroups lie in \mathcal{F} . It is said to be *universal* (respectively *H -universal*) if the fixed point set of each $K \in \mathcal{F}$ is contractible (respectively acyclic).

The following proposition relates the two previous concepts in our situation.

PROPOSITION 4.3. — *Let X denote a 2-dimensional acyclic G -complex without fixed points. Let $\mathcal{F} = \{H \subset G \mid X^H \neq \emptyset\}$. Then \mathcal{F} is a separating family of subgroups of G , and X is an H -universal (G, \mathcal{F}) -complex.*

Given a family of subgroups \mathcal{F} , let $N(\mathcal{F})$ denote the nerve of \mathcal{F} (regarded as a poset via inclusion) with a G -action induced by conjugation. Given any set \mathcal{H} of subgroups in G , we let $\mathcal{F}_{\geq \mathcal{H}}$ denote the poset of those subgroups in \mathcal{F} which contain some element of \mathcal{H} . For a single subgroup H we use the notation $\mathcal{F}_{\geq H}$ and $\mathcal{F}_{> H}$ to denote the posets of subgroups containing H or strictly containing H , respectively. We denote $X^{\mathcal{H}} = \cup_{H \in \mathcal{H}} X^H$.

The following are two key technical lemmas which will be required:

LEMMA 4.4. — *If X denotes a universal (H -universal) (G, \mathcal{F}) -complex then there exists a G -map $X \rightarrow N(\mathcal{F})$ which induces a homotopy equivalence (homology equivalence) between $X^{\mathcal{H}}$ and $N(\mathcal{F}_{\geq \mathcal{H}})$.*

LEMMA 4.5. — *Let \mathcal{F} be any family of subgroups of G , and let $\mathcal{F}_0 \subseteq \mathcal{F}$ be any subfamily such that $N(\mathcal{F}_{> H}) \simeq *$ for all $H \in \mathcal{F} - \mathcal{F}_0$. Then any $(H-)$ universal (G, \mathcal{F}_0) -complex is also an $(H-)$ universal (G, \mathcal{F}) -complex; and $N((\mathcal{F}_0)_{\geq \mathcal{H}}) \simeq N(\mathcal{F}_{\geq \mathcal{H}})$ for any set \mathcal{H} of subgroups of G .*

A complex Y is said to be homologically m -dimensional if $H_n(X, \mathbb{Z}) = 0$ for all $n > m$ and $H_m(X, \mathbb{Z})$ is \mathbb{Z} -free. For later use we observe that, for $m \geq 1$, if X is an m -dimensional acyclic complex, then any subcomplex of X is homologically $(m - 1)$ -dimensional and that the intersection of a finite number of homologically $(m - 2)$ -dimensional complexes is also homologically $(m - 2)$ -dimensional.

The following is a crucial criterion for the constructions we are seeking.

PROPOSITION 4.6. — *Let G be any finite group and let \mathcal{F} be a separating family for G . Then the following are equivalent:*

- *There is a (finite) 2-dimensional H -universal (G, \mathcal{F}) -complex.*
- *$N(\mathcal{F}_{> H})$ is homologically 1-dimensional for each subgroup $H \in \mathcal{F}$.*

- $N(\mathcal{F}_{\geq \mathcal{H}})$ is homologically 1-dimensional for every set \mathcal{H} of subgroups of G .

Given a separating family \mathcal{F} of subgroups of G , we say that $H \in \mathcal{F}$ is a *critical* subgroup if $N(\mathcal{F}_{>H})$ is not contractible. Given the above, we can concentrate our attention on the family \mathcal{SLV} and its subfamily of critical subgroups, denoted \mathcal{SLV}_c .

First we record conditions which allow one to show that certain subgroups in a family are not critical.

LEMMA 4.7. — *Let \mathcal{F} be any family of subgroups of G which has the property that $H \subseteq H' \subseteq H''$ and $H, H'' \in \mathcal{F}$ imply that $H' \in \mathcal{F}$. Fix a subgroup $H \in \mathcal{F}$; then $N(\mathcal{F}_{>H})$ is contractible if any of the following holds:*

- H is not an intersection of maximal subgroups in \mathcal{F} .
- There is a subgroup $\hat{H} \in \mathcal{F}$ properly containing H and such that $H \subsetneq K \cap \hat{H}$ for all $H \subsetneq K \in \mathcal{F}_c$

We can now state a simple sufficient condition for the existence of a 2-dimensional H -universal (G, \mathcal{F}) -complex:

PROPOSITION 4.8. — *Let \mathcal{F} be any separating family of subgroups of G . Assume for every non-maximal critical subgroup $1 \neq H \in \mathcal{F}$, that $N_G(H) \in \mathcal{F}$, and that $H \subsetneq K \cap N_G(H)$ for all non-maximal critical subgroups $K \in \mathcal{F}$ properly containing H . Then there exists a 2-dimensional H -universal (G, \mathcal{F}) -complex.*

We can in fact give a concrete description of the complex. For this we must introduce an integer associated to $H \in \mathcal{F}$.

DEFINITION 4.9. — If $H \in \mathcal{F}$, a family of subgroups of G , we define

$$i_{\mathcal{F}}(H) = \frac{1}{[N_G(H) : H]} \cdot (1 - \chi(N(\mathcal{F}_{>H}))).$$

Now let M_1, \dots, M_n be conjugacy classes representatives for the maximal subgroups of \mathcal{F} , and let H_1, \dots, H_k be conjugacy class representatives for all non-maximal critical subgroups of \mathcal{F} . Then there is a 2-dimensional H -universal (G, \mathcal{F}) -universal complex X which consists of one orbit of vertices of type G/M_i for each $1 \leq i \leq n$, $[-i_{\mathcal{F}}(H_j)]$ -orbits of 1-cells of type G/H_j for each $1 \leq j \leq k$, and free orbits of 1- and 2-cells. If G is simple⁶ then X can be constructed to contain exactly $i_{\mathcal{F}}(1)$ free orbits of 2-cells, and no free orbits of 1-cells.

5. EXPLICIT ACTIONS

In this section we will outline the construction of fixed point free actions on acyclic 2-dimensional complexes for the simple groups $PSL_2(2^k)$, for $k \geq 2$; $PSL_2(q)$ for $q \equiv \pm 3 \pmod{8}$ and $q \geq 5$; and for $Sz(2^k)$ for odd $k \geq 3$.

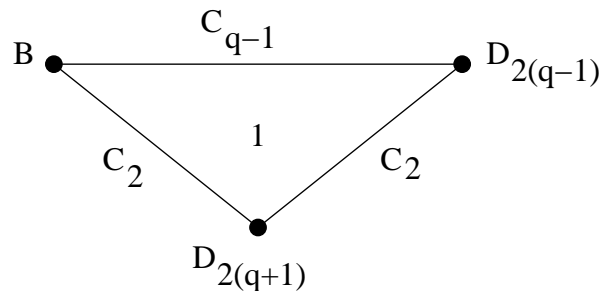
⁶In fact G must satisfy an additional technical condition which does not affect the results here.

EXAMPLE 5.1. — Let $G = PSL_2(q)$, where $q = 2^k$ and $k \geq 2$. Then there is a 2-dimensional acyclic fixed point free G -complex X all of whose isotropy subgroups are solvable. The complex X can be constructed with three orbits of vertices, with isotropy subgroups isomorphic to $B = \mathbb{F}_q \rtimes C_{q-1}$, $D_{2(q-1)}$ and $D_{2(q+1)}$; three orbits of edges with isotropy subgroups isomorphic to C_{q-1} , C_2 and C_2 ; and one free orbit of 2-cells.

Here B denotes a Borel subgroup, expressed as a semi-direct product isomorphic to $(C_2)^k \rtimes C_{q-1}$, identified with the subgroup of projectivized upper triangular matrices. In our notation C_r denotes the cyclic group of order r and D_r denotes the dihedral group of order r . In fact $D_{2(r-1)}$ can be identified with the subgroup of *monomial matrices*.

This example can be explained from the following analysis. The conjugacy classes of maximal solvable subgroups of G are represented precisely by the groups B , $D_{2(q-1)}$ and $D_{2(q+1)}$. The non-maximal critical subgroups must be intersections of maximal subgroups, one can check that up to conjugacy we get C_{q-1} , C_2 and 1. The precise numbers of orbits which appear is determined by calculating the integers $i_{SLV}(H)$ for the isotropy subgroups.

This example can actually be constructed directly using the 1-skeleton Y_1 of the coset complex Y for the triple of subgroups $(K_1, K_2, K_3) = (B, D_{2(q-1)}, D_{2(q+1)})$ in $G = PSL_2(\mathbb{F}_q)$ given by the maximal solvable subgroups. We can describe Y as the G -complex with vertex set $G/K_1 \sqcup G/K_2 \sqcup G/K_3$, where G acts by left translation, and with a 1-simplex for every pair of cosets with non-empty intersection and a 2-simplex for every triple of cosets with non-empty intersection. The following picture describes the orbit space Y/G :



It is not hard to see that as $G = \langle K_1, K_2, K_3 \rangle$, the complex Y is connected; however (as shown in [2], §9) it is not acyclic for $k \geq 3$, where $q = 2^k$. However, one can show that the module $H_1(Y_1, \mathbb{Z})$ is stably free – this involves a geometric argument based on the fact that Y_1 is a graph such that the fixed point sets Y_1^H are either contractible or empty for all subgroups $1 \neq H \subset G$ and contractible for all p -subgroups in G . The fact that G is a nonabelian simple group implies that the module must in fact be free (for a proof see [15], Prop. C.4.). Now we can simply attach a single free G -cell to Y_1 to kill its homology, yielding the acyclic complex X .

Carrying out this analysis in the classical case $G = A_5$ yields an acyclic complex X (which in this case is actually identical to the complex Y) whose cellular chains give a

complex of the form⁷

$$\begin{array}{ccccccc}
 & & \mathbb{Z}[A_5/C_2] & & \mathbb{Z}[A_5/A_4] & & \\
 & & \oplus & & \oplus & & \\
 0 & \rightarrow & \mathbb{Z}[A_5] & \rightarrow & \mathbb{Z}[A_5/C_2] & \rightarrow & \mathbb{Z}[A_5/D_6] \rightarrow \mathbb{Z} \rightarrow 0. \\
 & & \oplus & & \oplus & & \\
 & & \mathbb{Z}[A_5/C_3] & & \mathbb{Z}[A_5/D_{10}] & &
 \end{array}$$

EXAMPLE 5.2. — Let $G = PSL_2(\mathbb{F}_q)$, where $q = p^k$, $q \geq 5$ and $q \equiv \pm 3 \pmod{8}$. Then there exists a 2–dimensional acyclic fixed point free G –complex X , all of whose isotropy subgroups are solvable. More precisely, X can be constructed to have four orbits of vertices with isotropy subgroups isomorphic to $\mathbb{F}_q \rtimes C_{(q-1)/2}$, D_{q-1} , D_{q+1} and A_4 ; four orbits of edges with isotropy subgroups isomorphic to $C_{(q-1)/2}$, $C_2 \times C_2$, C_3 and C_2 ; and one free orbit of 2–cells.

These examples are slightly more complicated as the structure of the complex will depend on the value of q modulo 8.

Before explaining the final set of examples, we briefly recall the structure of the Suzuki groups $Sz(q)$ (see [6], [11], [19] for details). Fix $q = 2^{2k+1}$ and let $\theta \in \text{Aut}(\mathbb{F}_q)$ be the automorphism $x^\theta = x^{2^{k+1}} = x^{\sqrt{2q}}$ (note that $(x^\theta)^\theta = x^2$). For $a, b \in \mathbb{F}_q$ and $\lambda \in (\mathbb{F}_q)^*$, define the elements

$$S(a, b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ a & 1 & 0 & 0 \\ b & a^\theta & 1 & 0 \\ a^{2+\theta} + ab + b^\theta & a^{1+\theta} + b & a & 1 \end{pmatrix}$$

and

$$M(\lambda) = \begin{pmatrix} \lambda^{1+2^k} & 0 & 0 & 0 \\ 0 & \lambda^{2^k} & 0 & 0 \\ 0 & 0 & \lambda^{-2^k} & 0 \\ 0 & 0 & 0 & \lambda^{-1-2^k} \end{pmatrix}, \quad \tau = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Let $S(q, \theta) = \langle S(a, b) | a, b \in \mathbb{F}_q \rangle$, $T = \langle M(\lambda) | \lambda \in (\mathbb{F}_q)^* \rangle \cong C_{q-1}$ and

$$B = M(q, \theta) = S(q, \theta) \rtimes T \text{ and } N = \langle T, \tau \rangle \cong D_{2(q-1)}.$$

Then $Sz(q) \cong \langle M(q, \theta), \tau \rangle$, and under this identification the following hold

- (1) $S(q, \theta)$ is the 2–Sylow subgroup of $Sz(q)$.
- (2) There are four conjugacy classes of maximal subgroups in $Sz(q)$ which are solvable: (B) , (N) , (M_+) and (M_-) , where

$$M_+ \cong C_{q+\sqrt{2q+1}} \rtimes C_4 \text{ and } M_- \cong C_{q-\sqrt{2q+1}} \rtimes C_4.$$

⁷If we consider the original construction discussed in §1 of an acyclic A_5 –complex, then one can obtain the cellular structure below by subdividing each pentagon into a union of ten triangles.

These are all the maximal solvable subgroups in $Sz(q)$.

(3) $|Sz(q)| = q^2(q-1)(q^2+1) = q^2(q-1)(q+\sqrt{2q}+1)(q-\sqrt{2q}+1)$; note that the four factors in this expression are relatively prime.

We now describe the third set of examples.

EXAMPLE 5.3. — Let $q = 2^{2k+1}$, for any $k \geq 1$. Then there is a 2-dimensional acyclic fixed point free $Sz(q)$ -complex X , all of whose isotropy subgroups are solvable. X can be constructed to have four orbits of vertices with isotropy subgroups isomorphic to $M(q, \theta)$, $D_{2(q-1)}$, $C_{q+\sqrt{2q}+1} \rtimes C_4$ and $C_{q-\sqrt{2q}+1} \rtimes C_4$; four orbits of edges with isotropy subgroups isomorphic to C_{q-1} , C_4 , C_4 and C_2 ; and one free orbit of 2-cells.

6. NON-EXISTENCE OF FIXED POINT FREE ACTIONS

In this section we outline methods for showing that *most* finite simple groups cannot act on an acyclic 2-dimensional complex without fixed points. The first result in this direction is due to Segev [17].

THEOREM 6.1. — *If G is the alternating group A_n , with $n \geq 6$, then there is no fixed point free action of G on any acyclic 2-dimensional complex.*

Later this was substantially extended by Aschbacher–Segev [2], who proved:

THEOREM 6.2. — *If G is a finite simple group which acts on an acyclic 2-dimensional complex without fixed points, then G must be isomorphic to either a group of Lie type and Lie rank one, or isomorphic to the sporadic simple group J_1 (the first Janko group).*

We will now sketch the key arguments used to establish these results, which (by the Classification Theorem) rule out most of the finite simple groups. The following lemma will be referred to as the *four subgroup criterion*.

LEMMA 6.3. — *Let G be a finite group and X a 2-dimensional acyclic G -complex. If $H_1, H_2, H_3, H_4 \subset G$ are subgroups such that $X^{\langle H_i, H_j, H_k \rangle} \neq \emptyset$ for any i, j, k then*

$$X^{\langle H_1, H_2, H_3, H_4 \rangle} \neq \emptyset.$$

Proof. Suppose that in fact $X^{\langle H_1, H_2, H_3, H_4 \rangle} = \emptyset$. Let $\mathcal{H} = \{H_1, H_2, H_3, H_4\}$. Now $X^{\mathcal{H}}$ is the union of the acyclic subcomplexes X^{H_i} , which are such that any two or three of them have acyclic intersection, but the four have empty intersection. The homology of $X^{\mathcal{H}}$ is isomorphic (see [5], pg.168) to that of the nerve of the corresponding acyclic covering; yielding $H_2(X^{\mathcal{H}}, \mathbb{Z}) \cong H_2(\mathbb{S}^2, \mathbb{Z}) \cong \mathbb{Z}$. However we know that $X^{\mathcal{H}}$ must be homologically 1-dimensional, which yields a contradiction. \square

We now apply this result to multiply transitive groups.

PROPOSITION 6.4. — *Suppose that G acts 4-transitively on a set S with a point stabilizer $H \subset G$. If X is a 2-dimensional acyclic G -complex such that $X^H \neq \emptyset$, then $X^G \neq \emptyset$.*

Proof. If $|S| = 4$ then G is an extension of the form $1 \rightarrow Q \rightarrow G \rightarrow K \rightarrow 1$, where $K \subseteq \Sigma_4$ and $Q \subset H$. By Theorem 3.1, $\emptyset \neq X^Q$ must be acyclic, and as K is solvable its action on X^Q must have a fixed point and we are done. So assume that $|S| \geq 5$, and fix four elements $s_1, s_2, s_3, s_4 \in S$. For each $i = 1, 2, 3, 4$, let $H_i \subset G$ be the subgroup of elements which fix s_j for all $j \neq i$. For each $\{i, j, k, r\} = \{1, 2, 3, 4\}$, $\langle H_i, H_j, H_k \rangle$ is the point stabilizer of s_r and therefore fixes a point in X by assumption. Hence by Lemma 6.3 (where $G = \langle H_1, H_2, H_3, H_4 \rangle$), $X^G \neq \emptyset$. \square

We apply this to show that the alternating groups A_n for $n \geq 6$ do not admit fixed point free actions on acyclic 2-complexes. Note that A_n is $(n-2)$ -transitive on $\{1, 2, \dots, n\}$, with point stabilizer A_{n-1} and that m -transitivity implies k -transitivity for $k \leq m$; hence A_n is 4-transitive on $\{1, \dots, n\}$ for all $n \geq 6$ (see [1], page 56). If X is a 2-dimensional acyclic A_n -complex, then by our previous proposition, $X^{A_n} \neq \emptyset$ if $X^{A_{n-1}} \neq \emptyset$. Hence by induction we are reduced to considering the case when $G = A_6$; assume that X is a 2-dimensional acyclic G -complex with $X^G = \emptyset$. Using the subgroups $H_i = \langle (i, 5, 6) \rangle$ for $i = 1, 2, 3, 4$ we can show by contradiction that $X^H = \emptyset$ for $H = \text{Alt}\{1, \dots, 5\}$ (using the covering argument as before). Using an outer automorphism we can thus establish that $X^H = \emptyset$ for any $H \subset G$ with $H \cong A_5$. Next we consider the collection of subgroups

$$\mathcal{M} = \{ \langle (12)(36) \rangle, \langle (12)(45) \rangle, \langle (12)(34) \rangle, \langle (25)(36) \rangle, \langle (26)(35) \rangle \};$$

again applying the covering arguments and comparing with the homology of the nerve of this covering we see that $H_2(X^{\mathcal{M}}, \mathbb{Z}) \neq 0$, a contradiction. We refer to [17], page 39 for details.

This method can also be applied to the Mathieu groups M_n ; for $n = 11, 12, 23, 24$ they all act 4-transitively on a set with point stabilizer M_{n-1} . Now M_{10} contains A_6 as a subgroup of index two, hence every action of M_{11} or M_{12} on an acyclic 2-complex must have a fixed point. To obtain the same result for M_{23} and M_{24} , it suffices to establish it for M_{22} , which we will do subsequently.

The case of simple groups of Lie type, and of Lie rank at least equal to 2 can also be handled with these arguments (see [6] for background). We start with a basic lemma about parabolic subgroups.

LEMMA 6.5. — *Let G be a finite simple group of Lie type. Let Σ be the root system associated with G and let Σ_+ and Σ_- be the sets of positive and negative roots. Fix a set J of simple roots which does not contain all of them, and let L_J be the subgroup generated by the diagonal subgroup H together with the root subgroups X_r for all $r \in \langle J \rangle$. Let U_J and V_J be the subgroups generated by all X_r for roots $r \in \Sigma_+$ or $r \in \Sigma_-$, respectively, which are not in $\langle J \rangle$. Then $U_J \triangleleft P_J = U_J L_J$ and $V_J \triangleleft P'_J = V_J L_J$, U_J and V_J are nilpotent and $\langle U_J, V_J \rangle = G$.*

In our context we obtain the following fixed point theorem

LEMMA 6.6. — *Let G be a finite simple group of Lie type, and let $P \subsetneq G$ be one of the parabolic subgroups P_J or $P_{J'}$ in the previous lemma. Then for any action of G on an acyclic 2-complex X , $X^P \neq \emptyset$.*

Proof. Let us assume that $X^G = \emptyset$. Then there are subgroups $U_J \triangleleft P_J$, $V_J \triangleleft P_{J'}$ and $L_J = P_J \cap P_{J'}$ such that U_J and V_J are nilpotent, $P_J = U_J L_J$, $P_{J'} = V_J L_J$, and $\langle U_J, V_J \rangle = G$. Note that X^{U_J} and X^{V_J} are acyclic, disjoint and L_J -invariant. Considering the subspaces $A = X^{U_J}$ and $B = X^{V_J}$ and the action of L_J , we see that $X^{L_J} \neq \emptyset$ (Corollary 3.3); similarly we conclude from Lemma 3.4 that X^{P_J} and $X^{P_{J'}}$ are non-empty. \square

Now we can prove

THEOREM 6.7. — *If G is a simple group of Lie type and Lie rank at least two, then every G -action on an acyclic 2-dimensional complex has a fixed point.*

Proof. Take a root system $\Sigma = \Sigma_+ \sqcup \Sigma_-$ for G and let $J_1 \sqcup J_2$ be a decomposition of the set of simple roots as a disjoint union of non-empty subsets. For each $i = 1, 2$, set $H_i^+ = \langle H, X_s \mid s \in J_i \rangle$, and $H_i^- = \langle H, X_{-s} \mid s \in J_i \rangle$. The subgroup generated by any three of the H_i^\pm is contained in one of the parabolic subgroups P_{J_i} or P'_{J_i} and so has non-empty fixed point set in X . But in fact one can verify that $\langle H_1^\pm, H_2^\pm \rangle = G$, since it contains all subgroups X_s, X_{-s} for simple roots s and hence $X^G \neq \emptyset$ by the four subgroup criterion. \square

In [2], Aschbacher and Segev were able to apply the four subgroup criterion to prove that any sporadic simple group other than the Janko group J_1 acting on an acyclic 2-complex has a fixed point. In [15] a different treatment is given, showing that all the sporadics can be handled using a consistent technique which relies on understanding the subgroup structure of these groups in some detail. The essential result is the following.

PROPOSITION 6.8. — *Let \mathcal{F} be a separating family for G and let $K_1, K_2, K_3 \in \mathcal{F}$ be three subgroups such that neither K_2 nor K_3 is conjugate to K_1 . Let $K_{ij} = K_i \cap K_j$ and $K = K_1 \cap K_2 \cap K_3$. Let $\mathcal{F}_0 \subseteq \mathcal{F}$ denote the subfamily consisting of \mathcal{F}_c together with all subgroups conjugate to any of the K_i, K_{ij} or K . Assume that the following conditions hold, where $G' = \langle K_1, K_2, K_3 \rangle$:*

- $\frac{1}{[K_{12}:K]} + \frac{1}{[K_{13}:K]} + \frac{1}{[K_{23}:K]} \leq 1 + \frac{1}{[K_1:K]} + \frac{1}{[K_2:K]} + \frac{1}{[K_3:K]} - \frac{1}{[G':K]}$
- K_1 is maximal in \mathcal{F} .
- There is no $H \in \mathcal{F}_0$ such that $K \subsetneq H \subsetneq K_{12}$ or $K_{12} \subsetneq H \subsetneq K_1$.
- $N_G(K_1) \cap N_G(K_{12}) \cap N_G(K) = K$
- The triples (K_1, K_{12}, K) and (K_1, K_{13}, K) are not G -conjugate.

Then $H_2(N(\mathcal{F}_{\geq(K)}), \mathbb{Z}) \neq 0$ and so there is no 2-dimensional, H -universal (G, \mathcal{F}) -complex.

This result can be proved as follows: the coset complex Y for the triple (K_1, K_2, K_3) must have $H_2(Y, \mathbb{Z}) \neq 0$ by the first hypothesis (this follows from a counting argument); the other conditions allow one to push a non-zero class non-trivially into

$$H_2(N((\mathcal{F}_0)_{\geq(K)}), \mathbb{Z}) \cong H_2(N(\mathcal{F}_{\geq(K)}), \mathbb{Z})$$

via the homomorphism induced by the G -equivariant simplicial map $Y^* \rightarrow N((\mathcal{F}_0)_{\geq(K)})$ sending each vertex in the barycentric subdivision Y^* of Y to its isotropy subgroup. By Proposition 4.6, this implies the stated result. This proposition can be applied systematically to yield

THEOREM 6.9. — *Let G be any of the sporadic simple groups; then there is no 2-dimensional acyclic G -complex without fixed points.*

We illustrate how this may be applied with two examples. Here we assume that we are given a 2-dimensional acyclic X with a fixed point free G -action, and take \mathcal{F} to be the separating family of $H \subset G$ with $X^H \neq \emptyset$.

EXAMPLE 6.10. — Let $G = M_{22}$, one of the Mathieu groups. We can take $K_3 \cong 2^4 : A_6$, the subgroup which leaves invariant some hexad in the Steiner system of order 22, and it has an obvious action on this set of order 6 (see [10], Thm 6.8). Then K_1 can be taken to be the stabilizer of a point z in the hexad, and K_2 the stabilizer of some pair of points in the hexad including z . In this case $K_1 \cong L_3(4)$, $K_2 \cong 2^4 : S_5$, $K_{12} \cong 2^4 : A_5$, $K_{13} \cong 2^4 : A_5$, $K_{23} \cong 2^4 : S_4$ and $K \cong 2^4 : A_4$. Note that the K_{12} and K_{13} are distinct parabolic subgroups in $L_3(4)$. The conditions in our previous proposition can be checked to hold (note that from our previous results we can see that the K_i all act with fixed points and hence are in \mathcal{F}) and so we have completed the verification that the Mathieu groups have no fixed point free actions on acyclic 2-complexes.

Next we deal with the case J_1 which was not covered by [2].

EXAMPLE 6.11. — Let $G = J_1$, the first Janko group. Take $K_1 \cong (C_2)^3 \rtimes G_{21}$, where G_{21} is the Frobenius group of order 21, i.e. $C_7 \rtimes C_3$. K_1 is a maximal subgroup in J_1 . Let $K_2 \cong C_7 \times C_6$ be the normalizer of a subgroup of order 7 in K_{12} , and let $K_3 \cong C_3 \times D_{10}$ be the centralizer in G of a subgroup of order 3 in K_2 . Then $K_{12} \cong C_7 \rtimes C_3$, $K_{13} \cong C_6 \cong K_{23}$, and $K = K_1 \cap K_2 \cap K_3 \cong C_3$. Note that all these subgroups are solvable, and so are in \mathcal{F} . We can verify that

$$\sum_{i < j} \frac{1}{[K_{ij} : K]} = \frac{1}{7} + \frac{1}{2} + \frac{1}{2} < 1 + \frac{1}{14} + \frac{1}{10} = 1 + \frac{1}{[K_2 : K]} + \frac{1}{[K_3 : K]}$$

while the other conditions are also easy to check, hence showing that J_1 has no fixed point free action on an acyclic 2-complex.

We now consider the finite groups of Lie type which have Lie rank exactly equal to one. There are four families of such groups: the two dimensional projective special linear groups $L_2(q)$, the three dimensional projective special unitary groups $U_3(q)$, the Suzuki groups $Sz(2^{2k+1})$, and the Ree groups $Ree(3^{2k+1}) = {}^2G_2(3^{2k+1})$.

The following propositions are used to handle these groups.

PROPOSITION 6.12. — *Let L be one of the simple groups $L_2(q)$ or $Sz(q)$, where $q = p^k$ and p is prime ($p = 2$ in the second case). Let $G \subset \text{Aut}(L)$ be any subgroup containing L and \mathcal{F} a separating family for G . Then there exists a 2-dimensional H -universal (G, \mathcal{F}) -complex if and only if $G = L$, $\mathcal{F} = \mathcal{SLV}$ and q is a power of 2 or $q \equiv \pm 3 \pmod{8}$.*

PROPOSITION 6.13. — *Let $G = U_3(q)$, or ${}^2G_2(3^{2k+1})$. Then there is no 2-dimensional acyclic G -complex without fixed points.*

These results are proved by combining our previous non-existence techniques with the following additional notion. For any family of subgroups \mathcal{F} and any maximal element $M \in \mathcal{F}$, we set $Lk_{\mathcal{F}_{>1}}(M) = N(\mathcal{F}_{>1}^{<M}) = N(\{H \in \mathcal{F} \mid 1 \neq H \subsetneq M\})$. Then we have

LEMMA 6.14. — *Let \mathcal{F} denote a separating family for G . Let $\mathcal{F}_0 \subset \mathcal{F}$ be any subfamily which contains \mathcal{F}_c , and such that each non-maximal subgroup in \mathcal{F}_0 is contained in two or more maximal subgroups. Assume that \mathcal{F} satisfies the following two conditions*

- (1) $N(\mathcal{F}_{>1})/G$ is connected and $H_1(N(\mathcal{F}_{>1})/G, \mathbb{Z}) = 0$.
- (2) There is a maximal subgroup $M \in \mathcal{F}$ such that $Lk_{(\mathcal{F}_0)}(M)$ is not connected.

Then there is no H -universal 2-dimensional (G, \mathcal{F}) -complex.

Roughly speaking the proof of this lemma goes as follows: if such a complex did exist, then by (1) the singular set must be acyclic; but the prescribed conditions imply that the links at all vertices must be connected – hence contradicting (2).

In many instances this allows one to prove non-existence of a fixed point free action by contradiction; assuming its existence we can then find a maximal subgroup in the separating family such that the corresponding link is not connected. This of course requires a rather intricate knowledge of the maximal subgroups and more generally the finer structure of the groups under consideration. We refer to [15] §6 for complete details.

7. PROOF OF THE MAIN THEOREM

We are now prepared to sketch a proof of the main theorem. We recall the statement.

THEOREM 7.1. — *If G is any finite group, then there is an essential fixed point free 2-dimensional finite acyclic G -complex if and only if G is isomorphic to one of the simple groups $PSL_2(2^k)$ for $k \geq 2$; $PSL_2(q)$ for $q \equiv \pm 3 \pmod{8}$ and $q \geq 5$; or $Sz(2^k)$ for odd $k \geq 3$. Moreover the isotropy subgroups of any such G -complex are all solvable.*

Proof. We know that if G has an essential action on an acyclic 2-complex X without fixed points, then there is a non-abelian simple normal subgroup $L \triangleleft G$ with a fixed point free action and such that $G \subseteq \text{Aut}(L)$. By the Classification Theorem, we know that L must be an alternating group, a group of Lie type, or a sporadic simple group. The results in the previous section rule out all groups on this list⁸ except possibly the ones in the statement of the theorem. However we have already seen that these groups do in fact act on an acyclic 2-complex without fixed points, and that the isotropy subgroups are all solvable. This completes the proof. \square

The work of Oliver and Segev has provided a complete picture for understanding fixed point free group actions on acyclic 2-dimensional complexes. There remains however the problem of considering actions on *contractible* 2-dimensional complexes. In fact Aschbacher and Segev [2] have raised the following

QUESTION 7.2. — If X is a finite contractible 2-dimensional G -complex, then is $X^G \neq \emptyset$?

This remains open. The results described here are a basic step towards investigating this question but it will probably require a substantially different approach.

REFERENCES

- [1] Aschbacher, M. – **Finite Group Theory**, Cambridge Studies in Advanced Mathematics 10, CUP (2000).
- [2] Aschbacher, M. and Segev, Y. – *A fixed point theorem for groups acting on finite 2-dimensional acyclic simplicial complexes*, Proc. London Math. Soc. (3) **67** (1993), 329–354.
- [3] Aschbacher, M. and Smith S. – *Work in progress: a classification of quasithin groups*, available at
<http://www.math.uic.edu/~smiths/papers/quasithin/quasithin.dvi>.
- [4] Bredon, G. – **Introduction to Compact Transformation Groups**, Academic Press (1972).
- [5] Brown, K. – **Cohomology of Groups**, Springer-Verlag GTM 87 (1982).
- [6] Carter, R. – **Simple Groups of Lie Type**, Wiley (1972).
- [7] Floyd, E. and Richardson R. – *An action of a finite group on an n -cell without stationary points*, Bull.AMS **65** (1959), 73–76.
- [8] Gorenstein, D. – **Finite Groups**, Harper and Row (1969).
- [9] Gorenstein, D. – **The Classification of Finite Simple Groups**, Plenum Press, New York (1983).

⁸In fact the Tits group ${}^2F_4(2)'$ must be handled separately because it is not the full Lie type group ${}^2F_4(2)$.

- [10] Griess, R. – **Twelve Sporadic Groups**, Springer-Verlag (1998).
- [11] Huppert, B. and Blackburn, N. – **Finite Groups III**, Springer-Verlag (1982).
- [12] Kirby, R. and Scharlemann, M. – *Eight faces of the Poincaré homology 3–sphere*, in **Geometric Topology**, Academic Press (1979).
- [13] Oliver, R. – *Fixed point sets of group actions on finite acyclic complexes*, Comm. Math. Helv. **50** (1975), 155–177.
- [14] Oliver, R. – *Smooth compact group actions on disks*, Math.Z. **149** (1976), 79–96.
- [15] Oliver, R. and Segev, Y. – *Fixed point free actions on acyclic 2–complexes*, Acta Mathematica (to appear).
- [16] Serre, J.-P. – **Trees**, Springer-Verlag (1980).
- [17] Segev, Y. – *Group actions on finite acyclic simplicial complexes*, Israel J.Math. **82** (1993), 381–394.
- [18] Smith, P.A. – *Fixed points of periodic transformations*, AMS Coll.Pub. XXVII (1942), 350–373.
- [19] Suzuki, M. – *On a class of doubly transitive groups*, Annals of Math. **75** (1962), 105–145.

Alejandro ADEM

Department of Mathematics

University of Wisconsin

Madison, Wisconsin 53706 - U.S.A.

E-mail : adem@math.wisc.edu

THE PCP THEOREM

[after Arora, Lund, Motwani, Safra, Sudan, Szegedy]

by Bernard CHAZELLE

1. INTRODUCTION

The notion of *interactive proof systems* evolved out of cryptography and computational group theory. The cryptographic context is best explained through a little tale (perhaps one day to come true). One fine morning, one of your esteemed colleagues wakes up with, in his head, a crisp, concise, complete proof of Riemann's Hypothesis! Wisdom being one of his many qualities, he is not about to post his proof on the internet. Paranoia being another one, he is not even willing to reveal a single bit of information about the proof; that is, besides its conclusion that the RH is true. Is there any way for your colleague to convince you and the rest of the mathematical community that, indeed, he has a correct proof? Of course, one needs to define what exactly is meant by not "revealing a single bit". That is the subject of *zero-knowledge* cryptography.

The *PCP theorem* addresses a simpler variant: Can your colleague write down his proof in such a way that, were you to peek into it at a constant number of randomly chosen spots, you would leave utterly convinced of its validity? In other words, can he encode the proof as a string of bits so that: (i) a correct proof will never fail to convince you; (ii) an incorrect one will fool you with only a negligible probability? The catch is, you will be allowed to look at only a constant number of bits chosen at random. The PCP theorem asserts the existence of such an encoding. It is striking that the number of lookups can be kept constant regardless of the length of the proof. In fact, if you can put up with a failure rate slightly above $1/2$, i.e., accept a wrong proof half the time, but still never reject a correct one, then the number of bits can be reduced to 3. On the other hand, if you are allowed to read as many bits as are needed to store, say, two lines of this article, the probability of failure drops to 10^{-100} . A key point is that the new proof can be derived from the old one purely syntactically. In other words, one can write a compiler to translate the proof mechanically without any knowledge of mathematics. Furthermore, the new proof is not much longer than the previous one.

A common initial reaction to the PCP theorem is that it must be either wrong or trivial. Why wrong? It seems to imply that any flaw in the proof should spread itself all over the place, so as to be caught immediately in a random peek. But, how can so much information be stored in so few bits? Here is how: If the proof is correct, print it as such; if it is wrong, then intersperse the statement $2 + 2 = 3$ at every other step. The problem with that encoding is that a correct proof will not convince anyone. The beauty of the PCP theorem is not that flaws are caught so easily: it is that the mere absence of a flaw is persuasive in and of itself. There is nothing amazing about catching a liar's lie. But it is quite a feat to hear a true story from a congenital liar and end up believing it.

2. THE PCP VIEW OF NP

A *Turing machine* is a computer model whose main feature, for our purposes, is to be universal: in particular, whatever it can compute in time polynomial in the length of the input is believed to constitute what is tractable in *any* (non-quantum) model. The class P consists of the sets for which membership can be decided by a Turing machine in polynomial time. For example, the set of singular square integer matrices is in P, because determinants can be computed in a polynomial number of steps. The class NP includes the sets for which membership can be verified in polynomial time. For instance, the set of polynomials in $\mathbf{Q}[X_1, \dots, X_n]$ with at least one zero in $\{0, 1\}^n$ is in NP. The reason is that, given a polynomial f and a point $x \in \{0, 1\}^n$, one can check if $f(x) = 0$ in time polynomial in the number of bits needed to represent f . To find such a zero from scratch seems more difficult (to put it mildly), and it is widely conjectured that $P \neq NP$. Within computer science, this open question dwarfs all others in importance.

A 3-CNF formula is a conjunction of clauses, each one consisting of three literals; for example, $(v_1 \vee \neg v_2 \vee v_3) \wedge (v_2 \vee v_3 \vee \neg v_4)$. It is satisfiable if some true/false assignment of the v_i 's makes the formula true. The one above is, whereas $(v_1 \vee v_1 \vee v_1) \wedge (\neg v_1 \vee \neg v_1 \vee \neg v_1)$ is not. The set of satisfiable 3-CNF formulas is called 3-SAT. A classical result of Cook and Levin says that 3-SAT is *NP-complete*, meaning that, not only it is in NP, but deciding membership in *any* NP set can be reduced to testing the satisfiability of a 3-CNF. The Cook-Levin theorem shows that to understand 3-SAT is to understand all of NP.

Many other sets are known to be NP-complete: for example, the set of 3-colorable graphs. (A graph is 3-colorable if its nodes can be colored red, white, and blue with no edge sharing the same color.) The existence of NP-complete sets brings breathtaking universality into the computing picture. It implies that anyone who can quickly color graphs can also solve algebraic equations over finite fields, factor integers, compute discrete logarithms, find short vectors in lattices, determine the largest clique in a graph, etc.

To formalize what a mathematical proof has to do with NP takes some effort, but the intuition is clear. In any reasonable axiomatic system, this set is in NP:

$$\{ \langle T.1^n \rangle \mid T \text{ is a theorem with a proof of size at most } n \},$$

where $\langle T.1^n \rangle$ denotes the 0/1 string formed by writing the theorem T in binary in the axiomatic system and appending n ones at the end. A prover can guess a proof of length at most n , and the verifier can then check it in time polynomial in its length.

The class NP can be described in the language of proofs. If $L \in \text{NP}$ then, given any $x \in L$, there exists a *short proof*, i.e., a polynomial-time computation, that x indeed belongs to L ; for example, the solution of an algebraic equation. Conversely, if $x \notin L$, then no proof can convince anyone that x is in L . *Probabilistically checkable proofs* (PCP) add a small twist to this view: randomization. A PCP system for a set L consists of a string of bits (the proof) and a Turing machine with access to random bits (the verifier). Given an input x of n bits, the verifier generates $r(n)$ random bits;¹ then it looks up $q(n)$ bits of the proof at locations of its choice. The lookups are done all at once nonadaptively. Finally, after a polynomial amount of (deterministic) computation, the verifier must either accept or reject the proof. The class of sets L that satisfy the two requirements below is denoted by $\text{PCP}[r(n), q(n)]$:

- Given any $x \in L$, there is a proof that causes the verifier to accept x with probability 1.
- Given any $x \notin L$, every proof is rejected with probability at least $1/2$.

The functions r and q are called the *random-bit complexity* and *query-bit complexity*, respectively. To alleviate the notation, both of them are understood up to a constant factor. If $r(n) = O(\log n)$, the number of distinct random strings is polynomial and, by running the verifier on all of them, it is immediate that $\text{PCP}[\log n, 1] \subseteq \text{NP}$. Proving the reverse inclusion requires a great deal of ingenuity. The purpose of this article is to explain the proof at a conceptual level, leaving mathematical technicalities aside. The *PCP theorem* states that

$$(1) \quad \text{NP} = \text{PCP}[\log n, 1].$$

Note that the proof size can be assumed to be polynomial since at most $q(n)2^{r(n)} = n^{O(1)}$ bits of the proof have a chance of ever being read. The PCP theorem can be restated in a way that highlights its “error-spreading” aspect. Given any 3-CNF formula Φ on n variables, there exists another one, denoted by Ψ , which contains $n^{O(1)}$ variables and is satisfiable if and only if Φ is. Furthermore, if Ψ is not satisfiable, then no truth assignment can satisfy more than a fraction $1 - \varepsilon$ of its clauses, for some constant $\varepsilon > 0$. Finally, Ψ can be derived from Φ in polynomial time.

¹Throughout our discussion, random points or numbers are drawn uniformly, independently from a set that is always clearly understood from the context; in this case the set is $\{0; 1\}$.

It is instructive to see how this follows from (1), because the argument anticipates aspects of the proof of the PCP theorem. Consider a PCP system for Φ . Among the $2^{r(n)}$ possible random strings, some lead to acceptance, others (possibly) to rejection. Given such a string s , let Π_1, \dots, Π_q be the bits of the proof read by the verifier. (The locations of these bits depend on s but not on the proof itself.) Let Φ_s be a Boolean formula that evaluates to true if and only if Π_1, \dots, Π_q lead to acceptance: Φ_s has $q = O(1)$ variables, each one corresponding to one of the bits read. It is routine to convert Φ_s into a constant size 3-CNF formula Φ'_s by adding a few auxiliary variables if necessary. The formula $\Psi = \bigwedge_s \Phi'_s$ fits the bill. To see why, consider the (only interesting) case: If Φ is not satisfiable, then regardless of the proof, i.e., of the truth assignment of the Φ_s 's variables, at least half of these formulas are false and, hence, so is a constant fraction of the clauses in Ψ . \triangleright

This characterization of the PCP theorem, which interestingly makes no mention of proofs, verifiers, or even randomization, points to the connection between PCP and inapproximability. Indeed, it implies that it is NP-complete to distinguish between a satisfiable formula and one for which no truth assignment satisfies at least a fraction $1 - \varepsilon$ of the clauses. Another way to look at this result is that if we set out to maximize the number of satisfied clauses in a formula, then we cannot hope to find an approximate solution within a factor $1 - \varepsilon$ of the maximum in polynomial time, unless $P = NP$. (Other applications are mentioned in the *Historical Notes* section.)

REMARK 2.1. — From a mathematician's perspective, the PCP theorem might appear to focus on the "uninteresting" part of mathematics. It is a restatement of NP, not of P; as such, it says nothing about the difficulty of finding proofs. Also, it treats readers as mere fact-checkers. But mathematicians read proofs not so much to find bugs in them but to understand the ideas behind them. This mental picture, so vital to mathematics, is absent from the PCP viewpoint. Within the restrictive framework of verification, the PCP theorem is an impressive statement nevertheless.

REMARK 2.2. — The proof of the PCP theorem is a mix of elementary algebra and probability theory; it is long and technical but not particularly difficult. Its originality lies elsewhere: in two places to be precise. One is its use of computational self-reducibility. Instead of keeping the usual separation between proving and verifying, the verifier's work is itself re-encoded as part of the proof: the reader of a proof is made partly its author! The other intriguing aspect of the PCP theorem is its ingenious use of error-correcting codes to express not just signals and bit streams (in typical coding theory fashion) but mathematical proofs, instead.

We close this section presenting a short, archetypical motif of the proof. Given a 3-CNF formula Φ , we wish to design a PCP system to verify its satisfiability. The idea is to construct a large family of multivariate polynomials f_i such that: if Φ is satisfiable, then

any satisfying truth assignment corresponds to a common zero to all the f_i 's; otherwise, no more than half of them have a common zero. Suppose that the proof falsely claims that Φ is satisfiable. The verifier asks the prover to present the value of the f_i 's at that common zero. If the prover obliges, then half of them will be nonzero and the verifier will easily catch the lie by random sampling. Therefore, the prover must cheat by substituting 0 for the actual values. The verifier's strategy is then to push the prover into the liar's standard pattern of generating new lies to cover up old ones.

Here is one way to do that: Encode any linear combination of the f_i 's at the claimed zero as the image of a linear form g at some point x , and ask the prover to present g by values, i.e., provide a table, indexed by x , of all the values of $g(x)$. If the prover lies at one spot $g(x)$, then it must lie all over the place as well, since the verifier can evaluate $g(x)$ as $g(x+y) - g(y)$ for a random y , and hence, quickly spot any inconsistency. Many other such error-detecting mechanisms are needed. They all share the same goal, which is to force the prover to present functions that are very close to some "canonical" functions. Canonical functions are chosen to satisfy certain functional equations. Furthermore, any family of functions that satisfy these equations yield a satisfying assignment for Φ and, hence, a contradiction.

NOTATION — $\mathcal{F}_{q,m}^d$ denotes the set of m -variate polynomials of total degree at most d with coefficients in \mathbf{F}_q , the finite field of q elements (not to be confused with the query time). We restrict ourselves exclusively to prime fields. We say that a function $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$ is δ -close to a (finite) family of functions if, for some g in that family, $\text{Prob}[f(x) \neq g(x)] \leq \delta$, for random $x \in \mathbf{F}_q^m$. The smallest such δ is called the distance of f to the family. Given any nonempty $H \subseteq \mathbf{F}_q$, we use $\sum_{H^m} f$ as shorthand for $\sum_{x_1, \dots, x_m \in H} f(x_1, \dots, x_m)$. All logarithms are to the base 2.

3. TESTING TOOLS

Intuitively, the encoding of a proof in a PCP system must be such that any local deviation from what the verifier expects should be visible nearly everywhere. The relation between a polynomial and the corresponding polynomial map shares this characteristic: Changing a polynomial map at a single point has a rippling effect visible almost everywhere. This analogy suggests a line of attack: encode proofs as polynomials. In this section we pursue this lead and build a number of algebraic tools to be used later when proving the PCP theorem. We specify a polynomial in two different ways. It can be presented, i.e., written down, *by coefficients* (with the obvious meaning) or *by values* (as mentioned earlier). The appeal of the presentation by values is that it is extremely redundant and, hence, provides a built-in error-correcting mechanism.

3.1. The Sumcheck Test

Given $f \in \mathcal{F}_{q,m}^d$, $c \in \mathbf{F}_q$, and some nonempty $H \subseteq \mathbf{F}_q$, how can a proof convince a verifier that $\sum_{H^m} f = c$? Of course, the verifier can compute the sum on its own without a proof, but that requires evaluating f at $O(|H|^m)$ points. Can it be done faster? If the field is big enough, say, $q > 2dm$, it can be done with a single evaluation of f . Write $f_a = f(a, x_2, \dots, x_m) \in \mathcal{F}_{q,m-1}^d$; if $m = 1$, $\sum_{H^m \Gamma_1} f_a$ denotes $f(a)$. For $m > 0$, the PC proof is defined recursively as follows:

PC PROOF THAT $\sum_{H^m} f = c$

[1] Present $g(x) = \sum_{H^m \Gamma_1} f_x$ by coefficients.

[2] For all $a \in \mathbf{F}_q$, write down the PC proof that $\sum_{H^m \Gamma_1} f_a = g(a)$.

The recursion bottoms out at $m = 0$: no PC proof needed there. The verifier adopts a two-pronged strategy: First, trust the proof and check that it supports the claim; then, test the proof for internal consistency. Accordingly, the verifier begins by checking that $\sum_H g = c$, rejecting the proof if this fails. Next, it verifies that $g(x)$ is, indeed, the polynomial it thinks it is. It picks a random $a \in \mathbf{F}_q$, and uses the PC proof in [2] to verify recursively that, indeed, $\sum_{H^m \Gamma_1} f_a$ is equal to $g(a)$. (If $m = 1$, the verifier does not need to go to [2], since $\sum_{H^m \Gamma_1} f_a$ is available via a single evaluation of the polynomial f .) If this succeeds, the verifier accepts the proof and its claim that $\sum_{H^m} f = c$; else, it rejects it.

To argue that this works, we first observe that if the proof is correct, the verifier always accepts it. On the other hand, if any test fails, rejection ensues. So, the only case worth considering is where the claim is not true but all the tests pass and, therefore, the proof is accepted. Important: all subsequent correctness proofs in this paper will be limited to this case, too, without a need to repeat why.

The case $m = 0$ is error-free. Note that it is the (only) place where the verifier can match the proof against its own knowledge of f . All other tests involve only the internal consistency of the PC proof. Assume now that $m > 0$. We prove by induction on m that the verifier wrongly accepts with probability at most dm/q . Since the first test passes, $g(x)$ cannot agree with the true $\sum_{H^m \Gamma_1} f_x$ everywhere (else their respective sums would be both equal to c). But, being univariate polynomials of degree at most d , they agree at $x = a$ with probability at most d/q . This agreement might lead the verifier to wrongly accept. If there is disagreement at $x = a$, however, the verifier is back to its old task, but now with only $m - 1$ variables. So, by induction, it is fooled with probability at most $d(m - 1)/q$. Adding the earlier bound of d/q completes the induction.

LEMMA 3.1. — Given any $f \in \mathcal{F}_{q,m}^d$, $c \in \mathbf{F}_q$, and nonempty $H \subseteq \mathbf{F}_q$, if $\sum_{x \in H^m} f(x) = c$, then there exists a proof of that fact that the verifier always accepts. Otherwise, no proof is accepted with probability greater than dm/q . The verifier reads $O(dm \log q)$ bits of the proof, needs $O(m \log q)$ random bits, and performs a single evaluation of the polynomial f .

The most remarkable feature of this test is that only one evaluation of the function f is necessary. Note how increasing the size of the field, by making the polynomial map f increasingly redundant, has the effect of decreasing the error probability. The sumcheck test can be used in conjunction with the simple fact below to test if a function that is close enough to a polynomial vanishes everywhere.

LEMMA 3.2. — Given $H \subseteq \mathbf{F}_q$ of size $h > 0$, it is possible to build polynomials R_1, \dots, R_{q^ℓ} in $\mathcal{F}_{q,\ell}^{h\ell}$, each one in time $(h + \ell)^{O(1)}$, so that for any nonzero function $g : H^\ell \mapsto \mathbf{F}_q$ and a random index $1 \leq i \leq q^\ell$, the probability that $\sum_{H^\ell} gR_i = 0$ is at most $h\ell/q$.

3.2. The Low-Degree Test

We wish to design a PCP system to convince a verifier that a function $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$ is close to being a polynomial. Restrictions to lines are particularly useful for that purpose. Given $t \in \mathbf{F}_q$, let $f_{a,b}(t)$ denote the univariate function $f(a + tb)$. We define $\Delta_\ell(f)$ as the expected distance of $f_{a,b}$ to $\mathcal{F}_{q,1}^d$ for random $a, b \in \mathbf{F}_q^m$.

LEMMA 3.3. — If q/md^3 is large enough, given any function $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$, the distance of f to $\mathcal{F}_{q,m}^d$ is at most $c\Delta_\ell(f)$, for some absolute constant $c > 0$.

Intuitively, this is saying that if the restriction of a function to a random line is close to a polynomial, so is the function itself. This suggests an obvious PCP system.

PC PROOF THAT f IS δ -CLOSE TO $\mathcal{F}_{q,m}^d$.

For every pair $a, b \in \mathbf{F}_q^m$, present by coefficients the polynomial $g_{a,b} \in \mathcal{F}_{q,1}^d$ that is closest to $f_{a,b}$.

The verifier chooses two small constant parameters $\delta, \varepsilon > 0$ and picks $k = \lceil c\delta^{-1} \log \varepsilon^{-1} \rceil$ random (a_i, b_i, t_i) , with $a_i, b_i \in \mathbf{F}_q^m$ and $t_i \in \mathbf{F}_q$. Next, it checks that $g_{a_i, b_i}(t_i) = f(a_i + t_i b_i)$, for each $1 \leq i \leq k$. If all k tests succeed, then the proof is accepted. Any failure means rejection.

Correctness is immediate: If $f \in \mathcal{F}_{q,m}^d$, then $g_{a,b}(t)$ coincides with $f(a + tb)$, for all a, b , and all tests succeed. Suppose now that the distance from f to $\mathcal{F}_{q,m}^d$ exceeds δ . For any fixed a_i, b_i and random t_i , the probability that $f(a_i + t_i b_i) \neq g_{a_i, b_i}(t_i)$ is precisely the

distance from f_{a_i, b_i} to $\mathcal{F}_{q,1}^d$. Averaging over all a_i, b_i , it follows from Lemma 3.3 that, for any fixed i and random a_i, b_i, t_i ,

$$\text{Prob}[f(a_i + t_i b_i) \neq g_{a_i, b_i}(t_i)] = \Delta_\ell(f) > \frac{\delta}{c};$$

therefore, all k tests succeed with probability less than $(1 - \delta/c)^k < \varepsilon$. This implies the following lemma.

LEMMA 3.4. — *Let δ, ε be two arbitrarily small positive constants. Given a function $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$, fix some integer d such that q/md^3 is large enough. If $f \in \mathcal{F}_{q,m}^d$, then there exists a proof that the verifier always accepts. If f is not δ -close to $\mathcal{F}_{q,m}^d$, then no proof is accepted with probability greater than ε . The verifier reads $O(d \log q)$ bits of the proof, needs $O(m \log q)$ random bits, and performs $O(1)$ evaluations of f .*

If f is δ -close to $\mathcal{F}_{q,m}^d$, then its nearest polynomial (unique for small enough δ) can be evaluated at a random point by using f . To evaluate it at an arbitrary point, however, requires a recovery mechanism. The striking feature of the result below is that a single evaluation of f is sufficient to recover $f(x)$ at k (non-necessarily random) points.

LEMMA 3.5. — *For $\delta > 0$ small enough, fix d such that $q\delta/kd$ is large enough. Given a function $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$, let f_o be a nearest polynomial in $\mathcal{F}_{q,m}^d$. Pick k arbitrary points, z_1, \dots, z_k in \mathbf{F}_q^m . If f is δ -close to $\mathcal{F}_{q,m}^d$, then f_o is unique and there exists a proof that allows the verifier to output $f_o(z_1), \dots, f_o(z_k)$. Otherwise, with probability $1 - O(\sqrt{\delta})$, the verifier either outputs the right values or rejects the proof. The verification reads $O(dk \log q)$ bits of the proof, needs $O(m \log q)$ random bits, takes $(mdk \log q)^{O(1)}$ time, and performs a single evaluation of the function f .*

3.3. The Linearity Test

How hard is it to tell whether a function is almost a linear form, i.e., $x \in \mathbf{F}_q^m \mapsto a^T x$, for some $a \in \mathbf{F}_q^m$? We restrict ourselves to the case $q = 2$, the only one of interest for our purposes. Consider a function $f : \mathbf{F}_2^m \mapsto \mathbf{F}_2$ such that, for random $x, y \in \mathbf{F}_2^m$, $\text{Prob}[f(x) + f(y) \neq f(x+y)] \leq \delta$, for some small enough $\delta > 0$. A simple argument shows that the function is 2δ -close to some linear form. By now, we trust that the reader can easily write a PC proof for linearity testing.

LEMMA 3.6. — *Given a function $f : \mathbf{F}_2^m \mapsto \mathbf{F}_2$, fix a small enough constant δ . If f is a linear form, then there exists a proof of that fact that the verifier always accepts. On the other hand, if its distance to any linear form exceeds δ , no proof is accepted with probability greater than δ . The verifier reads $O(1)$ bits of the proof, needs $O(m)$ random bits, and performs $O(1)$ evaluations of the function f .*

4. THE PCP THEOREM

The proof consists of three parts: the first two involve the design of suboptimal PCP systems for 3-SAT; the third provides a composition method that allows us to plug the two suboptimal schemes together to produce the desired PCP system.

4.1. Optimal Random Bit Complexity: $\text{NP} \subseteq \text{PCP} [\log n, (\log n)^{O(1)}]$

Let Φ be a 3-CNF formula consisting of m clauses C_1, \dots, C_n and n variables v_1, \dots, v_n . Since $m = O(n^3)$ and our PCP bounds in this section are all (poly)logarithmic, we might as well assume that $m = n$. We associate a polynomial with each clause in a fairly obvious way: $x_i(1 - x_j)x_k$ with $\neg v_i \vee v_j \vee \neg v_k$; plus seven other possibilities. It is clear that Φ is satisfiable if and only if all these n polynomials have a simultaneous zero over \mathbf{F}_q . Given a 0/1 assignment f of the x_i 's, we define a function $G^f : \{1, \dots, n\}^4 \mapsto \{0, 1\}$ as follows:

$$(2) \quad G^f(i, j, k, l) = \begin{cases} 0 & \text{if } v_i, v_j, v_k \text{ do not all appear in } C_l \text{ in that order;} \\ f(i)f(j)f(k) & \text{if } C_l \text{ is } \neg v_i \vee \neg v_j \vee \neg v_k; \\ f(i)f(j)(1 - f(k)) & \text{if } C_l \text{ is } \neg v_i \vee \neg v_j \vee v_k; \\ \text{etc.} & \text{(6 other cases).} \end{cases}$$

Φ is satisfiable if and only if there exists an assignment f such that the function G^f vanishes everywhere. To take advantage of the redundancy of polynomial maps mentioned earlier, we encode f and then G^f as polynomial maps.

Let $h = \Theta(\log n)$ and $m = \lceil h/\log h \rceil$, and define q to be a prime sufficiently larger than h^3m^4 . Without loss of generality, assume that $n = h^m$. Fix a bijection between $\{1, \dots, n\}$ and H^m , where $H = \{0, \dots, h - 1\}$; for example, write $i - 1$ in base h . From now on, we regard the index i of x_i as an element (y_1, \dots, y_m) of $H^m \subseteq \mathbf{F}_q^m$. The assignment f maps H^m to $\{0, 1\}$ and, by Lagrange interpolation, can be extended into a map defined by a polynomial of degree $(h - 1)m$ (still called f , for simplicity):

$$(3) \quad f(y_1, \dots, y_m) = \sum_{t_1, \dots, t_m \in H} f(t_1, \dots, t_m) \prod_{i=1}^m \prod_{u \in H \setminus \{t_i\}} \frac{y_i - u}{t_i - u}.$$

Similarly, we regard $G^f(i, j, k, l)$ as a function from H^{4m} to \mathbf{F}_q , which we extend into a polynomial G^f_{poly} in $\mathcal{F}_{q, 4m}^{9hm}$. This is how we do it. First, we express G^f in a more unified manner:

$$G^f(i_1, i_2, i_3, l) = \prod_{s=1}^3 a_s(i_s, l) (b_s(l) - f(i_s)),$$

with the obvious meaning of all these functions: s specifies one of the three literals in the clause C_l ; i_s is the index (viewed as an element of H^m) of the corresponding variable x_{i_s} and $f(i_s)$ is its 0/1 assignment; $b_s(l) = 0/1$ indicates if x_{i_s} is negated, etc. The polynomial

extension of G^f is defined as

$$G_{\text{poly}}^f(i_1, i_2, i_3, l) = \prod_{s=1}^3 a'_s(i_s, l) (b'_s(l) - f(i_s)),$$

where a'_s, b'_s are the polynomial extensions of a_s, b_s as defined in (3). It is immediate that G_{poly}^f is of degree less than $9hm$. (Note that we avoid Lagrange interpolation on G^f itself.)

PC PROOF THAT Φ IS SATISFIABLE

- [1] Present $f : \mathbf{F}_q^m \mapsto \mathbf{F}_q$ by values, where f is a polynomial extension of a satisfying assignment for Φ .
- [2] Write down the PC proof that f is ε -close to $\mathcal{F}_{q,m}^{hm}$. *(low-degree test)*
- [3] Form all R_i 's in Lemma 3.2 ($\ell = 4m$) and, for each $1 \leq i \leq q^\ell$, write down the PC proof that $\sum_{H^{4m}} G_{\text{poly}}^f R_i = 0$. *(sumcheck test)*

The verifier applies the low-degree test on f and rejects the proof if it fails. Otherwise, it picks a random $1 \leq i \leq q^{4m}$, and sumchecks that $\sum_{H^{4m}} G_{\text{poly}}^f R_i = 0$. It accepts the proof if this test succeeds and rejects it otherwise.

REMARK 4.1. — The verifier can compute R_i and a'_s, b'_s on its own and also evaluate G_{poly}^f anywhere by querying f at three places in the proof. Also, we said earlier that the sumcheck test requires that the verifier can trust its evaluations of f . But what if the prover cheated in the presentation of f ? This cannot happen: whatever is presented *is* what defines f .

Why does this protocol work? As usual, we do not have false negatives. If Φ is satisfiable, then the prover only has to stick to the scenario above and all tests will succeed. Suppose now that Φ cannot be satisfied. For the usual reasons, we assume that all tests (i.e., low-degree and sumcheck) succeed. We distinguish between three cases:

- (1) $f \in \mathcal{F}_{q,m}^{hm}$: Then, G_{poly}^f is a polynomial of degree at most $9hm$, which is nonzero because Φ is not satisfiable. So, by Lemma 3.2, the probability that the sum to check is 0 is at most $4hm/q$. Assume that it is not 0. The degree of $G_{\text{poly}}^f R_i$ does not exceed $13hm$; therefore, by Lemma 3.1, the probability that the sumcheck test succeeds is $O(hm^2/q)$. This bounds the probability of failure in this case by $O(hm^2/q)$.
- (2) f is ε -close (but not 0-close) to $\mathcal{F}_{q,m}^{hm}$: Let f_o be a nearest neighbor in $\mathcal{F}_{q,m}^{hm}$. The sumcheck test requires a single evaluation of G_{poly}^f and, hence, evaluations of f at three points. The probability that f and f_o agree at all three points is at least $1 - 3\varepsilon$. The agreement means that the sumcheck test is, in effect, carried out on $G_{\text{poly}}^{f_o} R_i$. The previous case now shows that the failure probability is at most $3\varepsilon + O(hm^2/q)$.

- (3) f is not ε -close to $\mathcal{F}_{q,m}^{hm}$: By Lemma 3.4, the low-degree test will fail to catch that fact with probability at most ε .

To summarize, the verifier might fail to spot an inconsistency with probability $O(\varepsilon + hm^2/q) < 1/2$, for a small enough constant $\varepsilon > 0$. Since $q = O(m^4 h^3)$, the number of proof bits read is $O(hm^2 \log q) = (\log n)^{O(1)}$, the running time is $(hm)^{O(1)} = (\log n)^{O(1)}$, the number of random bits needed is $O(m \log q) = O(\log n)$ (the motivation for our choice of h), and the number of places at which the candidate assignment function is evaluated is $O(1)$. This proves that, indeed, $\text{NP} \subseteq \text{PCP}[\log n, (\log n)^{O(1)}]$.

4.2. Optimal Query Bit Complexity: $\text{NP} \subseteq \text{PCP}[n^{O(1)}, 1]$

Adding random bits allows the verifier to do with fewer lookups; in fact, a constant number of them. The strategy is roughly the same as before. Since the verifier is given access to only $O(1)$ bits of the proof, however, the ground field must be of constant size, so we set $q = 2$. Let Φ be a 3-CNF formula consisting of m clauses C_1, \dots, C_m and n variables v_1, \dots, v_n . As usual, we model each clause, say, $\neg v_i \vee v_j \vee \neg v_k$ as $x_i(1 - x_j)x_k$. This defines m cubic polynomials G_1, \dots, G_m , and Φ is satisfiable if and only if all the G_i 's have a common zero over \mathbf{F}_2 . Given $r = (r_1, \dots, r_m) \in \mathbf{F}_2^m$, let $F_r(x) = \sum r_i G_i(x)$. Our interest in F_r comes from this (trivial) fact:

LEMMA 4.2. — *If Φ is satisfiable, all 2^m polynomials F_r have a common zero; otherwise, given any $a \in \mathbf{F}_2^n$ and a random $r \in \mathbf{F}_2^m$, $F_r(a) = 0$ with probability $1/2$.*

The PC proof of satisfiability is based on this simple test. The prover wants to convince the verifier that it knows a common zero a to all the F_r 's (whether that is true or not). To do that, the proof will list the values of $F_r(a)$, for all r , so that the verifier can test that, indeed, $F_r(a) = 0$. The prover must also provide a consistency check that satisfies the verifier that its evaluations of F_r are correct. The cubic polynomial F_r can be written as $F_r(x) = f_r + \sum_i f_r^i x_i + \sum_{i,j} f_r^{ij} x_i x_j + \sum_{i,j,k} f_r^{ijk} x_i x_j x_k$. The verifier can evaluate $F_r(a)$ by using the three linear forms

$$(4) \quad H_1^a(y) = \sum_i a_i y_i; \quad H_2^a(y) = \sum_{i,j} a_i a_j y_{ij}; \quad H_3^a(y) = \sum_{i,j,k} a_i a_j a_k y_{ijk}.$$

Forgive our (abusive) notation y to refer to a set of n , n^2 , and n^3 labeled variables, respectively. The PC proof will present each H_i^a by values ($i = 1, 2, 3$). Of course, no guarantee exists that the prover will not corrupt the presentation. To catch any cheating, the verifier relies on two sets of functional equations: the H_i^a 's are (i) linear and (ii) related by the identities

$$(5) \quad H_2^a(y \otimes y') = H_1^a(y)H_1^a(y') \quad \text{and} \quad H_3^a(y \otimes z) = H_1^a(y)H_2^a(z),$$

where $y, y' \in \mathbf{F}_2^n$ and $z \in \mathbf{F}_2^{n^2}$. If $y \in \mathbf{F}_2^s$ and $z \in \mathbf{F}_2^t$, then $y \otimes z$ denotes the vector $(y_i z_j) \in \mathbf{F}_2^{st}$.

PC PROOF THAT Φ IS SATISFIABLE

Present the functions H_1^a, H_2^a, H_3^a by values, where a is a common zero to G_1, \dots, G_m .

The verifier performs three basic sets of tests. Fix some small constant $\varepsilon > 0$.

- The first test is to check the linearity of each H_i^a , by using the criterion of Lemma 3.6, with $\delta = \varepsilon^2$. The verifier rejects the proof if any of these 3 tests fail. From now on, *any* evaluation of $H_i^a(y)$ is to be immediately confirmed by the following test: pick a random y' and check that $H_i^a(y) = H_i^a(y + y') - H_i^a(y')$. If this test ever fails, the proof is rejected.
- Next, the verifier checks that the H_i^a 's are related by the two identities (5). Each one is tested $O(1/\varepsilon)$ times for random pairs (y, y') and (y, z) , where $y, y' \in \mathbf{F}_2^n$ and $z \in \mathbf{F}_2^{n^2}$. Again, any failure implies rejection of the proof.
- Finally, the verifier picks a random r and evaluates $F_r(a)$. To do that, it computes on its own $y_1 = (f_r^i)$, $y_2 = (f_r^{ij})$, $y_3 = (f_r^{ijk})$, as well as f_r , and then looks up the proof at three places to compute the sum $F_r(a) = f_r + H_1^a(y_1) + H_2^a(y_2) + H_3^a(y_3)$. If $F_r(a) = 0$ and none of the previous tests have failed, the verifier accepts the claim that Φ is satisfiable.

Why does this work? If Φ is satisfiable, then the proof needs simply to conform to the directives of the verifier and it will be accepted. Suppose that Φ is not satisfiable and that, by contradiction, the proof is accepted. What is the probability of failure? If the linearity tests passes then, by Lemma 3.6, with probability at least $1 - 3\varepsilon^2$, there exists a linear form \widehat{H}_i , for $i = 1, 2, 3$, that disagrees with H_i^a over a fraction at most ε^2 of its domain. This means that, with probability $1 - O(\varepsilon)$, we can assume that all identity tests are performed with respect to the true values of \widehat{H}_i . If either identity fails to be satisfied by the \widehat{H}_i 's, then by Lemma 4.3 (and its omitted analog for the first identity) a conservative estimate of $O(\varepsilon)$ bounds the probability that the verifier fails to catch that fact. Therefore, with probability $1 - O(\varepsilon)$, the value $F_r(a)$ computed by the verifier is, indeed, $f_r + \widehat{H}_1(y_1) + \widehat{H}_2(y_2) + \widehat{H}_3(y_3)$, for some linear forms \widehat{H}_i defined by some vector a in accordance with the format specified by (4). Since Φ is not satisfiable, we know by Lemma 4.2 that the value of $F_r(a)$ is zero with probability $1/2$. Therefore, the verifier will accept a wrong proof with probability $1/2 + O(\varepsilon)$. By setting ε to a small enough constant and repeating the verification, we bring the failure probability below $1/2$. The number of random bits is $O(n^3)$ and the number of bit lookups in the proof is constant. This concludes the proof that $\text{NP} \subseteq \text{PCP}[n^{O(1)}, 1]$.

LEMMA 4.3. — *If $\widehat{H}_3(y \otimes z) - \widehat{H}_1(y)\widehat{H}_2(z)$ is nonzero then, with probability $1/4$, it evaluates to 1 at random y, z .*

4.3. Self-Reduction: $\text{NP} = \text{PCP} [\log n, 1]$

We have built two PCP systems: one, S_1 , needs $O(\log n)$ random bits; the other, S_2 , uses $O(1)$ queries. We now combine them to extract the best feature from each. The basic idea is simple. We caught a glimpse of it in Section 2. Recall the action of the verifier V_1 for S_1 . First, it generates a random bit string s ; then, it computes a set of addresses i_1, \dots, i_q to look up in the proof Π , where $q = (\log n)^{O(1)}$. Upon reading the corresponding bits, $\Pi_{i_1}, \dots, \Pi_{i_q}$, the verifier evaluates a predicate in time $q^{O(1)}$ to decide whether to accept or not.

Now comes the self-reducibility part. By the Cook-Levin theorem the predicate in question can be expressed as a 3-CNF formula Φ_s of size $q^{O(1)}$. The verifier accepts if and only if there exists a string X_s such that the concatenated string $\Pi_{i_1} \cdots \Pi_{i_q} X_s$ forms a satisfying truth assignment for Φ_s . The key idea is that a PCP system such as S_2 is exactly the sort of thing that can be used to check the satisfiability of Φ_s . So, instead of computing the verification predicate itself, V_1 can hand the problem over to the PC proof system S_2 . Its verifier V_2 will then consult its own proof to check whether Φ_s is satisfiable, and will accept or reject accordingly. A minor technical point: Of course, we cannot let *both* V_1 and V_2 err with probability $1/2$. By repeating the verifiers' runs (the standard trick), we can trivially lower the odds of an error to any small constant.

Even though the number of bits read in the proof can be arbitrary, it is important for composition purposes that the number of *entries* be $O(1)$. Polynomial extensions give us a convenient tool for achieving that. A proof Π of length N can be viewed as a function $f : \{1, \dots, N\} \mapsto \{0, 1\}$, where $f(i) = \Pi_i$. As we did in Section 4.1, we can change our point of view and regard f as a function from H^m to $\{0, 1\}$, where H^m is in bijection with $\{1, \dots, N\}$ and $H = \{0, \dots, h-1\}$, for some parameters h, m such that $N = h^m$. (Pad the proof with junk if N is inexpressible in this way.) Let \hat{f} be a polynomial extension of f in $\mathcal{F}_{q,m}^{hm}$. The proof Π is now rewritten as a presentation of \hat{f} by values, with all the bells and whistles needed to apply the low-degree test and the recovery mechanism (Lemmas 3.4 and 3.5). The verifier applies the low-degree test to check that the presentation is δ -close to $\mathcal{F}_{q,m}^{hm}$ (for some suitably small $\delta > 0$), and rejects the proof if it is not. Otherwise, it appeals to Lemma 3.5 to evaluate f at $k (= q)$ points by a single evaluation of \hat{f} . In this way, the verifier can gain access to $\Pi_{i_1} \cdots \Pi_{i_q}$ in $O(1)$ queries to the new proof. Note that an entry in this new proof is no longer a single bit but a field element represented as a bit string. For simplicity, we still call the new proof Π : the difference is that now $q = O(1)$.

The benefits of composing proof systems are now obvious. Let us try our hand at composing S_1 with itself, i.e., S_1 with S_2 , where S_2 denotes S_1 . The verifiers for S_1 and S_2 need $O(\log n)$ and $O(\log(\log n)^{O(1)})$ random bits, respectively, i.e., a total of $O(\log n)$ of them. Obviously, the number of queries remains $O(1)$. All the verification work, being now done by S_2 , amounts to $(\log(\log n)^{O(1)})^{O(1)}$, i.e., $(\log \log n)^{O(1)}$. This bound can be further reduced by iterating the composition, but this is not the way to go to make it

$O(1)$. For that, we take the previous system, call it S_3 , and compose it with S_2 . This requires $O(\log n) + (\log \log n)^{O(1)} = O(\log n)$ random bits, $O(1)$ queries and $O(1)$ amount of work. It follows that $O(1)$ bits are read in the proof, and the PCP theorem is proven.

But is it really? The task of V_1 is not *only* to check that Φ_s is satisfiable but that $\Pi_{i_1} \cdots \Pi_{i_q}$ is part of a satisfying assignment. Here is a simple illustration of the conundrum we face. Say, a prover claims to have a satisfying assignment for $\bigwedge_i C_i$. A verifier might want to check this by picking C_i at random and verifying that the assignment makes C_i true. But suppose that, instead, it chooses to delegate that task to some other PC proof system. A second verifier will then take C_i as input and check that it is satisfiable. But any disjunction of three literals, such as C_i , is always satisfiable. What needs to be checked is not whether C_i is satisfiable on its own, but whether it is by using the assignment *specified* by S_1 . Returning to V_2 , its job is not to check that Φ_s is satisfiable but that there exists X_s such that $\Pi_{i_1} \cdots \Pi_{i_q} \cdot X_s$ makes Φ_s true. To resolve this *consistency* issue is key to making self-reducibility work. This is easy to do; in fact, by reducing the number of queries to $O(1)$, we have done the hardest part already. We sketch what remains to be done.

For the verification to be delegated to V_2 , of course, it is necessary to encode $\Pi_{i_1} \cdots \Pi_{i_q} X_s$ into the format $\sigma(\Pi_{i_1} \cdots \Pi_{i_q} X_s)$ that V_2 expects. It might be tempting to simply append $\sigma(\Pi_{i_1} \cdots \Pi_{i_q} X_s)$ at the end of Π , but doing so would raise the consistency problem mentioned earlier. Instead, we must effectively *replace* Π (and not just add on to it) with the encoding σ of every possible string $\Pi_{i_1} \cdots \Pi_{i_q} X_s$. But to do so would cause the same Π_{i_j} to appear in different encodings throughout the proof, again raising consistency issues. The solution is to encode each Π_{i_j} separately. Specifically we replace each Π_{i_j} by $\sigma(\Pi_{i_j})$. Likewise, we encode X_s as $\sigma(X_s)$.

This solves one problem, consistency, only to create another one. In this scheme, V_2 does not have access to $\sigma(\Pi_{i_1} \cdots \Pi_{i_q} X_s)$, which is the only encoding it can read, but to $\sigma(\Pi_{i_1}) \cdots \sigma(\Pi_{i_q}) \sigma(X_s)$. Is that good enough? Instead of encoding a whole truth assignment $a_1 \cdots a_n$ via σ , suppose we encode it in chunks: first $\sigma(a_1 \cdots a_{i_1})$, then $\sigma(a_{i_1+1} \cdots a_{i_2})$, etc, and finally $\sigma(a_{i_{q-1}+1} \cdots a_n)$. Can the verifier deal with that sort of *split form* encoding? The answer is yes. It is, in fact, a rather simple exercise to modify V_2 accordingly.

This completes the proof of the PCP theorem or, at least, of its conceptual outline. The doubting reader can always sample the proof at random and see if that helps...

5. HISTORICAL NOTES

Following the seminal work of Goldwasser, Micali, and Rackoff [20] and Babai [5], which introduced the notion of interactive proofs, an important variant was introduced by Ben-Or et al. [9], in which the verifier interacts with not one but several provers. This framework led to early incarnations of PCP systems by Fortnow, Rompel and Sipser [14].

The idea of putting tight resource restrictions on both the verifier (query time) and the proof (size) originated in the works of Babai et al. [6] and Feige et al. [13]. The algebraic view of Boolean expressions gained currency in a series of papers that highlighted the enormous expressive power of interactive proofs [7, 24]. Turning to co-NP, Lund et al. [24] explained how a prover can convince a verifier that a graph is *not* 3-colorable. (By contrast, to convince someone that a graph is 3-colorable is trivial.) Finally, the ultimate power of interaction was resolved by Shamir [30], who proved that languages with interactive proofs are precisely those that can be decided in polynomial space.

The current notion of PCP itself, with its focus on randomness and query complexity, was formally introduced by Arora and Safra [4]. This development was spurred in large measure by the key insight of Feige et al. [13], which for the first time tied probabilistic proof systems to inapproximability. Babai, Fortnow, and Lund [7] established that PCP [$n^{O(1)}, n^{O(1)}$] coincides with the class of problems solvable in nondeterministic exponential time. Babai et al. [6] and Feige et al. [13] essentially showed that NP is contained in PCP [polylog, polylog] (the precise bounds being somewhat stronger). Arora and Safra [4] proved that $\text{NP} = \text{PCP} [\log n, \sqrt{\log n}]$, and introduced the powerful concept of proof composition. The PCP theorem itself, i.e., $\text{NP} = \text{PCP} [\log n, 1]$, was proven by Arora et al. [3]. Finetuning the constants followed in quick order. Håstad [23] proved the striking result that three queries are sufficient as long as we can tolerate an ε chance of rejecting a correct proof. Building on that result, Guruswami et al. [21] showed that such false-negatives can be avoided provided that the error probability for wrongly accepting is $1/2 + \varepsilon$.

The connection to inapproximability [13] blossomed into a plethora of hardness results, one of the most impressive being Håstad's proof [22] that to approximate the clique number of an n -node graph within a factor of $n^{1-\varepsilon}$ is impossible (unless NP coincides with the randomized version of P, i.e., the class of sets for which membership can be decided in expected polynomial time by a randomized, error-free Turing machine). At the other end of the spectrum, consider MaxCut, the problem of partitioning the node set of a graph into two subsets with the maximum number of edges joining them. It is possible to find a solution in polynomial time that has a number of edges at least 0.878 times the maximum possible [17]. On the other hand, to push that approximation factor above 0.942 would require that $\text{P} = \text{NP}$ [23, 33] (building on [8]). A comprehensive 1996 survey of approximation results was compiled by Arora and Lund [2].

Many of the tools for checking the internal consistency of proof systems originated in the area of *program checking* [10, 11, 29]. For example, the low degree test, due to Arora et al. [3], incorporates ideas from [4, 11, 16, 28]. The sumcheck and linearity tests are due respectively to Lund et al. [24] and Blum, Luby, and Rubinfeld [11]. Testing that a polynomial is nonzero (Lemma 3.2) is from [6, 13]. Essential tools in PCP-related work

also include the *parallel repetition theorem* by Raz [27], the *long code* by Bellare, Goldreich and Sudan [8], and Fourier transform techniques by Håstad [23]. For background material in complexity theory, the following texts [15, 31, 26, 12], listed in increasing order of technical depth, are excellent entry points. We also mention [1, 2, 25, 32] for in-depth coverage of proof verification and approximation algorithms, and [18, 19, 20] for an introduction to zero-knowledge cryptography.

ACKNOWLEDGMENTS — I wish to thank Lance Fortnow, Oded Goldreich, and Muli Safra for their helpful comments on this manuscript.

REFERENCES

- [1] S. ARORA — *Probabilistic Checking of Proofs and the Hardness of Approximation Problems*, Ph.D. Thesis, UC Berkeley, 1994. Also available as <http://www.cs.princeton.edu/~arora>
- [2] S. ARORA & C. LUND — *Hardness of approximations*, in “Approximation Algorithms for NP-hard Problems,” ed. D. Hochbaum, PWS Publishing, 1996.
- [3] S. ARORA, C. LUND, R. MOTWANI, M. SUDAN, & M. SZEGEDY — *Proof verification and the hardness of approximation problems*, J. ACM 45 (1998), 501–555.
- [4] S. ARORA & M. SAFRA — *Probabilistic checking of proofs: a new characterization of NP*, J. ACM 45 (1998), 70–122.
- [5] L. BABAI — *Trading group theory for randomness*, Proc. 17th Annual ACM Symp. Theory Comput. (1985), 421–429.
- [6] L. BABAI, L. FORTNOW, L. LEVIN & M. SZEGEDY — *Checking computations in polylogarithmic time*, Proc. 23rd Annual ACM Symp. Theory Comput. (1991), 21–31.
- [7] L. BABAI, L. FORTNOW & C. LUND — *Non-deterministic exponential time has two-prover interactive protocols*, Computational Complexity 1 (1991), 3–40.
- [8] M. BELLARE, O. GOLDREICH & M. SUDAN — *Free bits, PCPs and non-approximability — towards tight results*, SIAM J. Comput. 27 (1998), 804–915.
- [9] M. BEN-OR, S. GOLDWASSER, J. KILIAN & A. WIGDERSON — *Multi-prover interactive proofs: how to remove intractability*, Proc. 20th Annual ACM Symp. Theory Comput. (1988), 113–131.
- [10] M. BLUM & S. KANNAN — *Designing programs that check their work*, Proc. 21st Annual ACM Symp. Theory Comput. (1989), 86–97.
- [11] M. BLUM, M. LUBY & R. RUBINFELD — *Self-testing/correcting with applications to numerical problems*, J. Comp. Sys. Sci. 47 (1993), 549–595.
- [12] D.-Z. DU & K.-I. KO — *Theory of Computational Complexity*, Wiley-Interscience, 2000.

- [13] U. FEIGE, S. GOLDWASSER, L. LOVASZ, S. SAFRA & M. SZEGEDY – *Interactive proofs and the hardness of approximating cliques*, J. ACM 43 (1996), 268–292.
- [14] L. FORTNOW, J. ROMPEL & M. SIPSER – *On the power of multi-prover interactive protocols*, Theoret. Comput. Sci. 134 (1994), 545–557.
- [15] M. GAREY & D. JOHNSON – *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York, 1979.
- [16] P. GEMMELL, R. LIPTON, R. RUBINFELD, M. SUDAN & A. WIGDERSON – *Self-testing/correcting for polynomials and approximate functions*, Proc. 23rd Annual ACM Symp. Theory Comput. (1991), 32–42.
- [17] M.X. GOEMANS & D.P. WILLIAMSON – *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM 42 (1995) 1115–1145.
- [18] O. GOLDREICH – *Modern Cryptography, Probabilistic Proofs and Pseudorandomness*, Algorithms and Combinatorics 17, Springer, 1999.
- [19] O. GOLDREICH, S. MICALI & A. WIGDERSON – *Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems*, J. ACM 38 (1991), 691–729.
- [20] S. GOLDWASSER, S. MICALI & C. RACKOFF – *The knowledge complexity of interactive proof systems*, SIAM J. Comput. 18 (1989), 186–208.
- [21] V. GURUSWAMI, D. LEWIN, M. SUDAN & L. TREVISAN – *A tight characterization of NP with 3 query PCPs*, Proc. 39th Annual IEEE Symp. Found. Comput. Sci. (1998), 8–17. Also available as ECCC Technical Report TR98-034.
- [22] J. HÅSTAD – *Clique is hard to approximate within $n^{1-\epsilon}$* , Acta Mathematica 182 (1999) 105–142.
- [23] J. HÅSTAD – *Some optimal inapproximability results*, Proc. 29th Annual ACM Symp. Theory Comput. (1997), 1–10. Also available as ECCC Technical Report TR97-037.
- [24] C. LUND, L. FORTNOW, H. KARLOFF & N. NISAN – *Algebraic methods for interactive proof systems*, J. ACM 39 (1992), 859–868.
- [25] E.W. MAYR, H.J. PROMEL & A. STEGER – *Lectures on Proof Verification and Approximation Algorithms*, LNCS 1367, Springer Verlag, 1998.
- [26] C.H. PAPADIMITRIOU – *Computational Complexity*, Addison Wesley, 1994.
- [27] R. RAZ – *A parallel repetition theorem*, SIAM J. Comput. 27 (1998), 763–803.
- [28] R. RUBINFELD & M. SUDAN – *Self-testing polynomial functions efficiently and over rational Domains*, Proc. 3rd Annual ACM/SIAM Symp. Discrete Algorithms (1992), 23–32.
- [29] R. RUBINFELD & M. SUDAN – *Robust characterizations of polynomials with applications to program testing*, SIAM J. Comput. 25 (1996), 252–271.
- [30] A. SHAMIR – *IP = PSPACE*, J. ACM 39 (1992), 869–877.

- [31] M. SIPSER – *Introduction to the Theory of Computation*, PWS Publishing, 1997.
- [32] M. SUDAN – *Efficient Checking of Polynomials and Proofs and the Hardness of Approximation Problems*, ACM Distinguished Dissertation Series for 1993, Springer, 1996.
- [33] L. TREVISAN, G.B. SORKIN, M. SUDAN & D.P. WILLIAMSON – *Gadgets, Approximation, and Linear Programming*, SIAM J. Comput. 29 (2000), 2074-2097.

Bernard CHAZELLE

Princeton University

Department of Computer Science

PRINCETON, NJ 08544

USA

E-mail : `chazelle@cs.princeton.edu`

IWASAWA ALGEBRAS AND ARITHMETIC

by **John COATES**

INTRODUCTION

Let p be a prime number, and G a compact p -adic Lie group. We recall that the Iwasawa algebra of G is defined by

$$\Lambda(G) = \varprojlim_U \mathbb{Z}_p[G/U]$$

where U runs over the open normal subgroups of G . Any compact \mathbb{Z}_p -module on which G acts continuously on the left has a unique structure as a left $\Lambda(G)$ -module, extending the G -action. Thanks to this remark, modules over $\Lambda(G)$, where G is usually the image of Galois in a finite dimensional p -adic Galois representation, abound in arithmetic geometry. K. Iwasawa [Iw] was the first to study the structure theory of finitely generated $\Lambda(G)$ -modules in the special case when $G = \mathbb{Z}_p$, and deduced from it his celebrated asymptotic formula for the growth of the order of the p -primary subgroup of the ideal class group in a \mathbb{Z}_p -extension of a number field. Almost immediately, J-P. Serre [Se1], [Se2] pointed out that, when $G = \mathbb{Z}_p^d$ for any integer $d \geq 1$, $\Lambda(G)$ is isomorphic to the local ring $\mathbb{Z}_p[[T_1, \dots, T_d]]$ of formal power series in d variables with coefficients in \mathbb{Z}_p , and that Iwasawa's structure theorem for $\Lambda(G)$ -modules could be re-proven for $d = 1$, and generalized to all $d \geq 1$, by using classical arguments in commutative algebra about the structure theory of modules up to pseudo-isomorphism ([B-CA], Chap. VII, §4.4, Theorems 4 and 5).

Intuitively, one might expect that the structure theory for $\Lambda(G)$ -modules would be very different in the commutative and non-commutative cases, but the aim of this seminar is to report on joint work of the author, P. Schneider and R. Sujatha [CSS] proving that, surprisingly, the two cases appear to be parallel in many ways. The first step towards elucidating the non-commutative theory was made by O. Venjakob [Ve1], [Ve2], who exploited ideas of J. Björk [Bj] to define in general the notion of a pseudo-null $\Lambda(G)$ -module. If G is pro- p and has no element of order p , Venjakob defines a finitely generated left $\Lambda(G)$ -module M to be pseudo-null if it is $\Lambda(G)$ -torsion (i.e. each element of M is

annihilated by some non-zero element of $\Lambda(G)$), and, in addition, $\text{Ext}_{\Lambda(G)}^1(M, \Lambda(G)) = 0$. To establish our structure theory up to pseudo-isomorphism, we need to impose further conditions on G , and we are grateful to B. Totaro for pointing out to us that probably the most natural hypothesis is that G should possess a p -valuation in the sense of M. Lazard [La]. We recall that a p -valuation on G is a function $\omega : G \rightarrow (0, \infty]$ satisfying the following axioms for all x and y in G :

- (i) $\omega(1) = \infty$, and $\frac{1}{p-1} < \omega(x) < \infty$ for $x \neq 1$;
- (ii) $\omega(xy^{-1}) \geq \min\{\omega(x), \omega(y)\}$;
- (iii) $\omega(x^{-1}y^{-1}xy) \geq \omega(x) + \omega(y)$;
- (iv) $\omega(x^p) = \omega(x) + 1$.

We say that G is p -valued if it possesses a p -valuation. If G is p -valued, we remark that the compactness of G guarantees that G is complete with respect to the p -valuation ω in the following sense. For each $u > 0$, let G_u denote the subgroup of G consisting of all g such that $\omega(g) \geq u$. As J-P. Serre observed to us, G_u is open in G because, choosing $N > u$, G_u contains the subgroup of G generated by the p^N -th powers, and it is well known that this latter subgroup is a neighbourhood of the identity in a p -adic Lie group. Hence the family $\{G_u : u > 0\}$ form an open basis for the topology of G since their intersection is trivial, and the natural map from G to $\varprojlim G/G_u$ is an isomorphism because of the compactness of G . Moreover, Lazard [La] established the following basic facts. Any closed subgroup of a p -valued group is also p -valued. If G is p -valued, then it is pro- p , and has no element of order p . The classic example of a p -valued group is the group of matrices in $GL_n(\mathbb{Z}_p)$ which are congruent to the identity modulo p (resp. mod 4) if p is odd (resp. if $p = 2$). If $p > n + 1$, any pro- p closed subgroup of $GL_n(\mathbb{Z}_p)$ is p -valued.

THEOREM 1.1 ([CSS]). — *Let G be a p -valued compact p -adic Lie group, and let M be a finitely generated torsion $\Lambda(G)$ -module. Let M_0 be the maximal pseudo-null submodule of M . Then there exist non-zero left ideals L_1, \dots, L_m , and a $\Lambda(G)$ -injection*

$$\phi : \bigoplus_{i=1}^m \Lambda(G)/L_i \longrightarrow M/M_0,$$

with Coker (ϕ) pseudo-null.

The special case of Theorem 1.1 in which M/M_0 is killed by some power of p was proven earlier by O. Venjakob [Ve1], [Ve2], and S. Howson [Ho]. In §2, we shall give a sketch of a proof of Theorem 1.1 taken from [CSS], which is remarkably parallel to the classical commutative proof in [B-CA], and which exploits the fact that $\Lambda(G)$ is a filtered ring to which one can apply the techniques of the algebraic theory of micro-localization (see, for example, [LO]). After finding this proof, we also realized that Theorem 1.1 can be derived from the work of M. Chamarie [Ch1], [Ch2], on modules over maximal orders (see [CSS] for the details).

We assume for the rest of this exposé that G is a p -valued compact p -adic Lie group. In particular, it follows that $\Lambda(G)$ is Noetherian, and has no zero divisors. Let $\text{Mod}(G)$ denote the category of all finitely generated left $\Lambda(G)$ -modules, and $C^1(G)$ the subcategory whose objects are the pseudo-null modules ($C^1(G)$ is closed under taking subobjects, quotients, and extensions). To discuss questions about the uniqueness of the decomposition in Theorem 1.1, we have to pass to the quotient category

$$\mathbf{M}(G) = \text{Mod}(G)/C^1(G).$$

We write $Q : \text{Mod}(G) \rightarrow \mathbf{M}(G)$ for the canonical functor. If M is an object of $\text{Mod}(G)$, we define its annihilator, which we denote by $\text{ann}_{\Lambda(G)}(M)$, to be the two sided ideal consisting of all r in $\Lambda(G)$ such that $r.M = 0$. We then define the annihilator of the object $Q(M)$ of the quotient category $\mathbf{M}(G)$, which we denote by $\text{ann}(Q(M))$, to be the sum of all the ideals $\text{ann}_{\Lambda(G)}(N)$, where N runs over all objects of $\text{Mod}(G)$ such that $Q(N)$ is isomorphic to $Q(M)$ in $\mathbf{M}(G)$. In fact, a lemma of Robson [Ro] shows that

$$\text{ann}(Q(M)) = \text{ann}_{\Lambda(G)}(M/M_0),$$

where, as above, M_0 denotes the maximal pseudo-null submodule of M . Yet another description of $\text{ann}(Q(M))$ can be given in terms of the left ideals L_1, \dots, L_m appearing in Theorem 1.1. Let J_i be the maximal two-sided ideal of $\Lambda(G)$ which is contained in L_i , and let $J = \bigcap_{i=1}^m J_i$. Then $J_i = \text{ann}_{\Lambda(G)}(\Lambda(G)/L_i)$, and we have

$$J = \text{ann}(Q(M));$$

in particular, we see that $J \neq 0$ if and only if $J_i \neq 0$ for $i = 1, \dots, m$.

It is in questions of annihilators that we find a basic difference between the commutative and non-commutative case. R. Greenberg (unpublished) has given an example of a p -valued open subgroup of $GL_2(\mathbb{Z}_p)$, and a finitely generated torsion $\Lambda(G)$ -module M such that $\text{ann}(Q(M)) = 0$. Following Chamarie [Ch2], we therefore define $Q(M)$ to be **bounded** (resp. **completely faithful**) if $\text{ann}(Q(M)) \neq 0$ (resp. if $\text{ann}(Q(N)) = 0$ for every torsion $\Lambda(G)$ -module N such that $Q(N)$ is a non-zero subquotient of $Q(M)$). It is proven in [Ch2] that, for every finitely generated torsion $\Lambda(G)$ -module M , we have a canonical decomposition

$$Q(M) = Q(U) \oplus Q(V),$$

where $Q(U)$ is completely faithful and $Q(V)$ is bounded. Very little is known about completely faithful objects in $\mathbf{M}(G)$ beyond the fact that Greenberg's example shows that they exist, and also it is shown in [Ch2] that they are cyclic, i.e. isomorphic in $\mathbf{M}(G)$ to $Q(\Lambda(G)/L)$ where L is a non-zero left ideal. However, Y. Hachimori and O. Venjakob [HV] have recently given examples of completely faithful $\Lambda(G)$ -modules which arise naturally in arithmetic geometry, and one suspects that their occurrence in number theory may be rather common.

We write $\mathbf{M}_b(G)$ for the full subcategory of $\mathbf{M}(G)$ consisting of the bounded objects, and we define $\mathcal{D}_b(G)$ to be the Grothendieck group of $\mathbf{M}_b(G)$. In fact, $\mathbf{M}_b(G)$ is also an abelian category in which every object has finite length, and the Jordan-Hölder theorem shows that $\mathcal{D}_b(G)$ is the free abelian group on the set of isomorphism classes of simple objects in $\mathbf{M}_b(G)$. It is natural to ask whether we can relate $\mathcal{D}_b(G)$ to a natural group of divisors of the ring $\Lambda(G)$, parallel to the classical theory for commutative, integrally closed, integral domains ([B-CA], Chap. VII, §4.5, Proposition 11). As we shall now explain, this is indeed the case. Let $K(G)$ denote the skew field of fractions of $\Lambda(G)$, which is well known to exist because $\Lambda(G)$ is Noetherian and has no divisors of zero. Then $\Lambda(G)$ is a **maximal order** (this is the non-commutative analogue of being integrally closed) in the sense that, if B is any intermediate ring with $\Lambda(G) \subset B \subset K(G)$ such that there exist non-zero elements u, v in $K(G)$ with $uBv \subset \Lambda(G)$, then necessarily $B = \Lambda(G)$ (see [CSS], Lemma 2.6). For any left (resp. right) $\Lambda(G)$ -module M , we put $M^* = \text{Hom}_{\Lambda(G)}(M, \Lambda(G))$ for the dual right (resp. left) $\Lambda(G)$ -module, and we say M is **reflexive** if the natural map from M to M^{**} is an isomorphism. A non-zero left (resp. right) $\Lambda(G)$ -submodule L of $K(G)$ is called a **fractional** left (resp. right) ideal if there is a non-zero v in $K(G)$ such that $L \subset \Lambda(G)v$ (resp. $L \subset v\Lambda(G)$). A **fractional ideal** of $\Lambda(G)$ is a subset I of $K(G)$ which is both a fractional left and a fractional right ideal. Finally, we define a fractional c -ideal of $\Lambda(G)$ to be a reflexive fractional ideal of $\Lambda(G)$, and we write $\mathcal{C}(G)$ for the set of fractional c -ideals. As a special case of general results about maximal orders, Asano ([As]) has shown that $\mathcal{C}(G)$ is an abelian group with respect to the product $I.J = (IJ)^{**}$. We recall that a two-sided ideal \mathfrak{p} of $\Lambda(G)$ is said to be **prime** if, whenever x and y are elements of $\Lambda(G)$ such that $x\Lambda(G)y \subset \mathfrak{p}$, we always have x is in \mathfrak{p} or y is in \mathfrak{p} . It is then also proven in [As] that $\mathcal{C}(G)$ is the free abelian group on the set \mathcal{P} of all non-zero prime c -ideals, and that every prime c -ideal has height 1 (i.e. is a minimal non-zero prime ideal). There would be great interest in giving an explicit description of this set \mathcal{P} (for example, when G is a p -valued open subgroup of $SL_2(\mathbb{Z}_p)$).

Our aim is to construct a canonical homomorphism

$$\chi : \mathcal{D}_b(G) \rightarrow \mathcal{C}(G),$$

and for this we need to localize $\Lambda(G)$ at the prime ideals in \mathcal{P} . We recall that a multiplicatively closed subset S of non-zero elements of $\Lambda(G)$ is said to be a right and left Ore set if, for each s in S and a in $\Lambda(G)$ both $aS \cap s\Lambda(G)$ and $Sa \cap \Lambda(G)s$ are non-empty. For each \mathfrak{p} in \mathcal{P} , let $S(\mathfrak{p})$ denote the set of all elements of $\Lambda(G)$ whose residue class in $\Lambda(G)/\mathfrak{p}$ is not a zero divisor. Chamarie [Ch1] has proven that $S(\mathfrak{p})$ is a left and right Ore set, and that the localization of $\Lambda(G)$ by $S(\mathfrak{p})$, which we denote by $\Lambda(G)_{\mathfrak{p}}$, is a bounded maximal order, with Jacobson radical $\mathfrak{p}\Lambda(G)_{\mathfrak{p}}$. Moreover, every left and right ideal in $\Lambda(G)_{\mathfrak{p}}$ is principal. Now, up to isomorphism, the objects of finite length in $\mathbf{M}_b(G)$ are of the form $Q(M)$, where M is a finitely generated torsion $\Lambda(G)$ -module. Hence the localization $M_{\mathfrak{p}} = \Lambda(G)_{\mathfrak{p}} \otimes_{\Lambda(G)} M$ is a finitely generated torsion $\Lambda(G)_{\mathfrak{p}}$ -module, which has

finite length because $\Lambda(G)_p$ is a principal ideal domain. Denoting the length of M_p by $\ell_p(Q(M))$, we then define

$$\chi(Q(M)) = \prod_{p \in \mathcal{P}} \mathfrak{p}^{\ell_p(Q(M))},$$

and we call $\chi(Q(M))$ the **characteristic ideal** of $Q(M)$. This is well defined, as it is proven in [Ch2], Lemma 4.2.1 that, for any finitely generated torsion $\Lambda(G)$ -module M , we have that $M_p = 0$ for all \mathfrak{p} in \mathcal{P} if and only if $Q(M)$ is completely faithful. Moreover, it is shown in [CSS] that $\text{ann}(Q(M))$ is a c -ideal such that $\chi(Q(M)) \subset \text{ann}(Q(M))$, and such that $\chi(Q(M))$ and $\text{ann}(Q(M))$ have the same prime factors in \mathcal{P} . In addition, the exactness of localization and the additivity of the length function shows that χ induces a homomorphism from $\mathcal{D}_b(G)$ to $\mathcal{C}(G)$.

THEOREM 1.2 ([CSS]). — *The homomorphism*

$$\chi : \mathcal{D}_b(G) \rightarrow \mathcal{C}(G)$$

is an isomorphism. In particular, χ induces a bijection between the set of isomorphism classes of simple objects in $\mathbf{M}_b(G)$ and the set \mathcal{P} of all non-zero prime c -ideals of $\Lambda(G)$.

As far as all finitely generated $\Lambda(G)$ -modules are concerned, only the following partial result is proven in [CSS]. If M is an arbitrary finitely generated $\Lambda(G)$ -module, we write M_t for its $\Lambda(G)$ -torsion submodule. It is shown in [Ve2] that the natural map from $M/M_t \rightarrow (M/M_t)^{**}$ is injective and has pseudo-null cokernel.

THEOREM 1.3 ([CSS]). — *Let M be a finitely generated $\Lambda(G)$ -module such that $Q(M_t)$ is bounded. Then we have an isomorphism*

$$Q(M) \xrightarrow{\mathbf{S}} Q(M_t) \oplus Q(M/M_t),$$

*where $Q(M/M_t)$ is reflexive in the sense that it is isomorphic to $Q((M/M_t)^{**})$ in the quotient category $\mathbf{M}(G)$.*

The fundamental question left open by [CSS] is whether every prime ideal \mathfrak{p} in \mathcal{P} is principal. This is true when $G = \mathbb{Z}_p^d$, for any integer $d \geq 1$, thanks to the Weierstrass Preparation Theorem, and we strongly suspect that it remains true for all compact, p -valued p -adic Lie groups G .

2. SKETCH OF THE PROOF OF THEOREM 1.1

One of the nicest parts of Bourbaki's treatise on commutative algebra is his elegant proof of the analogue of Theorem 1.1 for all finitely generated torsion modules over any Noetherian, integrally closed, integral domain (see [B-CA], Chap. VII, §4.4, Theorem 5). We now briefly explain how simple ideas from the algebraic theory of micro-localization allow one to extend these arguments to modules over a wide class of non-commutative

filtered rings. Following the spirit of Bourbaki, we proceed axiomatically, and leave to the end of this section the verification that our concrete ring $\Lambda(G)$ satisfies the axioms we impose.

Let A be an associative ring with unit elements, which will not, in general, be commutative. We assume that A is endowed with a filtration $F.A = \{F_n A : n \in \mathbb{Z}\}$, but we shall be unfaithful to Bourbaki and always assume our filtrations are increasing, i.e. $F_n A \subset F_{n+1} A$ for all n in \mathbb{Z} . This filtration will always be assumed to be exhaustive (i.e. $\bigcup_{n \in \mathbb{Z}} F_n A = A$) and separated (i.e. $\bigcap_{n \in \mathbb{Z}} F_n A = 0$). We write

$$\text{gr}.A = \bigoplus_{n \in \mathbb{Z}} F_n A / F_{n-1} A, \quad \widehat{A} = \varprojlim A / A_n$$

for the associated graded ring, and the completion of A with respect to the filtration, respectively. We follow the non-commutative literature ([LO], Chap. II, Theorem 2.2), and define A to be a **Zariski ring** if $\text{gr}.A$ is left and right Noetherian, and \widehat{A} is a faithfully flat left and right A -module.

For the rest of this section, we shall assume that the ring A satisfies the following axioms:

- (C1) A is complete with respect to $F.A$, i.e. the natural injection from A to \widehat{A} is an isomorphism;
- (C2) $\text{gr}.A$ is isomorphic as a graded ring to $k[T_1, \dots, T_r]$, the ring of polynomials in a finite number of variables with coefficients in a field k , graded by assigning to each of the variables a strictly negative integer as its degree.

Axioms (C1) and (C2) imply that A is left and right Noetherian and has no zero divisors, and that A is a Zariski ring.

For the remainder of the proof, M will denote an arbitrary finitely generated torsion A -module. We endow M with a **good filtration** $F.M = \{F_n M : n \in \mathbb{Z}\}$. This means that we have

$$F_n M = \sum_{i=1}^r F_{n-k_i} A \cdot w_i \quad (n \in \mathbb{Z}),$$

where w_1, \dots, w_r is some fixed set of A -generators of M , and k_1, \dots, k_r are fixed integers. Since $F.M$ is a good filtration, basic properties of Zariski rings show that not only is $F.M$ separated, but also any submodule N of M is closed in the filtration topology [LO]). Also, defining the $\text{gr}.A$ module $\text{gr}.M$ as usual by

$$\text{gr}.M = \bigoplus_{n \in \mathbb{Z}} F_n M / F_{n-1} M,$$

we have that $\text{gr}.M$ is a finitely generated $\text{gr}.A$ -module, which is plainly $\text{gr}.A$ -torsion.

The starting point of our proof is to apply the classical commutative theory to the finitely generated torsion $\text{gr}.A$ -module $\text{gr}.M$. We write $\text{Ass}(\text{gr}.M)$ for the set of prime ideals \mathfrak{p} in $\text{gr}.A$ which are the exact annihilator of some non-zero element of $\text{gr}.M$. Note

that the zero ideal is not in $\text{Ass}(\text{gr}.M)$, because $\text{gr}.M$ is $\text{gr}.A$ -torsion. As $\text{gr}.M$ is a graded $\text{gr}.A$ -module, every ideal in $\text{Ass}(\text{gr}.M)$ is graded. We define

$$W(M) = \{\mathfrak{p}_1, \dots, \mathfrak{p}_m\}$$

to be the set of prime ideals of height 1 in $\text{Ass}(\text{gr}.M)$. It is not difficult to see that $W(M)$ is independent of the very non-canonical choice of the good filtration $F.M$. We define $S = S(M)$ to be the set of all non-zero homogeneous elements of $\text{gr}.A$ which do not belong to $\mathfrak{p}_1 \cup \dots \cup \mathfrak{p}_m$. We can localize $\text{gr}.A$ with respect to the multiplicative set S of non-zero homogeneous elements, and obtain in this way the graded ring $S^{-1}\text{gr}.A$ ([B-CA], Chap. II, §2.9), which we denote by $(\text{gr}.A)_S$. The following lemma (see [CSS], Proposition 3.4) is easily established.

LEMMA 2.1. — *The non-zero graded prime ideals of $(\text{gr}.A)_S$ are precisely the $S^{-1}\mathfrak{p}_i$ ($1 \leq i \leq m$), and these all have height 1. Every proper graded ideal of $(\text{gr}.A)_S$ is contained in one of the $S^{-1}\mathfrak{p}_i$ ($1 \leq i \leq m$).*

COROLLARY 2.2. — *Every graded ideal in $(\text{gr}.A)_S$ is principal.*

To deduce the corollary, we first note that, as A is factorial by axiom (C2), so is its localization $(\text{gr}.A)_S$. Now, if \mathfrak{b} is any non-zero graded ideal of $(\text{gr}.A)_S$, every element of $\text{Ass}((\text{gr}.A)_S/\mathfrak{b})$ must be graded prime ideal of $(\text{gr}.A)_S$, and therefore of height 1 by Lemma 2.1. But then (see [B-CA], Chap. VII, §1.6, Proposition 10) the ideal \mathfrak{b} is divisorial, and so principal because $(\text{gr}.A)_S$ is factorial.

The heart of our proof is the following observation on Ore sets, which we understand goes back to Kashiwara [Ka]. As usual, if $x \in F_n A \setminus F_{n-1} A$, we define its principal symbol $\sharp(x)$ by $\sharp(x) = x + F_{n-1} A$. Let us define

$$T = \{t \in A : \sharp(t) \in S\};$$

here S is the multiplicative set of non-zero elements of A defined above.

PROPOSITION 2.3. — *T is a left and right Ore set in A .*

We omit the proof (see [Li] or the last part of [WK]), which uses the fact that A is a Zariski ring. In view of Proposition 2.3, it makes sense to take either the left or the right localization of A with respect to T . As they are isomorphic, we write A_T for either localization. Moreover, as is explained in [Li], A_T is endowed with a natural separated and exhaustive filtration $F.A_T$, with the property that $\text{gr}.(A_T) = (\text{gr}.A)_S$. Even though we have imposed the axiom that A is complete, it will not in general be true that A_T is complete with respect to the filtration $F.A_T$. Nevertheless, the filtration $F.A_T$ makes A_T into a Zariski ring (see [Li]), and this weaker result suffices for our purposes. Finally, if N is any finitely generated A -module endowed with a good filtration $F.N$, then its localization $N_T = A_T \otimes_A N$ is a finitely generated A_T -module, which (see [Li]) is also endowed with a natural filtration $F.N_T$ such that $\text{gr}.(N_T) = (\text{gr}.N)_S$.

PROPOSITION 2.4. — *Every left and right ideal in A_T is principal.*

To prove this, we can, by symmetry, restrict our attention to left ideals L of A_T . Taking any such left ideal, we endow it with the induced filtration $F.L$ given by $F_n L = L \cap F_n A_T$. By a basic property of Zariski rings, $F.L$ is again a good filtration. Plainly $\text{gr}.L$ is a graded ideal in $\text{gr}.A_T = (\text{gr}.A)_S$, and so is principal by Corollary 2.2. Thus we can find a homogeneous z in $\text{gr}.L$ such that $\text{gr}.L = (\text{gr}.A)_S \cdot z$. Now pick w to be any element of L such that $\sharp(w) = z$. But, thanks to a remarkable property of Zariski rings ([LO], Chap. I, §5, Corollary 5.5), we conclude that $L = A_T w$ (note that we are able to carry out this last step without passing to the completion of A_T).

Once we have established that A_T is a principal ideal domain, we can rapidly complete the proof of Theorem 1.1, following closely the classical commutative argument. As the localized module $M_T = A_T \otimes_A M$ is a finitely generated torsion A_T -module, an old result of Jacobson (see [Ja], Chap. 3, Theorem 10) shows that there exist elements w_1, \dots, w_m in M_T such that

$$M_T = A_T w_1 \bigoplus \cdots \bigoplus A_T w_m.$$

Let $\psi : M \rightarrow M_T$ be the canonical A -homomorphism given by $\psi(m) = 1 \otimes m$, and let $M' = \text{Im}(\psi)$, $N = \text{Ker}(\psi)$. Since N is precisely the set of T -torsion elements of M , we have $N_T = 0$, and so N is pseudo-null by Proposition 2.5 below. Now M' is an A -submodule of M_T with $M'_T = M_T$. We are clearly free to multiply any of the elements w_1, \dots, w_m above by any element of T , and thus we can assume that w_1, \dots, w_m all belong to M' . We then define the A -submodule M'' of M' by

$$M'' = A w_1 \bigoplus \cdots \bigoplus A w_m,$$

where the sum is clearly direct because w_1, \dots, w_m are even linearly independent over A_T . But $N' = M'/M''$ is a quotient of M with $N'_T = 0$, and so N' is also pseudo-null by Proposition 2.5. Now the map $a \mapsto a w_i$ induces an isomorphism of A -modules from A/L'_i to $A w_i$, where L'_i is the annihilator of w_i in A . The composed map

$$\rho : \bigoplus_{i=1}^m A/L'_i \xrightarrow{\sim} \bigoplus_{i=1}^m A w_i = M'' \subset M' = M/N$$

is an injective A -homomorphism with pseudo-null cokernel. Let L_i be the unique left ideal such that L_i/L'_i is the maximal pseudo-null submodule of A/L'_i . We deduce easily that ρ induces an injection of A -modules

$$\varphi : \bigoplus_{i=1}^m A/L_i \longrightarrow M/M_0,$$

where M_0 is the maximal pseudo-null submodule of M , and $\text{Coker}(\varphi)$ is pseudo-null. Thus we have established Theorem 1.1 with $\Lambda(G)$ replaced by our ring A satisfying axioms (C1) and (C2).

PROPOSITION 2.5. — *Let U be any A -subquotient of the A -torsion module M . Then $U_T = 0$ implies that U is pseudo-null. The converse statement holds if A is assumed to be Auslander regular.*

In fact, the notion of a module over an arbitrary ring A with identity element being pseudo-null is defined in [CSS]. We refer the reader to [CSS] for the somewhat delicate proof of Proposition 2.5, as well as a discussion of the notion of Auslander regularity.

The fact that $\Lambda(G)$ satisfies axioms (C1) and (C2) when G is a compact p -valued p -adic Lie group is a consequence of the following result, which is essentially contained in the important but difficult paper of M. Lazard [La]. The explanation given in [CSS] of how to derive this result from Lazard's work was given to us by B. Totaro. The last assertion of Proposition 2.6 is due to O. Venjakob [Ve1], [Ve2].

PROPOSITION 2.6. — *Assume that G is a compact p -adic Lie group, which is p -valued. Then $\Lambda(G)$ possesses a complete, separated and exhaustive filtration $F.\Lambda(G)$ such that $gr.\Lambda(G)$ is isomorphic as a graded ring to the polynomial ring $\mathbb{F}_p[X_0, \dots, X_d]$ in $d + 1$ variables, where d is the dimension of G ; here the grading on $\mathbb{F}_p[X_0, \dots, X_d]$ is given by assigning to each of the variables X_i a strictly negative integer degree. In particular, $\Lambda(G)$ satisfies axioms (C1) and (C2). In addition, $\Lambda(G)$ is Auslander regular.*

3. ARITHMETIC EXAMPLES

Concrete examples of finitely generated torsion $\Lambda(G)$ -modules, which are of great arithmetic interest, abound in arithmetic geometry. Because of lack of space, we only discuss two classes of examples. In both cases, G is non-commutative, and is the image of Galois in a 2-dimensional p -adic Galois representation; thus, by Lazard [La], G is automatically p -valued provided G is pro- p and $p \geq 5$.

Example 1. Let $p \geq 5$, and let μ_{p^n} ($1 \leq n \leq \infty$) denote the group of all p^n -th roots of unity. We write F for any finite extension of \mathbb{Q} containing μ_p , and define

$$F^{cyc} = F(\mu_{p^\infty}), \quad \Gamma = G(F^{cyc}/F).$$

Now fix a non-zero element α of F , which is not a root of unity, and define

$$K_\infty = F^{cyc}(\alpha^{1/p^n} : n = 1, 2, \dots), \quad G = G(K_\infty/F).$$

If we define H to be $G(K_\infty/F_{cyc})$, then both H and Γ are isomorphic to \mathbb{Z}_p , so that G is a p -adic Lie group of dimension 2, which is p -valued. Moreover, G is not commutative. Let $\psi : \Gamma \rightarrow \mathbb{Z}_p^\times$ be the character giving the action of Γ on μ_{p^∞} . Then, as α belongs to F , Kummer theory shows that the natural action of Γ on H via inner automorphism is given by the character ψ . One can study many left $\Lambda(G)$ -modules which are of arithmetic interest, but the simplest is probably the following. Let L_∞ denote the maximal

unramified abelian p -extension of K_∞ , and put $X = G(L_\infty/K_\infty)$. As usual, there is a continuous left action of G on X via inner automorphism (if σ is in G and x in X , we define $\sigma \cdot x = \tilde{\sigma}x\tilde{\sigma}^{-1}$, where $\tilde{\sigma}$ denotes any lifting of σ to the Galois group of L_∞ over F). Y. Ochi [Oc] has proven that X is a finitely generated torsion $\Lambda(G)$ -module. At present, very little else is known about the module $\Lambda(G)$; in particular, it seems that at present no example is known in which we can prove that X is not pseudo-null as a $\Lambda(G)$ -module. In the special case $F = \mathbb{Q}(\mu_p)$ and $\alpha = p$, one can easily show that $X \neq 0$ if and only if p is an irregular prime. Moreover, in this case, O. Venjakob [Ve3] has shown that if X is pseudo-null, then the p -primary subgroup of the ideal class group of K_∞ is zero.

Example 2. Let F be a finite extension of \mathbb{Q} , and E an elliptic curve defined over F , with $\text{End}_{\mathbf{Q}}(E) = \mathbb{Z}$. Let $p \geq 5$, and let E_{p^n} ($1 \leq n \leq \infty$) denote the group of p^n -division points on E . We define

$$F_\infty = F(E_{p^\infty}), \quad G = G(F_\infty/F).$$

The action of G on E_{p^∞} defines an injection of G into $\text{Aut}(E_{p^\infty}) \xrightarrow{\sim} GL_2(\mathbb{Z}_p)$, and, by a theorem of Serre [Se3], the image of G is open in $GL_2(\mathbb{Z}_p)$. By the Weil pairing, $F^{cyc} = F(\mu_{p^\infty})$ is contained in F_∞ , and we put

$$H = G(F_\infty/F^{cyc}), \quad \Gamma = G(F^{cyc}/F).$$

We shall assume from now on that G is pro- p (this can always be achieved, if necessary, by replacing F by a finite extension, e.g. by $F(E_p)$). Hence Γ is pro- p , and so is isomorphic to \mathbb{Z}_p .

For each intermediate field L with $F \subseteq L \subseteq F_\infty$, we recall that the **Selmer group** of E over L is defined by

$$S(E/L) = \text{Ker}(H^1(G(\overline{\mathbb{Q}}/L), E_{p^\infty}) \longrightarrow \prod_v H^1(G(\overline{L}_v/L_v), E(\overline{L}_v))),$$

where v runs over all finite places of L , and L_v denotes the union of the completions at v of all finite extensions of F contained in L . As usual, we have the exact sequence

$$0 \rightarrow E(L) \otimes \mathbb{Q}_p/\mathbb{Z}_p \rightarrow S(E/L) \rightarrow \text{III}(E/L)(p) \rightarrow 0,$$

where $\text{III}(E/L)(p)$ denotes the p -primary subgroup of the Tate-Shafarevich group of E over L . We write

$$X(E/L) = \text{Hom}(S(E/L), \mathbb{Q}_p/\mathbb{Z}_p)$$

for the compact Pontrjagin dual of the discrete p -primary module $S(E/L)$. If L is Galois over F , then the Galois group $G(L/F)$ of L over F has a natural action on both $S(E/L)$ and $X(E/L)$, and it is easily seen that $X(E/L)$ is always a finitely generated $\Lambda(G(E/L))$ -module. We shall be primarily interested in the $\Lambda(G)$ -module $X(E/F_\infty)$. If E has good ordinary reduction at all places v of F dividing p , old conjectures due to B. Mazur [Ma] and M. Harris [Ha] affirm, respectively, that $X(E/F^{cyc})$ is $\Lambda(\Gamma)$ -torsion and $X(E/F_\infty)$

is $\Lambda(G)$ -torsion. It is easy to see that the validity of Mazur's conjecture for all finite extensions L of F contained in F_∞ implies the validity of Harris' conjecture for F_∞ , but this is of little use in practice since the number of cases in which we can prove Mazur's conjecture remains very limited (the best result to date is K. Kato's [Kat] theorem that $X(E/F^{cyc})$ is $\Lambda(\Gamma)$ -torsion when E is an elliptic curve defined over \mathbb{Q} with good ordinary reduction at p , and F is an abelian extension of \mathbb{Q}). In [CH], an alternative approach is given which does enable one to give the first proven examples of Harris' conjecture, and to deduce new examples of Mazur's conjecture.

THEOREM 3.1. — *Assume that (i) $p \geq 5$, (ii) G is pro- p , (iii) E has good ordinary reduction at all places v of F dividing p , and (iv) $X(E/F^{cyc})$ is a finitely generated \mathbb{Z}_p -module. Then $X(E/F_\infty)$ is a finitely generated $\Lambda(H)$ -module, where $H = G(F_\infty/F^{cyc})$. In particular, $X(E/F_\infty)$ is a torsion $\Lambda(G)$ -module.*

Remark 1. Every $\Lambda(G)$ -module, which is finitely generated over $\Lambda(H)$, is automatically $\Lambda(G)$ -torsion. This is because $\Lambda(G)$ is not finitely generated over $\Lambda(H)$, since G/H is infinite.

Remark 2. In the general framework and notation of Theorem 1.1, we say that the $\Lambda(G)$ -module M has μ -invariant zero if none of the left ideals L_1, \dots, L_m appearing in Theorem 1.1 is of the form $\Lambda(G)p^k$ for some integer $k \geq 1$. If Γ is a group isomorphic to \mathbb{Z}_p , it follows from Theorem 1.1 and the Weierstrass preparation theorem that a $\Lambda(\Gamma)$ -module Y is a finitely generated \mathbb{Z}_p -module if and only if Y is $\Lambda(\Gamma)$ -torsion and has μ -invariant zero. Note also that if M is a finitely generated $\Lambda(H)$ -module, then it must have μ -invariant zero, because $\Lambda(G)/\Lambda(G)p^k$ is not a finitely generated $\Lambda(H)$ -module when $k \geq 1$.

A second important result about $X(E/F_\infty)$ is due to Y. Ochi and O. Venjakob [OV].

THEOREM 3.2. — *Assume that hypotheses (i), (ii), (ii), and (iv) of Theorem 3.1 are valid. Then $X(E/F_\infty)$ contains no non-zero pseudo-null submodule, the $\Lambda(H)$ -torsion submodule of $X(E/F_\infty)$ is zero, and $X(E/F_\infty)$ has strictly positive $\Lambda(H)$ -rank.*

To prove the last assertion of Theorem 3.2, we must use the fact that always $X(E/F_\infty) \neq 0$ (this was first remarked by R. Greenberg, and a proof is given in the Appendix of [CH]). Assume now that E over F satisfies hypotheses (i), (ii), (ii), and (iv) of Theorem 3.1. We conclude from the above results and Theorem 1.1 that there exist non-zero left ideals L_1, \dots, L_m of $\Lambda(G)$ such that we have an exact sequence of $\Lambda(G)$ -modules

$$0 \rightarrow \bigoplus_{i=1}^m \Lambda(G)/L_i \rightarrow X(E/F_\infty) \rightarrow D \rightarrow 0,$$

where D is pseudo-null. We stress that there is great arithmetic interest in studying the left ideals L_1, \dots, L_m , even in particular numerical examples. Of course, one imagines

that these left ideals must be related to the values at $s = 1$ of the twists of the complex L -function of E over F by Artin characters of G (i.e. those which factor through finite quotients of G). On a much more elementary level, one can ask whether or not $\text{ann}_{\Lambda(G)}(X(E/F_\infty)) \neq 0$. This still has not been settled in a single numerical example when $F_\infty = F(E_{p^\infty})$. Surprisingly, when K_∞ and G are as defined in Example 1 above, Hachimori and Venjakob [HV] have recently given many examples of elliptic curves E over K_∞ such that the dual of the Selmer group of E over K_∞ is a finitely generated torsion $\Lambda(G)$ -module, which is not pseudo-null, but which is completely faithful. For example, they prove that this is the case for the elliptic curve $E = X_1(11)$ given below, when $p = 5$, and K_∞ is the field obtained by adjoining to \mathbb{Q} all 5-power roots of unity and all 5-power roots of 11.

We end by discussing two specific numerical examples of elliptic curves over their field of p -power division points.

Numerical example 1. I am grateful to T. Fisher for first pointing out this example to me. Let E be the elliptic curve over \mathbb{Q}

$$E : y^2 + xy = x^3 - x - 1.$$

This is the curve $B1$ of conductor 294 in [Cr1]. Take $F = \mathbb{Q}(\mu_7)$ and $p = 7$. Although E has bad reduction at 7 over \mathbb{Q} , it is easily seen that E has good ordinary reduction at the unique prime of F above 7. Moreover, μ_7 is a Galois submodule of E_7 , whence we see easily that $F_\infty = \mathbb{Q}(E_{7^\infty})$ is a pro-7 extension of F . Fisher [Fil] has shown that $S(E/F) = 0$. One can then use arguments from Iwasawa theory ([CS], p. 83) to deduce that we also have $S(E/F^{cyc}) = 0$. Hence hypotheses (i), (ii), (iii), and (iv) of Theorem 3.1 are valid for E over F and $p = 7$. We conclude that $X(E/F_\infty)$ is a torsion $\Lambda(G)$ -module, with μ -invariant equal to zero, and with no non-zero pseudo-null submodule. Moreover, $X(E/F_\infty)$ is finitely generated over $\Lambda(H)$, its $\Lambda(H)$ -torsion submodule is zero, and it has positive $\Lambda(H)$ -rank. Here $G = G(\mathbb{Q}(E_{7^\infty})/\mathbb{Q}(\mu_7))$ and $H = G(\mathbb{Q}(E_{7^\infty})/\mathbb{Q}(\mu_{7^\infty}))$.

Numerical example 2. Let E be the elliptic curve $X_1(11)$ over \mathbb{Q} , namely

$$E : y^2 + y = x^3 - x^2.$$

Then E has good ordinary reduction at 5. Take $F = \mathbb{Q}(\mu_5)$ and $p = 5$. As $(0, 0)$ is a rational point of order 5 on E , $F_\infty = \mathbb{Q}(E_{5^\infty})$ is a pro-5 extension of F . Indeed, it is well known and easy to see that the image of G in $\text{Aut}(E_{5^\infty})$ can be identified with the subgroup of all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $GL_2(\mathbb{Z}_5)$ with $a \equiv d \equiv 1 \pmod{5}$, and $c \equiv 0 \pmod{5^2}$, and this group in turn is isomorphic to the group of all matrices in $GL_2(\mathbb{Z}_5)$, which are congruent to the identity modulo 5. Finally, it is well known that, in this case, $S(E/F^{cyc}) = 0$ (see [CS], Chap. 5). Hence hypotheses (i), (ii), (iii), and (iv) of Theorem 3.1 hold for E over F ,

and we conclude that $X(E/F_\infty)$ is a torsion $\Lambda(G)$ -module, with μ -invariant equal to zero, and with no non-zero pseudo-null submodule. In fact, the following stronger result is true.

PROPOSITION 3.3. — *Take $E = X_1(11)$, $p = 5$, $F = \mathbb{Q}(\mu_5)$, and $H = G(F_\infty/\mathbb{Q}(\mu_{5^\infty}))$. Then $X(E/F_\infty)$ is a finitely generated $\Lambda(H)$ -module of rank 4, its $\Lambda(H)$ -torsion submodule is zero, but it is not a free $\Lambda(H)$ -module.*

For each finite Galois extension L of F which is contained in F_∞ , we write

$$G_L = G(F_\infty/L), \quad H_L = G(F_\infty/L^{cyc}),$$

so that G_L ranges over the open normal subgroups of G , and H_L ranges over the open normal subgroups of H . The proof of Proposition 3.3 hinges on the remarkable fact that one can use ideas of Y. Hachimori and K. Matsuno [HM] to determine the exact \mathbb{Z}_5 -rank of the H_L -coinvariants of $X(E/F_\infty)$ for every finite Galois extension L of F contained in F_∞ . In particular, the following result is proven in §7 of [CH].

PROPOSITION 3.4. — *For each finite Galois extension L of F contained in F_∞ , $X(E/L^{cyc})$ is a free \mathbb{Z}_5 -module of rank 4. $[L^{cyc} : F^{cyc}] - \tau_L$, where τ_L denotes the number of prime of L^{cyc} above 11. In particular, Mazur's conjecture is true for E over L^{cyc} , and $E(L^{cyc})$ is a finitely generated abelian group of rank at most $4 \cdot [L^{cyc} : F^{cyc}] - \tau_L$.*

One deduces easily from this proposition that $X(E/F_\infty)_{H_L}$ has \mathbb{Z}_5 -rank equal to $4 \cdot [L^{cyc} : F^{cyc}]$ for all L . At the same time, one shows that the \mathbb{Z}_5 -torsion subgroup of $X(E/F_\infty)_{H_L}$ is never zero, which shows that $X(E/F_\infty)$ is not a free $\Lambda(H)$ -module.

By contrast, it is a deep arithmetic problem to determine the exact \mathbb{Z}_5 -rank of the G_L -coinvariants of $X(E/F_\infty)$ for all L . In particular, the following lemma is not difficult to prove using the methods of [CH].

LEMMA 3.5. — *Let L be any finite Galois extension of F contained in F_∞ . Then $X(E/F_\infty)_{G_L}$ is finite if and only if both $E(L)$ and the 5-primary subgroup of the Tate-Shafarevich group of E over L are finite.*

We have already used the classical fact that $X(E/F_\infty)_G$ is finite. By some remarkable explicit descent calculations, T. Fisher [Fi2] has recently shown that $X(E/F_\infty)_{G_L}$ is finite for L ranging over the following three cyclic extensions of F of degree 5 which are contained in F_∞ :

$$L_1 = \mathbb{Q}(X_1(11)_5), \quad L_2 = \mathbb{Q}(X_2(11)_5), \quad L_3 = F\mathbb{Q}(\mu_{11})^+;$$

here $X_2(11)$ denotes the unique elliptic curve of conductor 11 over \mathbb{Q} which has no non-zero rational point, and $\mathbb{Q}(\mu_{11})^+$ denotes the maximal real subfield of $\mathbb{Q}(\mu_{11})$. The field L_1 is the splitting field of the polynomial

$$x^5 + 2x^4 + 6x^3 - 2x^2 + 4x - 1.$$

Then T. Fisher's descent calculations [Fi2] show that $E(L_1)$ is finite, and $\lll(E/L_1)(5) = (\mathbb{Z}/5\mathbb{Z})^2$. However, it does not seem possible to extend his explicit calculations even to the field $K_2 = L_1(\mu_{5^2})$, which has degree 100 over \mathbb{Q} . Nevertheless, if

$$K_n = L_1(\mu_{5^{n+1}}) \quad (n = 0, 1, \dots),$$

T. Fisher, R. Greenberg and myself have shown that simple theoretical arguments from the Iwasawa theory of elliptic curves do enable one to prove that, for all integers $n = 0, 1, \dots$, $E(K_n)$ is finite and $\lll(E/K_n)(5)$ is finite of order 5^{16n+2} .

When this exposé was given in November 2001, I naively imagined that perhaps $X_1(11)$ had no points of infinite order in the whole tower F_∞ of 5-division points. I am very grateful to K. Matsuno (unpublished) for producing overwhelming numerical evidence that this is not true. Take the field L_3 above, and let $H_2 = L_3(\mu_{5^2})$, so that H_2 is an abelian extension of \mathbb{Q} of degree 100. K. Matsuno calculated the complex L -function of $X_1(11)$ over H_2 , and proved that it has a zero at $s = 1$ of order 4. Thus, unless the conjecture of Birch and Swinnerton-Dyer is false (which I do not for one moment believe), $X_1(11)$ should have rank 4 over H_2 . So far, no point of infinite order has been found, nor has Fisher been able to extend his explicit descent calculations to $X_1(11)$ over H_2 . Thus, in conclusion, two important questions remain unanswered about the arithmetic of $X_1(11)$ over its field F_∞ of 5-power division points. Is the dual of the Selmer group of $X_1(11)$ over F_∞ a completely faithful $\Lambda(G)$ -module? What is the \mathbb{Z} -rank of the group $X_1(11)(F_\infty)$ of F_∞ -rational points modulo its torsion subgroup?

REFERENCES

- [As] Asano, K.: *Zur Arithmetik in Schieftringen*, Osaka J. Math. **2**, 98–134 (1949).
- [Bj] Björk, J.-E.: *Filtered Noetherian Rings*. In Noetherian rings and their applications, Math. Survey Monographs **24**, pp. 59–97, AMS 1987.
- [B-CA] Bourbaki, N.: *Algèbre Commutative*, Paris, Hermann, 1972.
- [Ch1] Chamarie, M.: *Anneaux de Krull non commutatifs*, J. Algebra **72**, 210–222 (1981).
- [Ch2] Chamarie, M.: *Modules sur les anneaux de Krull non-commutatifs*. In Sémin. d'Algèbre P. Dubreil et M.-P. Malliavin 1982, Springer Lecture Notes **1029**, pp. 283–310, Springer, 1983.
- [Cr1] Cremona, J.: *Algorithms for modular elliptic curves*, Second ed., Cambridge University Press, 1997.
- [CH] Coates, J., Howson, S.: *Euler characteristics and elliptic curves II*, J. Math. Soc. Japan **53**, 175–235 (2001).
- [CS] Coates, J., Sujatha, R.: *Galois cohomology of elliptic curves*, TIFR-AMS Lecture Notes, **88**, 2000.
- [CSS] Coates, J., Schneider, P., Sujatha, R., *Modules over Iwasawa algebras*, to appear.

- [Fi1] Fisher, T.: *On 5 and 7 descents for elliptic curves*, Ph.D. thesis, Cambridge University, 2000.
- [Fi2] Fisher, T.: *Descent calculations for the elliptic curves of conductor 11*, to appear.
- [Ha] Harris, M.: *p-adic representations arising from descent on abelian varieties*, *Compositio Math.* **30**, 177–245 (1979).
- [Ho] Howson, S.: *Structure of central torsion Iwasawa modules*, to appear.
- [HM] Hachimori, Y., Matsuno, K.: *An analogue of Kida's formula for the Selmer groups of elliptic curves*, *J. Alg. Geometry* **8**, 581–601 (1999).
- [HV] Hachimori, Y., Venjakob, O., *Completely faithful Selmer groups over Kummer extensions*, in preparation.
- [Iw] Iwasawa, K.: *On Γ -extensions of number fields*, *Bull. AMS* **65**, 183–226 (1959).
- [Ja] Jacobson, H.: *Theory of rings*, *Math. Surveys* 2, AMS, Providence, 1943.
- [Ka] Kashiwara, M.: *Algebraic study of systems of partial differential equations*, Thesis, Tokyo University, 1971.
- [Kat] Kato, K.: *p-adic Hodge theory and values of zeta functions of modular forms*, to appear.
- [La] Lazard, M.: *Groupes analytiques p-adiques*, *Publ. Math. IHES* **26**, 380–603 (1965).
- [Li] Li Huishi: *Lifting Ore sets of Noetherian filtered rings and applications*, *J. Algebra* **179**, 686–703 (1996).
- [LO] Li Huishi, van Oystaeyen, F.: *Zariskian Filtrations*, Dordrecht: Kluwer 1996.
- [Ma] Mazur, B.: *Rational points of abelian varieties in towers of number fields*, *Invent. Math.* **18**, 183–266 (1992).
- [Oc] Ochi, O.: *Iwasawa modules via homotopy theory*, Ph.D. thesis, Cambridge University, 1999.
- [OV] Ochi, Y., Venjakob, O.: *On the structure of Selmer groups over p-adic Lie extensions*, to appear in *J. Alg. Geom.*
- [Ro] Robson, J.C.: *Cyclic and faithful objects in quotient categories with applications to noetherian simple or Asano rings*. In *Noncommutative Ring Theory*, Springer Lecture Notes **545**, pp. 151–172, Springer 1976.
- [Se1] Serre, J-P.: *Classes des corps cyclotomiques (d'après K. Iwasawa)*, *Sém. Bourbaki*, exp. n° 174, décembre 1958, *Collection hors série Astérisque* **5** (1995), 83–93.
- [Se2] Serre, J-P.: Letter to K. Iwasawa, dated August 27th 1958.
- [Se3] Serre, J-P.: *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, *Invent. Math.* **15** (1972), 259–331.
- [Ve1] Venjakob, O.: *Iwasawa Theory of p-adic Lie Extensions*, Thesis, Heidelberg University, 2000.
- [Ve2] Venjakob, O.: *On the structure of the Iwasawa algebra of a p-adic Lie group*, to appear in the *J. European Math. Soc.*

- [Ve3] Venjakob, O.: *The Weierstrass preparation theorem in non-commutative Iwasawa theory*, in preparation.
- [WK] Wexler-Kreindler, E.: *Microlocalisation, platitude et théorie de torsion*, *Comm. Algebra* **16**, 1813–1852 (1988).

John COATES

Cambridge University

DPMMS

Wilberforce Road

GB-CAMBRIDGE, CB3 0WB

E-mail : `J.H.Coates@dpmms.cam.ac.uk`

LES CONJECTURES DE MONODROMIE p -ADIQUES

par Pierre COLMEZ

Introduction

0.1. Survol des résultats

Le théorème local de monodromie ℓ -adique de Grothendieck peut s'énoncer de la manière suivante.

THÉORÈME 0.1. — *Soit K un corps complet pour une valuation discrète de corps résiduel k fini et de caractéristique p . Soient $\ell \neq p$ un nombre premier et V un \mathbf{Q}_ℓ -espace vectoriel de dimension finie sur lequel le groupe de Galois absolu \mathcal{G}_K de K agit continûment; alors l'action de l'inertie est quasi-unipotente : il existe un sous-groupe ouvert du sous-groupe d'inertie I_K de \mathcal{G}_K dont les éléments agissent de manière unipotente.*

Ce théorème, bien que quasi-trivial [44] (c'est une conséquence facile de l'existence d'une < structure de Frobenius > : le pro- ℓ -quotient de I_K est isomorphe à \mathbf{Z}_ℓ et, si $q = \text{card } k$, un Frobenius $\sigma \in \mathcal{G}_K$ opérant par conjugaison intérieure, agit sur ce pro- ℓ -quotient par multiplication par q), a des implications globales intéressantes concernant la cohomologie étale ℓ -adique des variétés algébriques (voir [49] pour une discussion et des variations sur ce thème).

L'énoncé correspondant étant complètement faux pour $\ell = p$ (le pro- p -quotient de I_K étant nettement plus compliqué que le pro- ℓ -quotient), il n'est pas très facile d'imaginer ce que peut en être un analogue p -adique et, de fait, il a fallu un certain temps, ne serait-ce que pour définir les objets à considérer. Il y a deux cas suivant que le corps K est de caractéristique 0 ou de caractéristique p . Dans le cas où K est de caractéristique 0, la difficulté pour arriver à formuler cette conjecture vient de ce que \mathcal{G}_K possède beaucoup trop de \mathbf{Q}_p -représentations et qu'il faut commencer par comprendre lesquelles on veut garder (représentations < de de Rham >). Le théorème de monodromie locale prend alors la forme suivante :

THÉORÈME 0.2. — *Toute \mathbf{Q}_p -représentation de de Rham de \mathcal{G}_K est potentiellement semi-stable.*

Si K est de caractéristique p , le groupe \mathcal{G}_K a < encore plus > de \mathbf{Q}_p -représentations et il faut remplacer ces dernières par des modules différentiels sur une couronne < infiniment fine > de rayon 1. Dans ce cadre, l'analogie du théorème de Grothendieck est l'énoncé suivant :

THÉORÈME 0.3. — *Tout module différentiel sur l'anneau de Robba, muni d'une structure de Frobenius, est quasi-unipotent.*

L'intérêt de l'énoncé du théorème 0.3 < conjecture de monodromie p -adique de Crew¹ > a été mis en évidence par Crew [29]. Cet énoncé permet, en particulier, d'obtenir des résultats de finitude [29] pour la cohomologie rigide. Ces résultats de finitude ont entre-temps été démontrés par Mebkhout [59] et Berthelot [6] (dans le cas des coefficients constants), le premier en utilisant ses résultats obtenus en collaboration avec Christol (cf. [16], [17], [18], [19] et [20]) et le second en utilisant les altérations de de Jong [50]. Pour d'autres applications, voir [53, 54, 72].

La conjecture de Crew vient d'être démontrée de manière quasi-simultanée par André [2], par Mebkhout [60] et par Kedlaya [52], par des méthodes très différentes. D'autre part, Berger [5] avait, peu auparavant, montré comment déduire le théorème 0.2 < conjecture de monodromie p -adique de Fontaine [39]² > de la conjecture de Crew.

0.2. Modules différentiels sur l'anneau de Robba

0.2.1. *La conjecture de Crew.* Soit L un corps complet pour une valuation p -adique de corps résiduel k . Soit \mathcal{R}_L (resp. \mathcal{E}_L^\dagger) l'anneau des fonctions à coefficients dans L qui sont analytiques (resp. analytiques bornées) sur une couronne $0 < v_p(T) \leq r$, où $r > 0$ dépend de la fonction que l'on considère. L'anneau \mathcal{R}_L est souvent appelé *anneau de Robba*. L'anneau \mathcal{E}_L^\dagger est un anneau valué non complet d'anneau résiduel $k((T))$.

Un Frobenius φ sur ces anneaux est un morphisme d'anneaux de \mathcal{E}_L^\dagger respectant la valuation et induisant le morphisme $x \rightarrow x^p$ sur $k((T))$ (ou, plus généralement $x \rightarrow x^q$, où q est une puissance de p); un tel morphisme s'étend par continuité à \mathcal{R}_L . Des exemples agréables de Frobenius sont $T \rightarrow T^p$ et $T \rightarrow (1 + T)^p - 1$. Un φ -module \mathcal{D} sur un de ces anneaux est un module libre de rang fini muni d'un endomorphisme φ -semi-linéaire $\varphi_{\mathcal{D}}$ dont la matrice dans une base est *invertible*.

Choisissons une dérivation continue ∂ de l'anneau \mathcal{E}_K^\dagger (comme $\frac{d}{dT}$, $T \frac{d}{dT}$ ou $(1 + T) \frac{d}{dT}$). Un ∂ -module sur un de ces anneaux est un module libre de rang fini muni d'une connexion

¹Où de Crew-Tsuzuki. Crew [29] est très prudent, et se cantonne aux modules différentiels provenant de la géométrie, et Tsuzuki [72] est nettement plus optimiste, mais ni Crew, ni Tsuzuki n'ont, formellement, présenté cet énoncé comme une conjecture.

²Cet énoncé n'est, lui non plus, pas formellement présenté comme une conjecture dans [39].

$\partial_{\mathcal{D}}$ au-dessus de ∂ , c'est-à-dire vérifiant

$$\partial_{\mathcal{D}}(f \cdot x) = \partial f \cdot x + f \cdot \partial_{\mathcal{D}}x \quad \text{si } f \in \mathcal{E}_K^{\dagger} \text{ (resp. } \mathcal{R}_K) \text{ et } x \in \mathcal{D}.$$

Un (φ, ∂) -module \mathcal{D} est un module qui est à la fois un φ -module et un ∂ -module de telle sorte que $\varphi_{\mathcal{D}}$ et $\partial_{\mathcal{D}}$ vérifient la relation de commutation

$$\partial_{\mathcal{D}} \circ \varphi_{\mathcal{D}} = \frac{\partial(\varphi(T))}{\varphi(\partial T)} \cdot \varphi_{\mathcal{D}} \circ \partial_{\mathcal{D}}.$$

Dans le cas particulier $\partial = (1 + T) \frac{d}{dT}$ et $\varphi(T) = (1 + T)^p - 1$, on a $\frac{\partial(\varphi(T))}{\varphi(\partial T)} = p$. Si on fixe une base de \mathcal{D} et si A (resp. B) est la matrice de $\varphi_{\mathcal{D}}$ (resp. $\partial_{\mathcal{D}}$) dans cette base, cette relation de commutation se traduit par

$$BA + \partial A = \frac{\partial(\varphi(T))}{\varphi(\partial T)} A \varphi(B).$$

Si E est une extension finie séparable de $k((T))$, il lui correspond des extensions $\mathcal{E}_L^{\dagger}(E)$ et $\mathcal{R}_L(E)$ de \mathcal{E}_L^{\dagger} et \mathcal{R}_L respectivement auxquelles les actions de φ et ∂ s'étendent canoniquement. La conjecture de Crew peut alors s'énoncer de façon précise sous la forme :

CONJECTURE 0.4. — *Si \mathcal{D} est un (φ, ∂) -module de rang d sur \mathcal{R}_L , alors il existe une extension finie séparable E de $k((T))$ telle que l'équation différentielle $\partial_{\mathcal{D}}X = 0$ admette d solutions linéairement indépendantes dans $\mathcal{R}_L(E)[\log T] \otimes \mathcal{D}$.*

Cette conjecture a une < reformulation filtrée >, faisant disparaître $\log T$, sous la forme : < si \mathcal{D} est un (φ, ∂) -module sur \mathcal{R}_L , alors il existe une extension finie séparable E de $k((T))$ et une filtration croissante de \mathcal{D} par des sous- ∂ -modules \mathcal{D}_i telles que, pour tout i , l'équation différentielle $\partial_{\mathcal{D}_i/\mathcal{D}_{i-1}}X = 0$ admette $d_i = \text{rang}(\mathcal{D}_i/\mathcal{D}_{i-1})$ solutions linéairement indépendantes dans $\mathcal{R}_L(E) \otimes (\mathcal{D}_i/\mathcal{D}_{i-1})$ >.

REMARQUE 0.5. — Il y a une subtilité cachée dans cette conjecture : la géométrie algébrique fournit naturellement des (φ, ∂) -modules sur \mathcal{E}_L^{\dagger} , mais, en rang ≥ 2 , il faut vraiment étendre les coefficients à $\mathcal{R}_L(E)$ [et pas seulement à $\mathcal{E}_L^{\dagger}(E)$]. Christol et Mebkhout se sont retrouvés confrontés au même problème (rem. 1.6). (On trouvera une explication de ce phénomène au n° 2.3 : les ensembles des \mathcal{E} -pentes et \mathcal{R} -pentes d'un (φ, ∂) -module sur \mathcal{E}_L^{\dagger} ne sont pas toujours égaux.)

0.2.2. *Les démonstrations.* Crew [28] avait établi la conjecture en rang 1, et avant cet été, les seuls résultats un peu généraux concernant cette conjecture étaient, d'une part, un théorème de Tsuzuki [73] montrant qu'un (φ, ∂) -module sur \mathcal{E}^{\dagger} isocline (i.e. n'ayant qu'une φ -pente³) devient trivial après une extension finie séparable de $k((T))$ et, d'autre

³Les φ -modules et les ∂ -modules ont chacun une notion de pente. Ces deux notions n'ont rien à voir ; en particulier elles se comportent de manières très différentes par produit tensoriel.

part, un théorème de Christol et Mebkhout [17] montrant qu'un (φ, ∂) -module de ∂ -pente 0 devient unipotent sur une extension modérée de $k((T))$ (i.e. une extension de la forme $k((T^{1/d}))$ avec $(d, p) = 1$).

Malgré ces résultats encourageants, les experts étaient plutôt sceptiques en ce qui concerne la conjecture de Crew. L'exemple [19, Ex. 3.0-11] du ∂ -module associé, en 2-adique, à l'opérateur différentiel $9T^3 \frac{d^2}{dT^2} + 9T^2 \frac{d}{dT} + \frac{4}{3} - T$ constituant un contre-exemple potentiel sérieux. La situation a radicalement changé avec l'article [1] dans lequel André montre que ce contre-exemple n'en est pas un. Sa démonstration dans le cas général [2] est une généralisation de celle qu'il a utilisée dans [1]; elle est purement existentielle et ne fournit aucun renseignement sur l'extension à faire⁴ pour rendre le module unipotent. Elle s'appuie de manière essentielle sur les résultats de Christol-Mebkhout et mêle adroitement la théorie de Galois différentielle et les contraintes combinatoires résultant de la théorie de Christol-Mebkhout.

La démonstration de Mebkhout [60] utilise aussi à fond les résultats de Christol-Mebkhout mais, contrairement à celle d'André, est constructive : elle fournit un algorithme (pour courageux) explicitant pas à pas les extensions séparables de $k((T))$ qu'il faut faire pour aboutir à la ∂ -pente 0.

Ces deux démonstrations n'utilisent la structure de Frobenius que de manière anecdotique et permettent de démontrer un résultat plus fort que la conjecture de Crew. La démonstration de Kedlaya [52] est totalement orthogonale. Il fait une étude poussée des φ -modules sur l'anneau de Robba (et ses généralisations) obtenant en particulier un analogue du théorème de Dieudonné-Manin. Ceci lui permet de munir un tel module d'une filtration stable par Frobenius telle que chaque morceau du gradué soit isocline. D'autre part, si on est parti d'un (φ, ∂) -module, la filtration est stable par ∂ , ce qui permet d'utiliser le théorème de Tsuzuki pour conclure.

0.2.3. *Compléments.* On peut réinterpréter la conjecture de Crew < à la Fontaine >. Supposons k algébriquement clos (cela évitera d'avoir à faire des extensions du corps des coefficients dans ce qui suit). Si E est une extension finie galoisienne de $k((T))$, l'anneau $\mathcal{R}_L(E)$ est muni d'actions de $\mathcal{G}_{k((T))} = \text{Gal}(k((T))^{\text{sep}}/k((T)))$ et de la dérivation $\mathcal{R}_L(E)$ -linéaire N normalisée par $N(\log T) = 1$; ces deux actions commutent entre elles et commutent aussi à ∂ .

Soit $\overline{\mathcal{R}}_{\log} = \bigcup_{E \subset k((T))^{\text{sep}}} \mathcal{R}_L(E)[\log T]$; c'est un anneau muni des dérivations N et ∂ et d'une action discrète de $\mathcal{G}_{k((T))}$, ces actions commutant deux à deux. Si \mathcal{D} est un (φ, ∂) -module de rang d , l'espace $V(\mathcal{D})$ des solutions de l'équation différentielle $\partial_{\mathcal{D}} X = 0$ dans $\overline{\mathcal{R}}_{\log} \otimes \mathcal{D}$ est un L -espace vectoriel de dimension d (d'après la conjecture de Crew) muni d'actions de N et $\mathcal{G}_{k((T))}$ commutant entre elles, l'action de $\mathcal{G}_{k((T))}$ se faisant à travers un quotient fini. Un tel objet sera appelé *un $(\mathcal{G}_{k((T))}, N)$ -module sur L* .

⁴Dans [1], André, après avoir prouvé l'existence d'une extension trivialisant le module, détermine explicitement cette extension par une méthode qu'il semble difficile de généraliser.

Réciproquement, si V est un $(\mathcal{G}_{k((T))}, N)$ -module sur L , le \mathcal{R}_L -module $\mathcal{D}(V)$ des points fixes de $\overline{\mathcal{R}}_{\log} \otimes_L V$ sous les actions de N et $\mathcal{G}_{k((T))}$ est un (φ, ∂) -module de rang d sur \mathcal{R}_L .

Il n'est pas difficile de montrer que les foncteurs $V \rightarrow \mathcal{D}(V)$ et $\mathcal{D} \rightarrow V(\mathcal{D})$ que l'on vient d'introduire sont inverses l'un de l'autre et donc que la catégorie des (φ, ∂) -modules sur \mathcal{R}_L est équivalente à celle des $(\mathcal{G}_{k((T))}, N)$ -modules sur L , ce qui en fournit une description particulièrement simple. (Pour une description conjecturale de cette catégorie en termes purement différentiels, voir [19, p. 671-672].)

D'autre part, à une représentation V de $\mathcal{G}_{k((T))}$ d'image finie (resp. à un (φ, ∂) -module \mathcal{D} , plus généralement à un ∂ -module *soluble*), on sait associer un entier, à savoir son conducteur de Swan $\text{Sw}(V)$ (resp. son irrégularité $\text{Irr}(\mathcal{D})$) et Tsuzuki a montré [71] (voir [30, 58] pour d'autres démonstrations) que l'on a

$$\text{Irr}(\mathcal{D}(V)) = \text{Sw}(V).$$

0.3. Anneaux de Fontaine et représentations p -adiques

0.3.1. *Notations.* On se fixe une clôture algébrique $\overline{\mathbf{Q}}_p$ de \mathbf{Q}_p et un système $(\varepsilon^{(n)})_{n \in \mathbf{N}}$ d'éléments de $\overline{\mathbf{Q}}_p$ vérifiant $\varepsilon^{(0)} = 1$, $\varepsilon^{(1)} \neq 1$ et $(\varepsilon^{(n+1)})^p = \varepsilon^{(n)}$ si $n \in \mathbf{N}$. Ceci fait de $\varepsilon^{(n)}$ une racine primitive p^n -ième de l'unité et on note F_n le corps $\mathbf{Q}_p(\varepsilon^{(n)})$ et F_∞ l'extension cyclotomique de \mathbf{Q}_p réunion des F_n . Soit \mathbf{C}_p le complété de $\overline{\mathbf{Q}}_p$ pour la valuation v_p .

On note $\mathcal{G}_{\mathbf{Q}_p}$ le groupe de Galois $\text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ et $\chi : \mathcal{G}_{\mathbf{Q}_p} \rightarrow \mathbf{Z}_p^*$ le caractère cyclotomique. Soit aussi $\mathcal{H}_{\mathbf{Q}_p}$ le noyau de la restriction de χ à $\mathcal{G}_{\mathbf{Q}_p}$ de telle sorte que $\mathcal{H}_{\mathbf{Q}_p} = \text{Gal}(\overline{\mathbf{Q}}_p/F_\infty)$ et soit $\Gamma_{\mathbf{Q}_p} = \mathcal{G}_{\mathbf{Q}_p}/\mathcal{H}_{\mathbf{Q}_p} = \text{Gal}(F_\infty/\mathbf{Q}_p)$.

(On peut remplacer \mathbf{Q}_p par un corps complet pour la valuation v_p , de corps résiduel parfait, et c'est ce qui est fait dans le texte principal, mais travailler avec \mathbf{Q}_p a l'avantage de simplifier certaines formules.)

0.3.2. *Le programme de Fontaine.* Soit G un groupe topologique (comme $\mathcal{G}_{\mathbf{Q}_p}$ ou $\mathcal{H}_{\mathbf{Q}_p}$). Pour mettre un peu d'ordre dans les \mathbf{Q}_p -représentations de G , la stratégie de Fontaine est de construire des anneaux topologiques munis d'une action de G et de structures additionnelles respectées par l'action de G . Chacun de ces anneaux permet de découper dans l'ensemble des \mathbf{Q}_p -représentations de G celles qui sont B -admissibles (une représentation V de dimension d est dite B -admissible si $B \otimes_{\mathbf{Q}_p} V$ est isomorphe à B^d en tant que G -module). Si V est une représentation B -admissible de \mathcal{G}_K , le B^G -module $\mathbf{D}_B(V) = (B \otimes V)^G$ est libre de rang $\dim_{\mathbf{Q}_p} V$ et est muni de toutes les structures additionnelles de B respectées par l'action de G . Ceci permet d'associer aux représentations de G des invariants plus maniables (en général des objets provenant de l'algèbre linéaire) et, si l'anneau B est assez fin (i.e. si on peut retrouver \mathbf{Q}_p à l'intérieur de B en utilisant les structures respectées par G), de classifier les représentations B -admissibles en termes de ces invariants. Cette approche a l'avantage de ramener l'étude de toutes les représentations B -admissibles à celle de l'anneau B ; tout l'art résidant dans la construction d'anneaux intéressants. Si

V est B-admissible, la matrice permettant de passer d'une base de $\mathbf{D}_B(G)$ sur B^G à une base de V sur \mathbf{Q}_p appartient à $GL_d(B)$; ses coefficients sont < les B-périodes > de la représentation V.

0.3.3. *Quelques anneaux de Fontaine.* La théorie du corps des normes de Fontaine et Wintenberger [43] et [76] permet d'associer à toute extension finie K de \mathbf{Q}_p un corps \mathbf{E}_K de caractéristique p, complet pour une valuation discrète. Elle permet de munir la clôture séparable $\mathbf{F}_p((T))^{\text{sep}}$ de $\mathbf{F}_p((T))$ d'une action naturelle de $\mathcal{G}_{\mathbf{Q}_p}$ identifiant $\mathcal{H}_{\mathbf{Q}_p}$ à $\mathcal{G}_{\mathbf{F}_p((T))}$; le groupe $\mathcal{G}_{\mathbf{Q}_p}$ agit alors sur $\mathbf{E}_{\mathbf{Q}_p} = \mathbf{F}_p((T))$ à travers $\Gamma_{\mathbf{Q}_p}$ par $\sigma(T) = (1 + T)^{\chi(\sigma)} - 1$.

On peut relever cette action en caractéristique 0 à $\mathbf{Z}_p[[T]]$ par la même formule et, si on fait agir le Frobenius φ par $T \rightarrow (1 + T)^p - 1$, les actions de φ et $\mathcal{G}_{\mathbf{Q}_p}$ sur $\mathbf{Z}_p[[T]]$ commutent. Ces actions s'étendent par continuité à $\mathcal{E}_{\mathbf{Q}_p}^\dagger$, $\mathcal{R}_{\mathbf{Q}_p}$ et $\mathcal{R}_{\mathbf{Q}_p}[\log T]$. On dispose d'autre part de < flèches de localisation de $\varepsilon^{(n)} - 1$ > commutant à l'action de Galois, $\mathcal{R}_{\mathbf{Q}_p}[\log T] \dashrightarrow F_n[[T]]$; cette flèche est en pointillés car elle n'est définie, par la formule

$$f \rightarrow \varphi^{-n}(f) = f(\varepsilon^{(n)}(1 + T)^{p^{-n}} - 1) = f\left(\varepsilon^{(n)} - 1 + \varepsilon^{(n)}\left(\sum_{k=1}^{+\infty} \binom{p^{-n}}{k} T^k\right)\right),$$

que sur le sous-anneau des $f \in \mathcal{R}_{\mathbf{Q}_p}[\log T]$ qui convergent en $\varepsilon^{(n)} - 1$. On obtient le diagramme d'anneaux

$$\mathcal{E}_{\mathbf{Q}_p}^\dagger \longrightarrow \mathcal{R}_{\mathbf{Q}_p} \longrightarrow \mathcal{R}_{\mathbf{Q}_p}[\log T] \dashrightarrow F_n[[T]] \longrightarrow F_n.$$

La procédure⁵ permettant de construire $\mathcal{E}_{\mathbf{Q}_p}^\dagger$, $\mathcal{R}_{\mathbf{Q}_p}$, ... à partir de $\mathbf{F}_p((T))$, permet, en partant de $\mathbf{F}_p((T))^{\text{sep}}$ d'obtenir le diagramme d'anneaux

$$\mathbf{B}^\dagger \longrightarrow \widetilde{\mathbf{B}}_{\text{rig}}^\dagger \longrightarrow \widetilde{\mathbf{B}}_{\text{log}}^\dagger \dashrightarrow \mathbf{B}_{\text{dR}}^+ \longrightarrow \mathbf{C}_p.$$

et comme on a muni $\mathbf{F}_p((T))^{\text{sep}}$ d'une action de $\mathcal{G}_{\mathbf{Q}_p}$ grâce à la théorie du corps des normes, tous ces anneaux sont munis d'une action continue de $\mathcal{G}_{\mathbf{Q}_p}$ commutant à celle de φ sur \mathbf{B}^\dagger , $\widetilde{\mathbf{B}}_{\text{rig}}^\dagger$ et $\widetilde{\mathbf{B}}_{\text{log}}^\dagger$.

0.3.4. *La hiérarchie [39] des représentations galoisiennes.* Une représentation de \mathcal{G}_K qui est $\mathbf{B}_{\text{dR}}^+[\frac{1}{\log(1+T)}]$ -admissible est dite *de de Rham*; elle est dite *crystalline* si elle est $\widetilde{\mathbf{B}}_{\text{rig}}^\dagger[\frac{1}{\log(1+T)}]$ -admissible, *semi-stable* si elle est $\widetilde{\mathbf{B}}_{\text{log}}^\dagger[\frac{1}{\log(1+T)}]$ -admissible⁶, et potentiellement semi-stable

⁵Pour construire $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ à partir de $\mathbf{F}_p((T))$, il y a un choix : le choix d'un relèvement de T ou, ce qui revient au même, d'un Frobenius sur $\mathcal{E}_{\mathbf{Q}_p}^\dagger$; une fois ce choix fait, la construction s'étend canoniquement à $\mathbf{F}_p((T))^{\text{sep}}$. Notre choix de T en fait un analogue p-adique de $e^{2i\pi} - 1$ et $\log(1 + T)$ devient un analogue p-adique de $2i\pi$.

⁶La définition habituelle fait intervenir des anneaux $\mathbf{B}_{\text{cris}}^+$ et \mathbf{B}_{st}^+ qui sont étroitement liés à $\mathbf{B}_{\text{rig}}^\dagger$ et $\mathbf{B}_{\text{log}}^\dagger$; en particulier les seuls éléments intéressants pour ces histoires de classification des représentations vivent dans des sous-F-espaces vectoriels de dimension finie stables par φ , et ceux-ci sont dans l'intersection $\mathbf{B}_{\text{rig}}^+ = \mathbf{B}_{\text{rig}}^\dagger \cap \mathbf{B}_{\text{cris}}^+$ (resp. $\mathbf{B}_{\text{log}}^+ = \mathbf{B}_{\text{log}}^\dagger \cap \mathbf{B}_{\text{st}}^+$). Passer de $\mathbf{B}_{\text{cris}}^+$ à $\mathbf{B}_{\text{rig}}^+$ revient à considérer la cohomologie

si ses périodes de de Rham appartiennent au sous-anneau de $\mathbf{B}_{\text{dR}}^+[\frac{1}{\log(1+T)}]$ engendré par $\widetilde{\mathbf{B}}_{\log[\frac{1}{\log(1+T)}]}^\dagger$ et $\overline{\mathbf{Q}}_p$.

Les implications \langle cristalline \Rightarrow semi-stable \Rightarrow potentiellement semi-stable \Rightarrow de Rham \rangle sont immédiates ; les deux premières sont strictes et la conjecture de monodromie p -adique de Fontaine est que la dernière implication est en fait une équivalence. Cette conjecture était connue en dimension 1 depuis longtemps ; Fontaine [40] l'avait récemment démontrée en dimension 2 et Hyodo [48] l'avait vérifiée pour les extensions de représentations semi-stables. D'autre part, Tsuji [69, 70] et Faltings [33] avaient démontré (à la suite de travaux de nombreuses personnes dont Tate, Bloch, Kato, Fontaine, Messing, Faltings, Hyodo etc.) la conjecture C_{st} de Fontaine, prouvant ainsi que les représentations provenant de la géométrie (du moins une bonne partie d'entre elles) sont potentiellement semi-stables (il faut en outre utiliser les altérations de de Jong).

0.3.5. *La théorie des (φ, Γ) -modules.* La théorie des (φ, Γ) -modules de Fontaine ([37]+[11]) établit une équivalence de catégories entre, d'une part, les \mathbf{Q}_p -représentations de $\mathcal{G}_{\mathbf{Q}_p}$ (i.e. les \mathbf{Q}_p -espaces vectoriels de dimension finie munis d'une action continue de $\mathcal{G}_{\mathbf{Q}_p}$) et, d'autre part, les (φ, Γ) -modules étales sur $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ (i.e. les $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ -espaces vectoriels de dimension finie munis d'actions semi-linéaires de φ et $\Gamma_{\mathbf{Q}_p}$ commutant entre elles tels que φ soit étale, c'est-à-dire de φ -pente 0). Cette équivalence de catégorie est obtenue par la procédure décrite plus pour le groupe $\mathcal{H}_{\mathbf{Q}_p}$, en utilisant l'anneau \mathbf{B}^\dagger : elle associe à V le $\mathcal{E}_{\mathbf{Q}_p}^\dagger = (\mathbf{B}^\dagger)^{\mathcal{H}_{\mathbf{Q}_p}}$ -espace vectoriel $\mathbf{D}^\dagger(V) = (\mathbf{B}^\dagger \otimes V)^{\mathcal{H}_{\mathbf{Q}_p}}$.

Comme $\Gamma_{\mathbf{Q}_p} \cong \mathbf{Z}_p^*$ est essentiellement procyclique, on peut retraduire l'équivalence de catégories ci-dessus en disant qu'une \mathbf{Q}_p -représentation de $\mathcal{G}_{\mathbf{Q}_p}$ de dimension d peut se décrire entièrement à l'aide de deux matrices A et B de $GL_d(\mathcal{O}_{\mathcal{E}, \mathbf{Q}_p}^\dagger)$ (où $\mathcal{O}_{\mathcal{E}, \mathbf{Q}_p}^\dagger$ désigne l'anneau des entiers de $\mathcal{E}_{\mathbf{Q}_p}^\dagger$) vérifiant la relation de commutation $A\varphi(B) = B\gamma(A)$, ce qui semble plus simple *a priori* que de décrire la représentation directement.

0.3.6. *Le (φ, ∇) -module attaché à une représentation galoisienne.* On peut transformer un (φ, Γ) -module en un φ -module avec connexion en considérant l'action ∇_∇ de l'algèbre de Lie de $\Gamma_{\mathbf{Q}_p}$ (i.e. l'action infinitésimale de $\Gamma_{\mathbf{Q}_p}$). Le passage à l'algèbre de Lie introduit des dénominateurs (la série définissant le logarithme n'est pas à coefficients entiers), et on doit étendre les coefficients de $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ à $\mathcal{R}_{\mathbf{Q}_p}$. L'étude du φ -module à connexion ainsi obtenu a été menée à bien par Berger [5], ce qui lui permet d'obtenir un certain nombre de résultats concernant la classification des représentations p -adiques.

Ce module différentiel ne rentre pas tout à fait dans le cadre du paragraphe précédent car, sur les coefficients, la dérivation est $\nabla = (1+T)\log(1+T)\frac{d}{dT}$ et $\log(1+T) \notin \mathcal{E}_{\mathbf{Q}_p}^\dagger$, ce qui fait que la connexion a une infinité de singularités régulières en les zéros de $\log(1+T)$, c'est-à-dire les $\zeta - 1$, $\zeta \in \mathbb{Z}_{p^\infty}$. Le résidu en chacune de ces singularités est l'opérateur de

rigide au lieu de la cohomologie cristalline : $\mathbf{B}_{\text{cris}}^+$ peut s'interpréter [36] comme le H^0 cristallin de l'anneau des entiers de $\overline{\mathbf{F}}$ et $\mathbf{B}_{\text{rig}}^+$ comme son H^0 rigide.

Sen (cf. [66, 22, 41]) dont les valeurs propres sont les < poids de Hodge-Tate généralisés > de V ; en particulier, ce résidu est nul si et seulement si V est \mathbf{C}_p -admissible.

D'autre part, utilisant la flèche $\mathcal{E}_{\mathbf{Q}_p}^\dagger \dashrightarrow F_n[[t]]$, on montre [41] que la représentation V est de de Rham si et seulement si les singularités ne sont qu'apparentes. Pour ramener la conjecture de Fontaine à celle de Crew, Berger résout simultanément toutes ces singularités en construisant un $\mathcal{R}_{\mathbf{Q}_p}$ -réseau de $\mathcal{R}_{\mathbf{Q}_p}[\frac{1}{\log(1+T)}] \otimes_{\mathcal{E}_{\mathbf{Q}_p}^\dagger} \mathbf{D}^\dagger(V)$ stable par $\partial_V = \frac{1}{\log(1+T)} \nabla_V$ et par φ et montre que le (φ, ∂) -module ainsi obtenu est quasi-unipotent si et seulement si V est potentiellement semi-stable.

1. ÉQUATIONS DIFFÉRENTIELLES p -ADIQUES

1.1. Anneaux de séries de Laurent

Soit L un corps de caractéristique 0 complet pour la valuation p -adique v_p supposée discrète (on pourrait remplacer cette hypothèse par maximalement complet). On suppose, pour se simplifier la vie, que le corps résiduel k de L est algébriquement clos (ça évitera d'avoir à étendre les scalaires à certains endroits). Remarquons que cette restriction n'est pas très sérieuse car, pour résoudre une équation différentielle, on peut toujours étendre les scalaires puis utiliser le théorème d'Ax-Sen-Tate (prop. 2.1) pour revenir au corps de base.

Soit \mathcal{E}_L l'ensemble des séries de Laurent $\sum_{k \in \mathbf{Z}} a_k T^k$ telle que la suite $(v_p(a_k))_{k \in \mathbf{Z}}$ soit minorée et vérifie $\lim_{k \rightarrow -\infty} v_p(a_k) = +\infty$.

Si $r > 0$, soit $\mathcal{E}_L^{[0,r]}$ (resp. $\mathcal{E}_L^{(0,r]}$) l'anneau des fonctions analytiques (resp. analytiques bornées) sur la couronne $0 < v_p(T) \leq r$. C'est aussi l'ensemble des séries de Laurent $\sum_{k \in \mathbf{Z}} a_k T^k$ vérifiant $\lim_{k \rightarrow -\infty} v_p(a_k) + kr = +\infty$ et $\lim_{k \rightarrow +\infty} v_p(a_k) + ks = +\infty$ quel que soit $0 < s \leq r$ (resp. la suite $(v_p(a_k))_{k \in \mathbf{Z}}$ est minorée). On peut aussi obtenir $\mathcal{E}_L^{[0,r]}$ en complétant $\mathcal{E}_L^{(0,r]}$ pour la topologie de Fréchet induite par la famille de valuations w_s , $0 < s \leq r$ définies par $w_s(\sum_{k \in \mathbf{Z}} a_k T^k) = \inf_{k \in \mathbf{Z}} (v_p(a_k) + sr)$.

On obtient alors \mathcal{E}_L^\dagger (resp. \mathcal{R}_L) comme limite inductive des espaces vectoriels topologiques $\mathcal{E}_L^{(0,r]}$ (resp. $\mathcal{E}_L^{[0,r]}$). Finalement, soient \mathcal{E}_L^+ et \mathcal{R}_L^+ les intersections respectives de \mathcal{E}_L et \mathcal{R}_L avec $L[[T]]$; ceci fait de \mathcal{R}_L^+ (resp. \mathcal{E}_L^+) l'anneau des fonctions analytiques (res. analytiques bornées) sur le disque $0 < v_p(T)$. Tout élément de \mathcal{R}_L peut alors s'écrire comme la somme d'un élément de \mathcal{E}_L^\dagger et d'un élément de \mathcal{R}_L^+ et on a $\mathcal{R}_L^+ \cap \mathcal{E}_L^\dagger = \mathcal{E}_L^+$.

De manière imagée, \mathcal{E}_L est l'anneau des fonctions analytiques sur une couronne vide de rayon 1 et \mathcal{R}_L (resp. \mathcal{E}_L^\dagger) est l'anneau des fonctions analytiques (resp. analytiques bornées) sur une couronne infiniment fine de rayon 1.

PROPOSITION 1.1. — (i) \mathcal{E}_L^+ et $\mathcal{E}_L^{(0,r]}$, $r > 0$, sont des anneaux principaux;

(ii) \mathcal{E}_L et \mathcal{E}_L^\dagger sont des corps;

- (iii) \mathcal{R}_L , \mathcal{R}_L^+ et $\mathcal{E}_L^{[0,r]}$, $r > 0$, sont des anneaux de Bézout;
 (iv) tout sous-module fermé d'un module libre M de rang fini sur \mathcal{R}_L , \mathcal{R}_L^+ ou $\mathcal{E}_L^{(0,r]}$, $r > 0$, est libre de rang $\leq \text{rang}(M)$.

Démonstration. — Les (i) et (ii) sont des conséquences de la théorie des polygones de Newton (cf. [32], par exemple) pour les séries de Laurent (et du fait que l'on a supposé L de valuation discrète). Le (iii) est un résultat de Lazard [55]; rappelons qu'un anneau de Bézout est un anneau dans lequel tout idéal de type fini est principal. Pour le (iv), voir [5]; ce résultat joue un grand rôle dans la réduction de la conjecture de Fontaine à celle de Crew.

On se fixe une dérivation ∂ et un Frobenius φ de \mathcal{E}_L^\dagger . Si E est une extension finie séparable de $k((T))$, il lui correspond des extensions $\mathcal{E}_L^\dagger(E)$ et $\mathcal{R}_L(E)$ de \mathcal{E}_L^\dagger et \mathcal{R}_L respectivement auxquelles les actions de φ et ∂ s'étendent canoniquement. Voir le n° 2.2 pour un point de vue plus général. Une propriété que nous aurons à utiliser est l'existence (cf. prop 2.3) d'un isomorphisme d'anneaux topologiques de \mathcal{R}_L sur $\mathcal{R}_L(E)$ [E est topologiquement isomorphe à $k((T))$]; en particulier, après extension séparable finie de $k((T))$, on se retrouve avec la même théorie.

1.2. Les résultats de Christol et Mebkhout

Les travaux de Christol et Mebkhout ont déjà fait l'objet d'un exposé à ce séminaire [56]; nous nous contenterons donc d'un résumé rapide. Pour une présentation détaillée de la théorie, nous renvoyons à [20].

1.2.1. *Valuation de convergence d'un module différentiel.* Rappelons que, si $0 < s \leq r$, on définit sur $\mathcal{E}^{[0,r]}$ une valuation w_s par la formule $w_s(\sum_{k \in \mathbf{Z}} a_k T^k) = \inf_{k \in \mathbf{Z}} v_p(a_k) + ks$. Si $f \in \mathcal{R}_L$, la fonction $s \rightarrow w_s(f)$ est définie sur un intervalle de la forme $]0, r]$, $r > 0$ et est une fonction concave de s , affine par morceaux. Une manière commode de voir w_s est d'introduire $\langle \text{le } \rangle$ point générique t_s de valuation s ; il vit dans une extension transcendante algébriquement close L_s de L , complète pour une valuation étendant v_p (et encore notée v_p), et est caractérisé (à automorphisme de corps valué près) par $v_p(t_s - a) = s$ quel que soit $a \in L$ de valuation s . On a alors $w_s(f) = v_p(f(t_s))$ si $f \in \mathcal{E}^{[0,r]}$ et $r \geq s$. (Pour construire L_s , on munit $L[X, X^{-1}]$ de la valuation $v_p(\sum_{k=k_0}^{k_1} a_k X^k) = \inf_{k \in \mathbf{Z}} v_p(a_k) + ks$; cette valuation est multiplicative et donc passe au corps des fractions; on peut alors compléter ce dernier, prendre une clôture algébrique et recompléter (maximalement si besoin est) pour obtenir L_s et prendre $t_s = X$.)

Si $G = (g_{i,j}) \in M_d(\mathcal{R}_L)$, on définit $w_s(G)$ comme le minimum des $w_s(g_{i,j})$, $1 \leq i, j \leq d$.

Soient \mathcal{D} un ∂ -module de rang d sur \mathcal{R}_L , e_1, \dots, e_d une base de \mathcal{D} sur \mathcal{R}_L et, si $n \in \mathbf{N}$, soit $G_n = (g_{i,j})$ la matrice définie par $\partial_{\mathcal{D}}^n e_j = \sum_{i=1}^d g_{i,j} e_i$. Si $s > 0$, la suite de terme

général $\frac{1}{n}w_s(G_n)$ a une limite (quand elle est définie) quand n tend vers $+\infty$ (norme spectrale d'un opérateur). Ceci permet d'introduire la fonction

$$\text{Val}(\mathcal{D}, s) = \sup\left(s, \lim_{n \rightarrow +\infty} \frac{v_p(n!) - w_s(G_n)}{n}\right) = \sup\left(s, \frac{1}{p-1} - \lim_{n \rightarrow +\infty} \frac{w_s(G_n)}{n}\right).$$

La quantité $\text{Val}(\mathcal{D}, s)$ s'interprète comme la valuation de convergence des solutions de l'équation différentielle $\partial_{\mathcal{D}}X = 0$ autour en t_s . (Une base des solutions autour d'un point x_0 est donnée par les séries $\sum_{n=0}^{+\infty} \partial_{\mathcal{D}}^n e_i \cdot \frac{(x_0 - T)^n}{n!}$, $1 \leq i \leq d$.)

Pour calculer cette fonction, on peut, dans les cas favorables, utiliser un résultat de Young [77] :

PROPOSITION 1.2. — Soient $\mathcal{L} = \partial^d + a_{d-1}(T)\partial^{d-1} + \dots + a_0(T)$, où $a_0(T), \dots, a_{d-1}(T) \in \mathcal{R}_L$ (et $\partial = \frac{d}{dT}$), $P = X^d + a_{d-1}(t_s)X^{d-1} + \dots + a_0(t_s)$ et u_1, \dots, u_d les valuations des racines de P (ces valuations se lisent sur le polygone de Newton de P), alors $\mathcal{L}X = 0$ admet une base de solutions f_1, \dots, f_d autour de t_s dont les valuations de convergence vérifient :

- (i) si $u_i < s$, la valuation de convergence de f_i est exactement $\frac{1}{p-1} + u_i$;
- (ii) si $u_i \geq s$, la valuation de convergence de f_i est $\leq \frac{1}{p-1} + s$.

Passer d'un module différentiel à un opérateur différentiel demande de trouver un vecteur cyclique (i.e. un vecteur X tel que $X, \partial_{\mathcal{D}}X, \dots, \partial_{\mathcal{D}}^{d-1}X$ forment une base de \mathcal{D} sur \mathcal{R}_L). Cela ne peut, en général, se faire sans passer au corps des fractions de \mathcal{R}_L , ce qui introduit quelques petits problèmes techniques pour se débarrasser des singularités apparentes que l'on récupère ainsi.

Un ingrédient technique très utile est l'antécédent de Frobenius. Soit φ un Frobenius sur \mathcal{E}_L^\dagger . Si \mathcal{D} est un ∂ -module sur \mathcal{R}_L , on note $\varphi^*\mathcal{D}$ le transformé de \mathcal{D} par φ : si la matrice de $\partial_{\mathcal{D}}$ dans une base est G , la matrice de $\partial_{\varphi^*\mathcal{D}}$ est $\partial(\varphi(T))\varphi(G)$. On appelle antécédent de Frobenius de \mathcal{D} un ∂ -module \mathcal{N} tel que l'on ait $\varphi^*\mathcal{N} = \mathcal{D}$. Il ne faut pas confondre cette notion avec celle de structure de Frobenius ; un ∂ -module admet une structure de Frobenius (i.e. de (φ, ∂) -module) si $\varphi^*\mathcal{D} \cong \mathcal{D}$. L'existence d'antécédents de Frobenius est garantie sous des conditions assez larges par un théorème de Christol et Dwork [15] :

PROPOSITION 1.3. — Soit \mathcal{D} un ∂ -module sur l'anneau $\mathcal{E}^{]r_1, r_2[}$ des fonctions analytiques sur la couronne $0 < r_1 < v_p(T) < r_2$ tel que $\text{Val}(\mathcal{D}, s) < \frac{1}{p} + s$ si $s \in]r_1, r_2[$. Il existe alors un (unique) φ -module \mathcal{N} sur $\mathcal{E}^{]pr_1, pr_2[}$ tel que $\mathcal{D} = \varphi^*\mathcal{N}$. De plus, on a $\text{Val}(\mathcal{N}, ps) = p\text{Val}(\mathcal{D}, s)$.

On remarquera que, si la valuation de convergence de \mathcal{D} est assez petite (ce qui sera le cas pour les modules solubles, cf. ci-dessous), on peut réitérer le procédé pour se ramener en un point où le théorème de Young peut être utilisé de manière efficace. Une application de ces techniques est la très utile proposition :

PROPOSITION 1.4 ([18]). — La fonction $s \rightarrow \text{Val}(\mathcal{D}, s)$ est convexe, affine par morceaux, et sa dérivée ne prend pour valeurs que des nombres rationnels de dénominateurs $\leq d$.

1.2.2. *La théorie des ∂ -pentes.* Par convexité, la limite $\lim_{s \rightarrow 0^+} \text{Val}(\mathcal{D}, s)$ existe (elle peut être infinie) et le module \mathcal{D} est dit *soluble* si $\lim_{s \rightarrow 0^+} \text{Val}(\mathcal{D}, s) = 0$. Il existe alors un nombre rationnel $\beta \geq 0$, de dénominateur $\leq d$, tel que $\text{Val}(\mathcal{D}, s) = (1 + \beta)s$, si s est assez petit. Ce nombre rationnel β est appelé *la plus grande pente* de \mathcal{D} . On dit que \mathcal{D} est *purement de pente β* si toutes les solutions de l'équation $\partial_{\mathcal{D}}X = 0$ (et pas seulement une solution générale) ont, en t_s (s proche de 0), une valuation de convergence égale à $(1 + \beta)s$.

THÉORÈME 1.5 (< de décomposition > [18, 19]). — *Un ∂ -module soluble \mathcal{D} sur \mathcal{R}_L possède une décomposition canonique $\mathcal{D} = \bigoplus_{\gamma \geq 0} \mathcal{D}_{\gamma}$, où \mathcal{D}_{γ} est purement de pente γ .*

REMARQUE 1.6. — Si on part d'un ∂ -module sur \mathcal{E}_L^{\dagger} , on pourrait espérer obtenir une décomposition sur \mathcal{E}_L^{\dagger} , mais l'exemple du ∂ -module associé à l'opérateur $\mathcal{L} = T^2 \frac{d}{dT} + (3T - \pi) \frac{d}{dT} + 1$ (où $\pi^{p-1} = -p$) montre qu'il n'en est rien ; un calcul montre qu'il possède deux pentes distinctes sur \mathcal{R}_L et que la série $\sum_{n=0}^{+\infty} \pi^{-n} n! T^n$ est une solution de $\mathcal{L}f = 0$ appartenant à \mathcal{R}_L mais pas à \mathcal{E}_L^{\dagger} , ce qui exclut une décomposition sur \mathcal{E}_L^{\dagger} .

THÉORÈME 1.7 (< de Hasse-Arf > [18]). — *Si \mathcal{D} est un ∂ -module soluble, alors la quantité $\text{Irr}(\mathcal{D}) = \sum_{\gamma \geq 0} \gamma \cdot \text{rang } \mathcal{D}_{\gamma}$ est un entier.*

Ce théorème est une étape dans la démonstration du *théorème de l'indice* qui était une des motivations de Christol et Mebkhout (généraliser en rang quelconque les résultats de Robba [61, 62, 63, 64, 21] en rang 1). L'irrégularité $\text{Irr}(\mathcal{D})$ de \mathcal{D} est définie comme l'indice d'un certain opérateur ; c'est donc un entier par définition et le théorème ci-dessus donne une formule permettant de la calculer.

COROLLAIRE 1.8. — *Un ∂ -module soluble irréductible, de rang fini sur \mathcal{R}_L , n'a qu'une seule pente, et le dénominateur de cette pente divise le rang du module.*

Ce qui précède s'applique en particulier aux (φ, ∂) -modules : l'existence d'un Frobenius permet de montrer que l'on a $\text{Val}(\mathcal{D}, s) \leq \frac{1}{p} \text{Val}(\mathcal{D}, ps)$ si s est assez petit, avec égalité si $\text{Val}(\mathcal{D}, ps) < \frac{p}{p-1} + ps$.

1.2.3. *∂ -modules de pente 0.* Si \mathcal{D} est un module différentiel de rang d sur une couronne non vide A du corps des nombres complexes et si $x_0 \in A$, les solutions locales de l'équation $\partial_{\mathcal{D}}X = 0$ autour de x_0 forment un \mathbf{C} -espace vectoriel de dimension d . On peut prolonger analytiquement une solution le long d'un chemin et, au bout d'un tour, on récupère une nouvelle solution autour de x_0 ; d'où un opérateur < de monodromie > N . Les valeurs propres de $\frac{1}{2i\pi} \log N$ s'appellent les exposants de \mathcal{D} ; ce sont des éléments de \mathbf{C}/\mathbf{Z} bien définis à l'ordre près.

En p -adique, on ne dispose pas du prolongement analytique, mais Frobenius sous toutes ses coutures constitue un substitut efficace. En utilisant les antécédents de Frobenius, Christol et Mebkhout [17] (voir aussi [31]) ont réussi à définir les exposants p -adiques

pour un ∂ -module \mathcal{D} sur \mathcal{R}_L purement de ∂ -pente 0 (on dit aussi parfois que \mathcal{D} vérifie la *condition de Robba*). Ces exposants appartiennent à $(\mathbf{Z}_p/\mathbf{Z})^d$ modulo une certaine relation d'équivalence assez compliquée, mais si les différences de ces exposants sont toutes non Liouville (un nombre p -adique α est non Liouville si $\lim_{n \rightarrow \pm\infty} \frac{1}{n} v_p(\alpha - n) = 0$, i.e. si α est entier ou pas trop proche de \mathbf{Z}), alors ces exposants sont des éléments de \mathbf{Z}_p/\mathbf{Z} bien définis à l'ordre près.

THÉORÈME 1.9 (< de la monodromie p -adique > [17]). — *Si \mathcal{D} est un ∂ -module de rang d sur \mathcal{R}_L de ∂ -pente 0, et si les différences de ses exposants sont non Liouville, alors l'équation différentielle $\partial_{\mathcal{D}} X = 0$ admet d solutions linéairement indépendantes dans $\mathcal{R}_L[\mathrm{T}^{\alpha_1}, \dots, \mathrm{T}^{\alpha_d}, \log \mathrm{T}] \otimes_{\mathcal{R}_L} \mathcal{D}$, où $\alpha_1, \dots, \alpha_d$ sont les exposants de \mathcal{D} .*

COROLLAIRE 1.10. — *Un (φ, ∂) -module sur \mathcal{R}_L de ∂ -pente 0 devient unipotent sur une extension de \mathcal{R}_L associée à une extension modérément ramifiée de $k((\mathrm{T}))$.*

Démonstration. — Comme \mathcal{D} est muni d'une structure de Frobenius, l'ensemble des exposants de \mathcal{D} est stable par $\alpha \rightarrow p\alpha$ et les exposants de \mathcal{D} sont des nombres rationnels de dénominateur premier à p .

1.3. Théorie de Galois différentielle

Ce n° est consacré à la démonstration d'André [2]. Celui-ci se place dans un cadre abstrait s'appliquant à d'autres situations que celles des ∂ -modules solubles sur l'anneau de Robba, et nous avons choisi de nous restreindre à ce cadre pour nous concentrer sur la construction qui est au cœur de la démonstration (il s'agit du module \mathcal{M}_ℓ^\otimes ci-dessous, obtenu par induction tensorielle) et son utilisation.

1.3.1. *Filtration des groupes de Galois différentiels.* On peut associer à un ∂ -module \mathcal{D} de rang d sur \mathcal{R}_L un groupe de Galois différentiel $G_{\mathcal{D}}$; c'est un sous-groupe algébrique de GL_d défini sur une extension finie L' de L ; ses représentations algébriques sont en correspondance avec les ∂ -modules de rang fini sur $\mathcal{R}_{L'}$ que l'on obtient à partir de \mathcal{D} et de son dual par produit tensoriel, sous-objet et quotient. Réciproquement, si \mathcal{D}' est un tel ∂ -module, son groupe de Galois différentiel est un quotient de $G_{\mathcal{D}}$. Nous utiliserons cette correspondance pour passer du langage des représentations à celui des ∂ -modules sans plus de commentaire.

Si $\gamma \geq 0$, on note $G_{\mathcal{D}}^\gamma$ le plus grand sous-groupe distingué de $G_{\mathcal{D}}$ tel que les représentations de $G_{\mathcal{D}}/G_{\mathcal{D}}^\gamma$ n'aient que des ∂ -pentes $\leq \gamma$. Comme $G_{\mathcal{D}}$ est algébrique, toute suite strictement décroissante de sous-groupes algébriques est de longueur finie. En particulier, la filtration de $G_{\mathcal{D}}$ par les ∂ -pentes n'a qu'un nombre fini de sauts (un saut est un réel $\gamma > 0$ tel $G_{\mathcal{D}}^{\gamma-\varepsilon} \neq G_{\mathcal{D}}^\gamma$ si $\varepsilon > 0$) $\gamma_1, \dots, \gamma_r$ et on appelle $\gamma_0 = 0, \gamma_1, \dots, \gamma_r$ les ∂ -pentes de $G_{\mathcal{D}}$.

REMARQUE 1.11. — (i) Si \mathcal{D}' est un sous- ∂ -module ou un quotient de \mathcal{D} , le théorème de décomposition montre que les ∂ -pentes de \mathcal{D}' sont incluses dans celles de \mathcal{D} ; en particulier, elles sont inférieures ou égales à la plus grande pente de \mathcal{D} .

(ii) Si \mathcal{D}_1 et \mathcal{D}_2 sont deux ∂ -modules sur \mathcal{R}_L , les pentes de $\mathcal{D}_1 \otimes \mathcal{D}_2$ sont inférieures ou égales à la plus grande des plus grandes ∂ -pentes de \mathcal{D}_1 et \mathcal{D}_2 : la valuation de convergence du produit de deux séries entières est inférieure ou égale au maximum des valuations de convergence de ces séries.

On déduit de cette remarque que les ∂ -pentes d'une représentation de $G_{\mathcal{D}}$ sont incluses dans celles de $G_{\mathcal{D}}$ et qu'elles sont $\leq \gamma$ si et seulement si $G_{\mathcal{D}}$ agit à travers $G_{\mathcal{D}}/G_{\mathcal{D}}^{\gamma}$.

1.3.2. *La conjecture de Crew.* Si \mathcal{D} est un ∂ -module sur \mathcal{R}_L et E est une extension finie séparable de $k((T))$, le $\mathcal{R}_L(E)$ -module $\mathcal{R}_L(E) \otimes_{\mathcal{R}_L} \mathcal{D}$ est un ∂ -module sur $\mathcal{R}_L(E)$ dont le groupe de Galois différentiel $G_{\mathcal{D}}(E)$ est un sous-groupe de $G_{\mathcal{D}}$. La conjecture de Crew (sous sa forme filtrée) peut se reformuler sous la forme : \langle si \mathcal{D} est un (φ, ∂) -module irréductible sur \mathcal{R}_L , il existe une extension finie séparable E de $k((T))$ telle que $G_{\mathcal{D}}(E) = \{1\} \rangle$. Le cas de rang 1 permet de démontrer que cet énoncé est vrai si $G_{\mathcal{D}}$ est résoluble et, plus généralement, il permet de prouver qu'il existe une extension finie séparable E de $k((T))$ telle que $G_{\mathcal{D}}(E)$ n'ait pas de quotient abélien non trivial. Quitte à faire une extension séparable finie de $k((T))$, on est donc ramené à vérifier le résultat suivant.

PROPOSITION 1.12. — *Si \mathcal{D} est un (φ, ∂) -module irréductible sur \mathcal{R}_L , alors tout quotient simple de $G_{\mathcal{D}}$ est abélien.*

La démonstration se fait par l'absurde. Le principe est de faire apparaître des dénominateurs dans les pentes en induisant à partir d'une extension modérée de $k((T))$ pour obtenir une contradiction avec le théorème de Hasse-Arf.

Soit H un quotient simple non abélien de $G_{\mathcal{D}}$. Comme H est simple, il a exactement une ∂ -pente γ , et celle-ci est non nulle le corollaire 1.10. Soit \mathcal{M} une représentation irréductible de H de dimension $d > 1$; alors \mathcal{M} est de pente γ . Soit $\ell \neq p$ un nombre premier ne divisant ni d , ni le numérateur de γ , ni l'ordre du groupe $\text{Out}(H)$, et soit ζ une racine primitive ℓ -ième de l'unité. Si $0 \leq i \leq \ell - 1$, soit $\mathcal{M}^{(i)}$ le $\mathcal{R}_L(T^{1/\ell})$ - ∂ -module obtenu à partir de \mathcal{M} via le changement de variable $T \rightarrow \zeta^i T^{1/\ell}$; si $x \in \mathcal{M}^{(0)}$, on note $x^{(i)}$ son image dans $\mathcal{M}^{(i)}$ via le changement de variable $T^{1/\ell} \rightarrow \zeta^i T^{1/\ell}$ (et donc $x^{(0)} = x$). Considérons les ∂ -modules \mathcal{M}_{ℓ} et $\mathcal{M}_{\ell}^{\otimes}$ définis par :

$$\mathcal{M}_{\ell} = \{x^{(0)}, \dots, x^{(\ell-1)}, x \in \mathcal{M}^{(0)}\},$$

et $\mathcal{M}_{\ell}^{\otimes}$ est le sous- \mathcal{R}_L -module de

$$\mathcal{M}^{(0)} \otimes_{\mathcal{R}_L(T^{1/\ell})} \cdots \otimes_{\mathcal{R}_L(T^{1/\ell})} \mathcal{M}^{(\ell-1)}$$

engendré par les $x^{(0)} \otimes \cdots \otimes x^{(\ell-1)}$, $x \in \mathcal{M}^{(0)}$. L'intérêt du passage de \mathcal{M} à \mathcal{M}_{ℓ} est que la pente a été divisée par ℓ (comme la dimension a été multipliée par ℓ , cela ne fournit pas de contradiction au théorème de Hasse-Arf). Quand on étend les scalaires de $\mathcal{R}_L(T)$ à $\mathcal{R}_L(T^{1/\ell})$, le module \mathcal{M}_{ℓ} devient isomorphe à la somme des $\mathcal{M}^{(i)}$; comme le groupe de Galois différentiel de chacun des $\mathcal{M}^{(i)}$ est isomorphe à H , le groupe de Galois différentiel

H_ℓ de \mathcal{M}_ℓ est un sous-groupe du produit en couronne de H par le groupe cyclique C_ℓ (ce produit en couronne U est un produit semi-direct $1 \rightarrow H^\ell \rightarrow U \rightarrow C_\ell \rightarrow 1$, où C_ℓ agit par permutation circulaire des facteurs). Comme H est simple, il y a a priori 3 cas :

- (I) H_ℓ est le produit direct de C_ℓ et H ;
- (II) H_ℓ est un produit semi-direct de C_ℓ et H ;
- (III) H_ℓ est égal au produit en couronne.

Dans tous les cas, une suite de sous-groupes distingués de H_ℓ est au plus de longueur 3 et H_ℓ a au plus 3 pentes. Deux de ces pentes sont nulles [une par définition et l'autre car le quotient C_ℓ de H_ℓ correspond à l'extension modérée $\mathcal{R}_L(T^{1/\ell})$ de $\mathcal{R}_L(T)$], et la troisième est la pente de \mathcal{M}_ℓ , c'est-à-dire γ/ℓ . Pour conclure, il suffit de montrer que ces trois cas sont exclus par notre choix de ℓ .

Dans le premier cas, l'action de C_ℓ découpe \mathcal{M}_ℓ en ℓ morceaux de rang d et ∂ -pente γ/ℓ , et comme $d\gamma/\ell$ n'est pas entier, cela contredit le théorème de Hasse-Arf. Le second cas est exclu par la condition ℓ ne divise pas l'ordre de $\text{Out}(H)$. Dans le troisième cas, le module \mathcal{M}_ℓ^\otimes est une représentation irréductible de H_ℓ de dimension d^ℓ ; comme l'action de H_ℓ ne se factorise pas à travers C_ℓ (car \mathcal{M} est irréductible, de dimension > 1), la ∂ -pente de ce module est γ/ℓ , ce qui contredit le théorème de Hasse-Arf puisque $d^\ell\gamma/\ell$ n'est pas entier.

1.4. Le théorème de Turrittin p -adique

Dans ce n° , nous esquissons la démonstration du < théorème de Turrittin p -adique > de Mebkhout [60], la conjecture de Crew en étant un cas particulier. Pour simplifier les calculs, on utilise la dérivation $\partial = T \frac{d}{dT}$ et on note π le < π de Dwork > (i.e. une racine $(p-1)$ -ième de $-p$).

1.4.1. *Construction de ∂ -modules solubles de rang 1.* Si P est un polynôme sans terme constant, on note $\chi(P)$ le module de rang 1 dont $\exp(\pi P(T^{-1}))$ est une base des sections horizontales.

PROPOSITION 1.13. — Soient $a \in L$ vérifiant $v_p(a) = 0$ et $\beta \in \mathbf{N} - \{0\}$. Si $v_p(\beta) = n$, il existe un polynôme $P_{\beta,a}(X) = \frac{a}{\beta}X^\beta + \frac{pb_1}{\beta}X^{\beta/p} + \dots + \frac{p^n b_n}{\beta}X^{\beta/p^n}$, où les β_i sont des entiers d'une extension convenable de K , tel que le module $\chi(P_{\beta,a})$ soit soluble. De plus, ce module

- (i) est de ∂ -pente β ;
- (ii) peut être muni d'une structure de Frobenius.

L'existence de $P_{\beta,a}$ est due à Robba [62]. Pour Robba, l'extension convenable dont il est question dans la proposition est assez imposante (c'est un corps algébriquement clos maximale complet), mais Matsuda [57] a montré que l'on peut prendre les β_i dans une extension finie de L , du moins si $p \neq 2$. Le (i) se démontre alors par un calcul direct et le (ii) est dû à Chiarellotto et Christol [13].

1.4.2. *Modules de ∂ -pente entière.* Pour rester dans le cadre < valuation discrète >, nous supposons désormais $p \neq 2$. La démonstration du théorème de Turritin p -adique repose sur l'énoncé suivant

PROPOSITION 1.14 ([60]). — (i) Si \mathcal{D} est un ∂ -module irréductible soluble sur \mathcal{R}_L de ∂ -pente entière $\beta > 0$, alors il existe $h \in \mathbf{N}$, une extension finie L' de L et $a \in L'$ de valuation 0 tel que le module $\mathcal{D} \otimes \chi(P_{\beta,a})$ soit de pente $< \beta$.

(ii) Si \mathcal{D} est un ∂ -module soluble de rang premier à p n'ayant qu'une seule ∂ -pente $\beta \in \mathbf{N}$, alors \mathcal{D} peut se casser en deux.

La démonstration de cette proposition est assez technique; ce qui suit n'en est qu'une esquisse. Soit $I =]\frac{1}{(p-1)\beta}, \frac{p}{(p-1)\beta}[$. On commence par prendre un antécédent de Frobenius \mathcal{D}_h d'ordre h suffisamment grand pour que l'on soit dans les conditions d'utilisation du théorème de Young (prop. 1.2) et que l'on ait $\text{Val}(\mathcal{D}_h, s) = (1 + \beta)s$ si $s \in I$; puis on choisit un vecteur cyclique de telle sorte que \mathcal{D}_h soit associé à un opérateur différentiel $Q(\partial) = \partial^d + a_{d-1}(T)\partial^{d-1} + \dots + a_0(T)$, où les a_i appartiennent au corps des fractions des fonctions holomorphes sur la couronne $v_p(T) \in I$.

Tordre \mathcal{D}_h par $\chi(aX^\beta)$ revient à remplacer $Q(\partial)$ par $Q(\partial - \frac{\pi\beta a}{T^\beta}) = \partial^d + \dots + b_0(T)$. Comme \mathcal{D}_h est purement de ∂ -pente β , on a

$$w_s(a_0(T)) = d\left(\frac{1}{p-1} - \beta s\right) \text{ et } w_s(a_i(T)) \geq (d-i)\left(\frac{1}{p-1} - \beta s\right) \quad \text{si } s \in I,$$

et on montre que l'on peut trouver a dans une extension finie de L de telle sorte qu'il existe $s_0 \in I$ tel que l'on ait $w_{s_0}(b_0(T)) > d(\frac{1}{p-1} - \beta s_0)$. Un petit calcul montre alors que, si

$$Q_1(\partial) = Q(\partial - \pi\partial(P_{\beta,a}(T^{-1}))) = \partial^d + c_{d-1}(T)\partial^{d-1} + \dots + c_0(T)$$

est l'opérateur différentiel correspondant au module tordu $\mathcal{D}_h \otimes \chi(P_{\beta,a})$, alors

(i) $w_s(c_i(T)) \geq (d-i)(\frac{1}{p-1} - \beta s)$ si $0 \leq i \leq d-1$ et $s \in I$;

(ii) il existe $s_0 \in I$ tel que $w_{s_0}(c_0(T)) > d(\frac{1}{p-1} - \beta s)$;

(iii) $w_s(c_{d-1}(T)) = \frac{1}{p-1} - \beta s$ si $(d, p) = 1$ et $s \in I$.

Le (i) assure que $\mathcal{D}_h \otimes \chi(P_{\beta,a})$ est de plus grande ∂ -pente $\leq \beta$ et le (iii) implique que l'on a égalité si $(d, p) = 1$. Le (ii), quant à lui, montre que $\mathcal{D}_h \otimes \chi(P_{\beta,a})$ n'est pas purement de ∂ -pente β (cf. th. 1.2). Il n'y a plus qu'à utiliser le théorème de décomposition et appliquer h fois Frobenius pour revenir au module initial et conclure.

REMARQUE 1.15. — La proposition 1.14 admet (cf. [45, prop. 3.3]) un analogue pour les représentations galoisiennes d'un corps local (à corps résiduel fini).

1.4.3. *Énoncé et démonstration du théorème.* Si \mathcal{D} est un ∂ -module soluble de rang d sur \mathcal{R}_L , nous dirons de façon informelle < si on ne rencontre pas de nombre de Liouville en cours de route > pour signifier < si la catégorie tannakienne engendrée par \mathcal{D} et les $\mathcal{R}_L(E)$, où E parcourt l'ensemble des extensions séparables de $k((T))$, n'admet que des objets dont les exposants et leurs différences sont non Liouville >.

THÉORÈME 1.16 ([60]). — Si \mathcal{D} est un ∂ -module soluble de rang d sur \mathcal{R}_L , et si on ne rencontre pas de nombre de Liouville en cours de route, alors il existe une extension séparable finie E de $k((T))$ et des entiers p -adiques non Liouville $\alpha_1, \dots, \alpha_d$ tels que l'équation différentielle $\partial_{\mathcal{D}}X = 0$ admette d solutions linéairement indépendantes dans $\mathcal{R}_L(E)[T^{\alpha_1}, \dots, T^{\alpha_d}, \log T]$.

REMARQUE 1.17. — On ne rencontre pas de nombre de Liouville en cours de route, si on part d'un (φ, ∂) -module, et le théorème ci-dessus admet comme conséquence la conjecture de Crew. D'un autre côté, bien qu'a priori il faille le faire exprès pour tomber sur un nombre de Liouville, on n'a aucun critère satisfaisant permettant de garantir que cela ne va pas arriver. En particulier, on ne sait pas démontrer que les exposants d'un module différentiel défini sur $\overline{\mathbf{Q}}(T)$ sont non Liouville.

REMARQUE 1.18. — Le théorème ci-dessus donne une description tout à fait satisfaisante des ∂ -modules solubles ; le cas non soluble reste, quant à lui, totalement mystérieux.

Pour démontrer le théorème (sous sa forme filtrée faisant disparaître $\log T$), il suffit de traiter le cas d'un module irréductible.

Si \mathcal{D} est de rang 1, sa ∂ -pente est entière et les propositions 1.13 et 1.14 montrent que l'on peut faire baisser cette ∂ -pente en tordant par un module de rang 1 avec structure de Frobenius. Une récurrence immédiate montre que \mathcal{D} est de la forme $\mathcal{D}_0 \otimes \mathcal{D}_f$, où \mathcal{D}_0 est de ∂ -pente 0 et \mathcal{D}_f admet une structure de Frobenius ; il n'y a plus qu'à utiliser le théorème de Crew (i.e. la conjecture de Crew en rang 1) pour trivialisier \mathcal{D}_f et le théorème 1.9 pour conclure.

Si \mathcal{D} est de rang premier à p , sa ∂ -pente a un dénominateur premier à p et donc devient entière après une extension modérée de $k((T))$. Une récurrence immédiate, utilisant le (ii) de la proposition 1.14 (et le fait que si $d = d_1 + d_2$ est premier à p , alors d_1 ou d_2 est premier à p), permet de montrer que \mathcal{D} acquiert un constituant de rang 1 après une suite d'extensions modérées de $k((T))$.

Si \mathcal{D} est de rang divisible par p , le module $\text{End}^0 \mathcal{D}$ des éléments de trace nulle est de rang premier à p . D'après ce qui précède, après une extension séparable finie E de $k((T))$, le module $\text{End}^0(\mathcal{D})$ acquiert un constituant de rang 1 et de ∂ -pente 0 ; il existe alors $\alpha \in \mathbf{Z}_p$ et un morphisme non trivial de \mathcal{D} dans $\mathcal{D} \otimes T^\alpha$. Quitte à faire une extension modérée, on se retrouve dans l'un des deux cas suivants :

(i) le module $\text{End}^0(\mathcal{D})$ a une section horizontale, et la dimension des endomorphismes de \mathcal{D} commutant à la connexion est ≥ 2 ; \mathcal{D} n'est donc pas irréductible ;

(ii) il existe $\alpha \in \mathbf{Z}_p$ irrationnel et un morphisme non trivial de \mathcal{D} dans $\mathcal{D} \otimes T^\alpha$, et comme \mathcal{D} et $\mathcal{D} \otimes T^\alpha$ ne sont pas isomorphes (considérer les déterminants), cela implique que \mathcal{D} n'est pas irréductible.

En conclusion, après une extension finie séparable, \mathcal{D} acquiert une composante irréductible de rang strictement inférieur à celui de \mathcal{D} . Le théorème s'en déduit par récurrence sur le rang de \mathcal{D} .

2. φ -MODULES

Soit L un corps complet pour une valuation discrète étendant v_p , et dont le corps résiduel est algébriquement clos.

2.1. La stratégie de Kedlaya

Si E est une extension finie galoisienne de $k((T))$ et $G = \text{Gal}(E/k((T)))$, l'anneau $\mathcal{R}_L(E)[\log T]$ est muni d'actions de φ , ∂ , G et N (où $N = \frac{d}{d \log T}$ est la dérivation $\mathcal{R}(E)$ -linéaire normalisée par $N(\log T) = 1$). L'action de G commute aux autres; celle de N commute à ∂ , et on a $N\varphi = p\varphi N$.

Maintenant, si \mathcal{D} est un (φ, ∂) -module sur \mathcal{R}_L , alors (modulo la conjecture de Crew), il existe une extension finie galoisienne de $k((T))$ telle que le L -espace vectoriel $V(\mathcal{D}) = (\mathcal{R}_L(E)[\log T] \otimes_{\mathcal{R}_L} \mathcal{D})^{\partial=0}$ soit de dimension d (cf. n° 0.2.3). Le L -espace vectoriel $V(\mathcal{D})$ est muni d'actions de G , φ et N , et G commute à φ et N , tandis que φ et N vérifient la relation $N\varphi = p\varphi N$.

D'après le théorème de Dieudonné-Manin, $V(\mathcal{D})$ admet une décomposition suivant les pentes de φ (i.e. les valuations de valeurs propres de φ); à cette décomposition est associée une (unique) filtration $0 = V_0 \subset V_1 \subset \dots \subset V_\ell = V(\mathcal{D})$ telle que V_i soit stable par φ , que φ n'ait qu'une seule pente r_i sur V_i/V_{i-1} , et que l'on ait $r_1 < \dots < r_\ell$. La relation $N\varphi = p\varphi N$ montre que $N(V_i) \subset V_{i-1}$ et donc que cette filtration est stable par N , et que N agit trivialement sur le gradué associé.

On en déduit l'existence d'une filtration $0 = \mathcal{D}_0 \subset \mathcal{D}_1 \subset \dots \subset \mathcal{D}_\ell = \mathcal{D}$ de \mathcal{D} par des sous- φ -modules sur \mathcal{R}_L , vérifiant les conditions suivantes :

- (i) il existe un sous- φ -module Δ^i sur \mathcal{E}_L^\dagger de $\mathcal{D}_i/\mathcal{D}_{i-1}$ tel que $\mathcal{D}_i/\mathcal{D}_{i-1} = \mathcal{R}_L \otimes_{\mathcal{E}_L^\dagger} \Delta^i$;
- (ii) il existe un φ -module W_i sur L tel que

$$\mathcal{R}_L(E) \otimes_L W_i = \mathcal{R}_L(E) \otimes_{\mathcal{R}_L} (\mathcal{D}_i/\mathcal{D}_{i-1}) \quad \text{et} \quad \mathcal{E}_L^\dagger(E) \otimes_L W_i = \mathcal{E}_L^\dagger(E) \otimes_{\mathcal{E}_L^\dagger} \Delta^i.$$

(On a $W_i = V_i/V_{i-1}$ et $\Delta^i = \Delta_i/\Delta_{i-1}$ avec $\Delta_i = (\mathcal{E}_L^\dagger(E)[\log T] \otimes_L V_i)^{N=0, G=1}$; pour vérifier que tout marche bien, il faut utiliser le fait que $\mathcal{E}_L^\dagger(E)[\log T]$ est un sous-anneau de $\mathcal{R}_L(E)[\log T]$ stable par ∂ , φ , G et N .)

Maintenant, si on part d'un φ -module sur \mathcal{R}_L (sans ∂ -structure), on peut se demander ce qui reste vrai. Le (ii) est manifestement trop fort (il est plus ou moins équivalent à l'existence d'une ∂ -structure pour laquelle W_i est le L -espace vectoriel des sections horizontales). Ceci amène Kedlaya à introduire deux notions de φ -pentes, les \mathcal{E} -pentes et les \mathcal{R} -pentes (< pentes génériques > et < pentes spéciales > chez Kedlaya); il remplace alors

(ii) par < (ii') les \mathcal{E} -pentes et les \mathcal{R} -pentes de Δ^i sont les mêmes > (c'est trivialement le cas si W_i existe), et montre qu'une φ -module quelconque sur \mathcal{R}_L a une filtration vérifiant (i) et (ii'). D'autre part, il montre que, si on est parti d'un (φ, ∂) -module, alors la filtration est stable par ∂ , ce qui permet de ramener la conjecture de Crew à un cas traité par Tsuzuki [73].

Si on essaie de faire en sens inverse le chemin ci-dessus, on tombe sur un os : on ne sait pas quelle extension galoisienne E de $k((T))$ va marcher. Cela oblige à les considérer toutes et donc à construire un anneau $\mathcal{R}(k((T))^{\text{sep}})$ (cf. n° 2.2); malheureusement cet anneau est beaucoup trop gros et il faut ensuite < décompléter > pour redescendre à une extension finie de $k((T))$. Toutes ces étapes sont assez techniques ...

2.2. Les < foncteurs > \mathcal{E} , \mathcal{E}^\dagger et \mathcal{R}

Soit E_0 un corps de caractéristique p complet pour une valuation discrète v_E de corps résiduel k parfait. Notre but dans ce n° est d'associer⁷ à une extension algébrique E de E_0 ou au complété d'une telle extension, des anneaux $\mathcal{E}(E)$, $\mathcal{E}^\dagger(E)$ et $\mathcal{R}(E)$. Si E est parfait, ces anneaux sont canoniques, mais, dans le cas général, la construction dépend du choix d'un Frobenius sur $\mathcal{E}^\dagger(T)$ et d'un isomorphisme de $k((T))$ sur E_0 ou, ce qui revient au même, d'une uniformisante de E_0 . Si $E_0 = k((T))$, on retombe sur les anneaux $\mathcal{E} = \mathcal{E}(T)$, $\mathcal{E}^\dagger = \mathcal{E}^\dagger(T)$ et $\mathcal{R} = \mathcal{R}(T)$ définis précédemment.

2.2.1. Généralités sur les corps valués complets.

PROPOSITION 2.1. — (i) Si K est un corps complet pour une valuation $v : K \rightarrow \mathbf{R} \cup \{+\infty\}$, et si \overline{K} est une clôture algébrique de K , alors v s'étend de manière unique à \overline{K} .

(ii) Le complété $\widehat{\overline{K}}$ de \overline{K} pour cette valuation est un corps algébriquement clos.

(iii) La clôture séparable K^{sep} de K dans \overline{K} est dense dans $\widehat{\overline{K}}$.

(iv) $\text{Gal}(K^{\text{sep}}/K)$ s'identifie au groupe des automorphismes continus de $\widehat{\overline{K}}$ laissant K fixe et, si H est un sous-groupe de $\text{Gal}(K^{\text{sep}}/K)$, le sous-corps de $\widehat{\overline{K}}$ fixé par H est le complété de la clôture radicielle de $(K^{\text{sep}})^H$.

REMARQUE 2.2. — Le (iii) n'a, bien évidemment, d'intérêt qu'en caractéristique p et le (iv) est le théorème d'Ax-Sen-Tate ([3], [68]); contrairement aux apparences, ce n'est pas une conséquence formelle de la théorie de Galois.

2.2.2. *Le cas E parfait.* Soit $F = W(k)[\frac{1}{p}]$ le corps des fractions de l'anneau des vecteurs de Witt à coefficients dans k , ce qui fait de F un corps complet pour la valuation v_p , d'anneau des entiers $\mathcal{O}_F = W(k)$ et de corps résiduel k_F .

Si E est un corps valué de caractéristique p , on note E^+ l'anneau de ses entiers.

⁷C'est une extension de la théorie des anneaux de Cohen [8] développée dans [52]; les anneaux $\mathcal{E}(E)$, $\mathcal{E}^\vee(E)$ et $\mathcal{R}(E)$ correspondent respectivement aux anneaux $\Gamma^E[\frac{1}{p}]$, $\Gamma_{\text{con}}^E[\frac{1}{p}]$ et $\Gamma_{\text{an,con}}^E$ de Kedlaya.

Soit maintenant $\tilde{\mathbb{E}}$ un corps parfait de caractéristique p muni d'une valuation $v_{\mathbb{E}}$. Soient $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}}) = W(\tilde{\mathbb{E}})$ et $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}}) = W(\tilde{\mathbb{E}}^+)$. Soient $\mathcal{E}(\tilde{\mathbb{E}}) = \mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})[\frac{1}{p}]$ et $\mathcal{E}^+(\tilde{\mathbb{E}}) = \mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})[\frac{1}{p}]$. Les anneaux $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}}) \subset \mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ sont, par construction, séparés et complets pour la topologie p -adique (*topologie forte*) et on a $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})/p\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}}) = \tilde{\mathbb{E}}$ et $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})/p\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}}) = \tilde{\mathbb{E}}^+$. Par ailleurs, $\mathcal{E}(\tilde{\mathbb{E}})$ est un corps complet pour la valuation p -adique (la topologie associée est la *topologie forte*), d'anneau de valuation $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ et de corps résiduel $\tilde{\mathbb{E}}$.

Si $x \in \tilde{\mathbb{E}}$, on note $[x]$ son représentant de Teichmüller dans $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$; rappelons que c'est l'unique élément de $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ ayant x pour réduction modulo p et possédant une racine p^n -ième dans $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ quel que soit $n \in \mathbf{N}$. Tout élément x de $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ (resp. $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})$) s'écrit donc de manière unique sous la forme $\sum_{k=0}^{+\infty} p^k [x_k]$, où $(x_k)_{k \in \mathbf{N}}$ est une suite d'éléments de $\tilde{\mathbb{E}}$ (resp. $\tilde{\mathbb{E}}^+$) et tout élément de $\mathcal{E}(\tilde{\mathbb{E}})$ ou $\mathcal{E}^+(\tilde{\mathbb{E}})$ sous la forme $\sum_{k \gg -\infty}^{+\infty} p^k [x_k]$.

La *topologie faible* sur $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ (resp. $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})$) est, par définition, la topologie qui fait de l'application $x = \sum_{k=0}^{+\infty} p^k [x_k] \rightarrow (x_k)_{k \in \mathbf{N}}$ un homéomorphisme de $\mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ [resp. $\mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})$] sur $\tilde{\mathbb{E}}^{\mathbf{N}}$ [resp. $(\tilde{\mathbb{E}}^+)^{\mathbf{N}}$] muni de la topologie produit ($\tilde{\mathbb{E}}$ et $\tilde{\mathbb{E}}^+$ étant munis de la topologie définie par la valuation $v_{\mathbb{E}}$); si $\bar{\pi} \in \tilde{\mathbb{E}}$ vérifie $v_{\mathbb{E}}(\bar{\pi}) > 0$, la topologie faible est aussi obtenue en prenant les $[\bar{\pi}]^k \mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}}) + p^{n+1} \mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ [resp. les $[\bar{\pi}]^k \mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}}) + p^{n+1} \mathcal{O}_{\mathcal{E}}^+(\tilde{\mathbb{E}})$, pour $k, n \in \mathbf{N}$], comme base de voisinages de 0. On munit $\mathcal{E}(\tilde{\mathbb{E}}) = \bigcup_{n \in \mathbf{N}} p^{-n} \mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ et $\mathcal{E}^+(\tilde{\mathbb{E}})$ de la topologie de la limite inductive.

Si $r > 0$, soit $\mathcal{O}_{\mathcal{E}}^{(0,r]}(\tilde{\mathbb{E}})$ le sous-anneau des $x = \sum_{k=0}^{+\infty} p^k [x_k] \in \mathcal{O}_{\mathcal{E}}(\tilde{\mathbb{E}})$ vérifiant $rv_{\mathbb{E}}(x_k) + k \geq 0$ quel que soit $k \geq 0$ et $\lim_{k \rightarrow +\infty} rv_{\mathbb{E}}(x_k) + k = +\infty$. On munit $\mathcal{O}_{\mathcal{E}}^{(0,r]}(\tilde{\mathbb{E}})$ de la topologie $[\bar{\pi}]$ -adique pour laquelle il est complet (et qui est plus faible que la topologie p -adique). On munit $\mathcal{E}^{(0,r]}(\tilde{\mathbb{E}}) = \mathcal{O}_{\mathcal{E}}^{(0,r]}(\tilde{\mathbb{E}})[\frac{1}{p}]$ et $\mathcal{E}^\dagger(\tilde{\mathbb{E}}) = \bigcup_{r>0} \mathcal{E}^{(0,r]}(\tilde{\mathbb{E}})$ de la topologie de la limite inductive.

Si $s > 0$, soit $w_s : \mathcal{E}^+(\tilde{\mathbb{E}}) \rightarrow \mathbf{R}$ la fonction définie par $w_s(\sum_{k \gg 0} p^k [x_k]) = \inf_{k \in \mathbf{Z}} sv_{\mathbb{E}}(x_k) + k$. Les w_s forment une famille de valuation sur $\mathcal{E}^+(\tilde{\mathbb{E}})$ et on définit $\mathcal{R}^+(\tilde{\mathbb{E}})$ comme le complété de $\mathcal{E}^+(\tilde{\mathbb{E}})$ pour la topologie de Fréchet définie par cette famille (une suite u_n tend vers u dans $\mathcal{R}^+(\tilde{\mathbb{E}})$ si et seulement si, quel que soit $s > 0$, $w_s(u_n - u) \rightarrow +\infty$ quand $n \rightarrow +\infty$).

De même, si $r > s > 0$, soit $w_s : \mathcal{E}^{(0,r]}(\tilde{\mathbb{E}}) \rightarrow \mathbf{R}$ la fonction définie par $w_s(\sum_{k \in \mathbf{Z}} p^k [x_k]) = \inf_{k \in \mathbf{Z}} sv_{\mathbb{E}}(x_k) + k$ et soit $\mathcal{E}^{[0,r]}(\tilde{\mathbb{E}})$ le complété de $\mathcal{E}^{(0,r]}(\tilde{\mathbb{E}})$ pour la topologie de Fréchet définie par cette famille de valuations. Finalement, soit $\mathcal{R}(\tilde{\mathbb{E}}) = \bigcup_{r>0} \mathcal{E}^{[0,r]}(\tilde{\mathbb{E}})$ muni de la topologie de la limite inductive.

L'anneau $\mathcal{E}^\dagger(\tilde{\mathbb{E}})$ est un sous-corps de $\mathcal{E}(\tilde{\mathbb{E}})$; ses éléments sont dits *surconvergents*. Les anneaux $\mathcal{E}^\dagger(\tilde{\mathbb{E}})$ et $\mathcal{R}^+(\tilde{\mathbb{E}})$ sont des sous-anneaux de $\mathcal{R}(\tilde{\mathbb{E}})$, tout élément de $\mathcal{R}(\tilde{\mathbb{E}})$ peut s'écrire comme la somme d'un élément de $\mathcal{E}^\dagger(\tilde{\mathbb{E}})$ et d'un élément de $\mathcal{R}^+(\tilde{\mathbb{E}})$ et $\mathcal{R}^+(\tilde{\mathbb{E}}) \cap \mathcal{E}^\dagger(\tilde{\mathbb{E}}) = \mathcal{E}^+(\tilde{\mathbb{E}})$.

L'action de φ sur $\mathcal{O}_{\mathcal{E}}(\tilde{E})$ s'étend par continuité aux anneaux $\mathcal{E}^{\dagger}(\tilde{E})$, $\mathcal{R}(\tilde{E})$ et $\mathcal{R}^+(\tilde{E})$. L'action de φ est bijective sur $\mathcal{E}^{\dagger}(\tilde{E})$, $\mathcal{R}(\tilde{E})$ et $\mathcal{R}^+(\tilde{E})$ et induit un isomorphisme de $\mathcal{E}^{(0,r]}(\tilde{E})$ sur $\mathcal{E}(\tilde{E})^{(0,\frac{r}{p}]}$ et de $\mathcal{E}^{[0,r]}(\tilde{E})$ sur $\mathcal{E}(\tilde{E})^{[0,\frac{r}{p}]}$.

2.2.3. *Le cas général.* Si E est une extension algébrique de E_0 ou le complété d'une telle extension, notons \hat{E} le complété de sa clôture radicielle.

Choisissons un Frobenius φ sur \mathcal{E}_L^{\dagger} et une uniformisante $\bar{\pi}$ de E_0 . On a alors $\varphi(T) = T^p + p \sum_{k \in \mathbb{Z}} a_k T^k$, où les a_k sont des éléments de \mathcal{O}_L vérifiant une condition de décroissance convenable, et il existe dans $\mathcal{O}_{\mathcal{E}}^{\dagger}(\tilde{E})$ un unique élément π dont la réduction modulo p est $\bar{\pi}$ et tel que l'on ait $\varphi(\pi) = \pi^p + p \sum_{k \in \mathbb{Z}} a_k \pi^k$. On note $\mathcal{E}(E_0)$ l'image de \mathcal{E} dans $\mathcal{E}(\tilde{E})$ par l'application $f \rightarrow f(\pi)$. Par construction, $\mathcal{E}(E_0)$ est un sous-corps de $\mathcal{E}(\tilde{E})$ stable par φ , complet pour la topologie forte, de corps résiduel E_0 .

Maintenant, si E est une extension finie de E_0 , alors $\mathcal{E}(\hat{E})$ contient une unique extension algébrique $\mathcal{E}(E)$ de $\mathcal{E}(E_0)$ dont le corps résiduel est E .

Si E est une extension algébrique de E_0 (resp. le complété d'une extension algébrique de E_0), on note $\mathcal{E}(E)$ l'adhérence dans $\mathcal{E}(\tilde{E})$ de $\bigcup_{E' \subset E} \mathcal{E}(E')$ pour la topologie forte (resp. faible), où E' parcourt les extensions finies de E_0 contenues dans E . On note $\mathcal{O}_{\mathcal{E}}(E) = \mathcal{O}_{\mathcal{E}}(\tilde{E}) \cap \mathcal{E}(E)$ l'anneau des entiers de $\mathcal{E}(E)$ et on a, dans tous les cas, $\mathcal{O}_{\mathcal{E}}(E)/p\mathcal{O}_{\mathcal{E}}(E) = E$.

Soit $\mathcal{E}^{\dagger}(E) = \mathcal{E}^{\dagger}(\tilde{E}) \cap \mathcal{E}(E)$ le sous-corps des éléments surconvergents de $\mathcal{E}(E)$. Plus généralement, si $r > 0$, soit $\mathcal{E}^{(0,r]}(E) = \mathcal{E}^{(0,r]}(\tilde{E}) \cap \mathcal{E}(E)$. On note $\mathcal{E}^{[0,r]}(E)$ l'adhérence de $\mathcal{E}^{(0,r]}(E)$ dans $\mathcal{E}^{[0,r]}(\tilde{E})$ et $\mathcal{R}(E)$ l'adhérence de $\mathcal{E}^{\dagger}(E)$ dans $\mathcal{R}(\tilde{E})$. On a $\mathcal{R}(\hat{E}) = \mathcal{R}(E)$, si \hat{E} est le complété de E pour la valuation v_E .

Si H est un sous-groupe fermé du groupe $\text{Aut}_{E_0} E$ des automorphismes continus de E laissant E_0 fixe, et X est l'un des foncteurs \mathcal{E} , \mathcal{E}^{\dagger} , $\mathcal{E}^{[0,r]}$... ci-dessus, alors H opère continûment sur $X(E)$, mais il faut faire un peu attention quand on prend les points fixes sous l'action de H . Par exemple, on a $\mathcal{E}(E)^H = \mathcal{E}(E^H)$ et $\mathcal{E}^{\dagger}(E)^H = \mathcal{E}^{\dagger}(E^H)$, mais $\mathcal{R}(E)^H = \mathcal{R}(\hat{E})^H = \mathcal{R}(\hat{E}^H)$ n'est pas, en général, égal à $\mathcal{R}(E^H)$.

Plus généralement, si L est une extension finie totalement ramifiée de F , munie d'une extension de φ , on peut tensoriser tous les anneaux précédents par \mathcal{O}_L au-dessus de $\mathcal{O}_F = W(k)$; on dénote cette opération en rajoutant un L en indice.

Si E est une extension finie E_0 de corps résiduel k' , soient $E'_0 = k' \cdot E_0$ et $L' = W(k') \otimes_{W(k)} L$. Si $\bar{\pi}_E$ est une uniformisante de E , si $\bar{P}(X) = X^d + \bar{a}_{d-1}X^{d-1} + \dots + \bar{a}_0 \in (E'_0)^+[X]$ est le polynôme minimal de $\bar{\pi}_E$ et si $a_i \in \mathcal{O}_{\mathcal{E},L}^{\dagger}(E'_0)$ est un relèvement surconvergent de \bar{a}_i , alors $P = X^d + a_{d-1}X^{d-1} + \dots + a_0$ (resp. $\varphi(P)$) a une unique racine π_E (resp. $\varphi(\pi_E)$) dans $\mathcal{O}_{\mathcal{E}}(\tilde{E})$ dont l'image dans E est $\bar{\pi}_E$ (resp. $\bar{\pi}_E^p$) et on a $\pi_E \in \mathcal{E}^{\dagger}(E)$. (Lemme de Hensel si E/E_0 est séparable; le cas général s'en déduit en appliquant φ le nombre de fois qu'il faut.) Ceci permet de montrer que φ s'étend de manière unique à tous les anneaux construits ci-dessus; son action commute à celle de $\text{Aut}_{E_0} E$.

La dérivation ∂ s'étend de manière unique à $\mathcal{E}_L^\dagger(E)$ et $\mathcal{R}_L(E)$; on a par exemple

$$\partial\pi_E = -\frac{\partial a_{d-1}\pi_E^{d-1} + \cdots + \partial a_0}{d\pi_E^{d-1} + (d-1)a_{d-1} + \cdots + a_1}.$$

Elle s'étend par continuité à $\mathcal{E}_L(E)$ si E est une extension *séparable* quelconque de E_0 . Par contre, *elle ne s'étend pas par continuité à $\mathcal{E}_L^\dagger(E)$ ou $\mathcal{R}_L(E)$ si E est une extension infinie < trop ramifiée > de E_0 .*

PROPOSITION 2.3. — *L'application $f(T) \rightarrow f(\pi_E)$ induit un isomorphisme de*

- (i) $\mathcal{E}_{L'}(T)$ sur $\mathcal{E}_L(E)$;
- (ii) $\mathcal{E}_{L'}^\dagger(T)$ sur $\mathcal{E}_L^\dagger(E)$;
- (iii) $\mathcal{R}_{L'}(T)$ sur $\mathcal{R}_L(E)$.

D'autre part, on a le résultat suivant généralisant le théorème de Lazard.

PROPOSITION 2.4 ([52]). — *Si E est une extension algébrique de E_0 ou le complété d'une telle extension, et si L est une extension finie de F , alors $\mathcal{R}_L(E)$ est un anneau de Bézout.*

2.3. φ -Modules sur l'anneau de Robba

Si A est un anneau muni d'un endomorphisme φ , un φ -module sur A est un A -module libre de rang fini muni d'une action semi-linéaire de φ telle que la matrice de φ dans une base soit inversible.

2.3.1. *Le théorème de Dieudonné-Manin.* Soit κ un corps algébriquement clos de caractéristique p et M une extension finie du corps des fractions $W(\kappa)[\frac{1}{p}]$ de l'anneau des vecteurs de Witt à coefficients dans κ . On suppose que l'action de φ sur $W(\kappa)$ s'étend à M . Le sous-corps de M fixe par φ est une extension finie totalement ramifiée de \mathbf{Q}_p , ayant même groupe de valuation que M ; il contient donc une uniformisante π de M .

Soit D un φ -module de rang d sur M . Soit v un vecteur cyclique (tel que $v, \varphi(v), \dots, \varphi^{d-1}(v)$ forment une base de D sur M) et soit $P(X) = X^d + a_{d-1}X^{d-1} + \cdots + a_0$ le polynôme défini par $\varphi^d(v) + a_{d-1}\varphi^{d-1}(v) + \cdots + a_0v = 0$. Soient r_1, \dots, r_d les valuations des racines de p dans une extension de M (ces valuations se lisent directement sur le polygone de Newton de P).

PROPOSITION 2.5. — *Les r_i ne dépendent, à l'ordre près, que de D et pas du choix de v et si, pour $1 \leq i \leq d$, il existe $\alpha_i \in M$ avec $v_p(\alpha_i) = r_i$, alors D admet une base (e_1, \dots, e_d) telle que l'on ait $\varphi(e_i) = \alpha_i e_i$ pour $1 \leq i \leq d$.*

La proposition ci-dessus est une version du théorème de Dieudonné-Manin. Remarquons que l'on peut assurer l'existence des α_i en adjoignant à M une racine d'ordre convenable de π . Les r_i s'appellent les φ -pentes de D .

Notons, pour le reste de cette partie, \tilde{E} le complété de la clôture algébrique de $k((T))$, où k est algébriquement clos de caractéristique p . On note $\tilde{\mathcal{R}}$, $\tilde{\mathcal{E}}^\dagger$ et $\tilde{\mathcal{E}}$ respectivement les anneaux $\mathcal{R}_L(\tilde{E})$, $\mathcal{E}_L^\dagger(\tilde{E})$ et $\mathcal{E}_L(\tilde{E})$.

La proposition 2.5 ci-dessus s'applique en particulier à un φ -module sur $\tilde{\mathcal{E}}$. Si $E \subset \tilde{E}$ et si D est un φ -module sur $\mathcal{E}^\dagger(E)$, on appelle \mathcal{E} -*pent*es de D les φ -pentes de $\tilde{\mathcal{E}} \otimes_{\mathcal{E}_L^\dagger(E)} D$. On dit que D est *isocline* s'il n'a qu'une \mathcal{E} -pente et *étale* si cette pente est nulle.

2.3.2. *Un analogue du théorème de Dieudonné-Manin sur l'anneau de Robba.*

THÉORÈME 2.6 ([52]). — *Si D est un φ -module de rang d sur $\tilde{\mathcal{R}}$, alors D possède une base de vecteurs propres sur $\mathcal{R}_{L'}(\tilde{E})$, où L' est une extension finie de L ; de plus, les valuations des valeurs propres correspondantes ne dépendent, à l'ordre près, que de D ; ces valuations s'appellent les \mathcal{R} -pentes de D .*

REMARQUE 2.7. — En rang 1, un φ -module sur $\tilde{\mathcal{R}}$ est défini sur $\tilde{\mathcal{E}}^\dagger$, et les \mathcal{E} -pente et \mathcal{R} -pente coïncident.

La démonstration de ce théorème s'apparente à un numéro de funambulisme. On commence par montrer que, si $v_p(\lambda)$ est assez grand, il existe $v \in D$ vérifiant $\varphi(v) = \lambda v$ (c'est loin d'être trivial). On utilise alors le théorème de Lazard généralisé (prop. 2.4) pour montrer que ce vecteur est multiple d'un vecteur propre w [pour une valeur propre de valuation $\leq v_p(\lambda)$] primitif (i.e. que l'on peut compléter en une base de D sur $\tilde{\mathcal{R}}$). Appliquant ceci au module $D/\langle w \rangle$, on fabrique une base de D dans laquelle la matrice de φ est triangulaire. La somme des valuations des termes diagonaux est égale à la \mathcal{E} -pente de $\det D$; en particulier, elle ne dépend pas de la base choisie.

Passer d'une matrice triangulaire à une matrice diagonale n'est pas une mince affaire. Par exemple, en dimension 2, si on part d'une base v_1, v_2 dans laquelle l'action de φ est donnée par $\varphi(v_1) = \lambda_1 v_1$ et $\varphi(v_2) = \lambda_2 v_2 + a v_1$, il y a deux cas suivant que $v_L(\lambda_2) - v_L(\lambda_1) \geq 0$ ou $v_L(\lambda_2) - v_L(\lambda_1) < 0$ (où v_L désigne la valuation normalisée de L). Dans le premier cas, l'équation $\varphi(b) - \frac{\lambda_2}{\lambda_1} b = \frac{a}{\lambda_1}$ a une solution dans $\tilde{\mathcal{R}}$ et $(v_1, v_2 + b v_1)$ est une base constituée de vecteurs propres; dans le second cas, l'équation précédente n'a, en général, pas de solution et il faut procéder autrement. Kedlaya montre (et c'est là le point le plus délicat) que, si $v_L(\lambda_2) - v_L(\lambda_1) \leq -2$, alors D contient un vecteur propre primitif pour une valeur propre λ'_1 avec $v_L(\lambda'_1) < v_L(\lambda_1)$. Si $v_L(\lambda_2) - v_L(\lambda_1)$ est pair (et on peut toujours se ramener à ce cas en remplaçant L par une extension quadratique), après un nombre fini d'étapes, on se ramène au cas $v_L(\lambda_2) - v_L(\lambda_1) \geq 0$.

Le cas général ne peut pas se ramener au cas de la dimension 2 car la construction précédente ne fera jamais apparaître que des pentes dont le dénominateur est une puissance de 2, et la démonstration dans le cas général est franchement technique.

2.3.3. *Filtration par les pentes.*

THÉORÈME 2.8 ([52]). — Si D est un φ -module sur \mathcal{R}_L , alors D admet une unique filtration $0 = D_0 \subset D_1 \subset \cdots \subset D_\ell = D$ par des sous- φ -modules sur \mathcal{R}_L vérifiant les conditions suivantes :

- (i) D_i/D_{i-1} est un φ -module sur \mathcal{R}_L n'ayant qu'une seule \mathcal{R} -pente s_i ;
- (ii) $s_1 < s_2 < \cdots < s_\ell$;
- (iii) D_i/D_{i-1} contient un sous- φ -module Δ_i sur \mathcal{E}^\dagger isocline de \mathcal{E} -pente s_i tel que l'on ait $D_i/D_{i-1} = \mathcal{R}_L \otimes_{\mathcal{E}_L^\dagger} \Delta_i$.

La démonstration de ce théorème est presque aussi acrobatique que celle du précédent. Soient e_1, \dots, e_d une base de D sur \mathcal{R}_L et A la matrice de φ dans cette base. D'après le théorème 2.6, il existe une matrice $M \in GL_d(\tilde{\mathcal{R}})$ telle que $M^{-1}A\varphi(M)$ soit diagonale. Approximant M par une matrice à coefficients dans $\mathcal{R}_L(E)$, où E est une extension finie de $k((T))$, on construit un sous- φ -module Δ sur $\mathcal{E}_L^\dagger(E)$ tel que $\mathcal{R}_L(E) \otimes_{\mathcal{E}_L^\dagger(E)} \Delta = \mathcal{R}_L(E) \otimes_{\mathcal{R}_L} D$ et ayant les mêmes ensembles de \mathcal{R} -pentés et \mathcal{E} -pentés. C'est cette dernière condition qui demande le plus de travail ; en effet, si D est un φ -module de rang d sur \mathcal{E}_L^\dagger , et si $r_1 \leq \cdots \leq r_d$ (resp. $s_1 \geq \cdots \geq s_d$) sont les \mathcal{E} -pentés (resp. les \mathcal{R} -pentés) de D , les seules relations que l'on ait en général entre ces ensembles sont l'égalité $s_1 + \cdots + s_d = r_1 + \cdots + r_d$ et les inégalité $s_1 + \cdots + s_i \geq r_1 + \cdots + r_i$, si $i < d$. Cette condition permet de montrer que la filtration croissante de Δ par les \mathcal{E} -pentés, qui n'est a priori définie que sur $\mathcal{E}_L(E) \otimes_{\mathcal{E}_L^\dagger(E)} \Delta$, se descend en une filtration sur Δ . Pour terminer, on étend cette filtration à $\mathcal{R}_L(E) \otimes_{\mathcal{R}_L} D$ et un peu de descente galoisienne permet de conclure.

2.3.4. *Application aux (φ, ∂) -modules.* Soit ∂ une dérivation sur \mathcal{E}_L^\dagger . Tsuzuki [73] (voir aussi [14]) a démontré le résultat suivant.

THÉORÈME 2.9 ([73]). — Si D est un (φ, ∂) -module isocline sur \mathcal{E}_L^\dagger , il existe une extension finie séparable E de $k((T))$ telle que $\mathcal{E}_L^\dagger(E) \otimes_{\mathcal{E}_L^\dagger} D$ possède une base e_1, \dots, e_d sur $\mathcal{E}_L^\dagger(E)$ vérifiant $\varphi(e_i) = e_i$ et $\partial e_i = 0$ si $1 \leq i \leq d$.

Les deux conditions $\varphi(x) = x$ et $\partial x = 0$ ne sont pas loin d'être équivalentes : les L -espaces vectoriels engendrés par les solutions sont les mêmes dans les deux cas. Notons que, en rang 1, cet énoncé est équivalent à la conjecture de Crew car un élément inversible de \mathcal{R}_L appartient à \mathcal{E}_L^\dagger et que, par force, un φ -module de rang 1 est isocline. Ceci nous fournit donc une autre démonstration du théorème de Crew.

Le théorème 2.8 permet de déduire la conjecture de Crew de cet énoncé. En effet, une récurrence immédiate montre qu'il suffit de prouver que le \mathcal{E}^\dagger -module isocline D_1 est stable par ∂ , ce qui suit, avec quelque effort, de ce que ∂ diminue les φ -pentés.

2.3.5. *Compléments.* Le résultat de Tsuzuki peut se réinterpréter en termes galoisiens en anticipant un peu sur les résultats des §§ suivants (cf. remarque 5.2). La dérivation ∂ s'étend par continuité à \mathcal{E} et à $\mathcal{E}(k((T))^{\text{sep}})$. D'autre part, les foncteurs

$$V \rightarrow D(V) = \left(\mathcal{E}(k((T))^{\text{sep}}) \otimes_{\mathbf{Q}_p} V \right)^{\mathcal{G}_{k((T))}} \quad \text{et} \quad D \rightarrow V(D) = \left(\mathcal{E}(k((T))^{\text{sep}}) \otimes_{\mathcal{E}} D \right)^{\varphi=1}$$

sont inverses l'un de l'autre, et établissent une équivalence de catégories entre la catégorie des \mathbf{Q}_p -représentations de $\mathcal{G}_{k((T))}$ et celle des φ -modules étales sur \mathcal{E} . On en déduit l'existence sur tout φ -module étale D sur \mathcal{E} d'une unique connexion ∂_D au-dessus de ∂ vérifiant la relation $\partial_D \circ \varphi = \frac{\partial(\varphi(T))}{\varphi(\partial(T))} \varphi \circ \partial_D$: en effet, une telle connexion s'étend de manière unique à $\mathcal{E}(k((T))^{\text{sep}}) \otimes_{\mathcal{E}} D = \mathcal{E}(k((T))^{\text{sep}}) \otimes_{\mathbf{Q}_p} V(D)$ et le fait que φ est étale et vérifie la relation de commutation ci-dessus implique que ∂_D est nul sur $V(D)$. En résumé, *tout φ -module étale D sur \mathcal{E} peut être muni d'une unique structure de (φ, ∂) -module.*

COROLLAIRE 2.10 ([73]). — *Si D est un (φ, ∂) -module étale sur \mathcal{E} les conditions suivantes sont équivalentes :*

- (i) ∂ est surconvergente (il existe une base de D sur \mathcal{E} dans laquelle la matrice de ∂ est à coefficients dans \mathcal{E}^\dagger);
- (ii) (l'inertie de) $\mathcal{G}_{k((T))}$ agit à travers un quotient fini sur $V(D)$.

3. REPRÉSENTATIONS p -ADIQUES ET COHOMOLOGIE GALOISIENNE

3.1. Généralités

Dans toute cette partie, G est un groupe profini; en particulier, G est compact. Un G -anneau est un anneau topologique muni d'une action continue de G respectant sa structure d'anneau (on demande que $(\sigma, x) \rightarrow \sigma(x)$ soit continue).

Si B est un G -anneau, une B -représentation de G est un B -module libre W de rang fini muni d'une action semi-linéaire continue de G .

Si on choisit une base e_1, \dots, e_d de W sur B et que l'on note $U_\sigma = (u_{i,j}^\sigma)$ la matrice des vecteurs $\sigma(e_1), \dots, \sigma(e_d)$ dans la base e_1, \dots, e_d (i.e. $\sigma(e_j) = \sum_{i=1}^d u_{i,j}^\sigma e_i$), la semi-linéarité de l'action de G se traduit par la relation de cocycle $U_{\sigma\tau} = U_\sigma \sigma(U_\tau)$ quels que soient $\sigma, \tau \in G$. En particulier, en prenant $\tau = \sigma^{-1}$, on en déduit le fait que U_σ est inversible et que $\sigma \rightarrow U_\sigma$ est un cocycle continu à valeurs dans $GL_d(B)$. D'autre part, si on choisit une autre base f_1, \dots, f_d de W sur B et que l'on note $M = (m_{i,j})$ la matrice de passage et U'_σ la matrice de $\sigma(f_1), \dots, \sigma(f_d)$ dans la base f_1, \dots, f_d , on a $U'_\sigma = M^{-1} U_\sigma \sigma(M)$, ce qui montre que les cocycles associés à deux bases différentes sont cohomologues. Ceci permet d'associer à toute B -représentation de G de rang d un élément de l'ensemble de cohomologie continue $H^1(G, GL_d(B))$.

Si V est une \mathbf{Q}_p -représentation de G , alors $B \otimes_{\mathbf{Q}_p} V$, muni de l'action diagonale de G , est une B -représentation de G ; on dit que V est B -admissible si la classe de cohomologie de $B \otimes_{\mathbf{Q}_p} V$ dans $H^1(G, GL_d(B))$ qui lui est associée est triviale ou, autrement dit, si $B \otimes_{\mathbf{Q}_p} V \cong B^d$ en tant que G -module. On remarquera que le groupe de cohomologie continue $H^1(G, GL_d(B))$ dépend fortement de la topologie que l'on a mise sur B mais que la B -admissibilité d'une \mathbf{Q}_p -représentation n'en dépend pas (tant que l'injection de \mathbf{Q}_p dans B est continue).

REMARQUE 3.1. — Si V est une \mathbf{Q}_p -représentation B -admissible, $\mathbf{D}_B(V) = (B \otimes_{\mathbf{Q}_p} V)^G$ est un B^G -module libre de rang d et l'application naturelle $\alpha_B : B \otimes_{B^G} \mathbf{D}_B(V) \rightarrow B \otimes_{\mathbf{Q}_p} V$ est un isomorphisme commutant à l'action de G . En particulier, si $B^{\text{structure}} = \mathbf{Q}_p$, on peut retrouver V en tant que \mathbf{Q}_p -représentation de G à partir de $\mathbf{D}_B(V)$ en prenant les éléments de $B \otimes_{B^G} \mathbf{D}_B(V)$ fixés par les structures.

3.2. Le théorème de Hilbert 90 et ses variantes

La discussion précédente montre que l'on a intérêt à étudier les ensembles $H^1(G, GL_d(B))$. En particulier, si cet ensemble est trivial, toutes les \mathbf{Q}_p -représentations de G sont B -admissibles et le module $\mathbf{D}_B(V)$ est un invariant de V pour tout V .

Pour ce faire, on dispose de deux outils classiques. D'une part la suite exacte \langle d'inflation-restriction \rangle : si Λ est un G -anneau et si H est un sous-groupe fermé distingué de G , alors la suite d'ensembles pointés

$$1 \longrightarrow H^1(G/H, GL_d(\Lambda^H)) \xrightarrow{\text{inf}} H^1(G, GL_d(\Lambda)) \xrightarrow{\text{res}} H^1(H, GL_d(\Lambda))$$

est exacte. D'autre part, du théorème de Hilbert 90 (et des techniques entrant dans sa démonstration : \langle séries de Poincaré \rangle , cf. [67] par exemple).

PROPOSITION 3.2 (Hilbert 90). — *Si L/F est une extension galoisienne de groupe de Galois G et si on munit L de la topologie discrète, alors*

- (i) $H^1(G, GL_d(L)) = \{1\}$ si $d \geq 1$;
- (ii) $H^1(G, L) = \{0\}$.

Le résultat suivant se déduit du théorème de Hilbert 90 par approximations successives.

PROPOSITION 3.3. — *Soit Λ un G -anneau et soit $\pi \in \Lambda$ tel que l'idéal engendré par π soit stable par G et Λ soit séparé et complet pour la topologie π -adique supposée plus forte que la topologie de Λ . Si $H^1(G, GL_d(\Lambda/\pi\Lambda)) = \{1\}$ quel que soit $d \geq 1$ ($\Lambda/\pi\Lambda$ étant muni de la topologie quotient), alors*

- (i) $H^1(G, GL_d(\Lambda)) = \{1\}$ si $d \geq 1$;
- (ii) $H^1(G, \Lambda) = \{0\}$.

PROPOSITION 3.4. — *Soient Λ un G -anneau principal et $\omega \in \Lambda$ tel que l'idéal (ω) soit stable par G . Si $\Lambda' = \Lambda[\frac{1}{\omega}] = \bigcup_{n \in \mathbf{N}} \omega^{-n} \Lambda$ est muni de la topologie de la limite inductive alors l'application naturelle $H^1(G, GL_d(\Lambda)) \rightarrow H^1(G, GL_d(\Lambda'))$ est surjective.*

Démonstration. — À un cocycle à valeurs dans $GL_d(\Lambda')$ correspond une Λ' -représentation de G et la compacité de G implique l'existence d'un sous- Λ -réseau stable.

3.3. La méthode de Sen

La méthode de Sen [66] permet, sous certaines *conditions de Tate-Sen*, de réduire beaucoup la complexité apparente des ensembles de cohomologie que l'on considère. On part des données suivantes :

— H est un sous-groupe fermé distingué de G tel que $\Gamma = G/H$ s'identifie à un sous-groupe de \mathbf{Z}_p^* (plus généralement, un groupe du type $\mathbf{Z}_p \times U$, où U est un groupe fini ferait l'affaire) ;

— $\tilde{\Lambda} \ni \pi$ est un anneau intègre, séparé et complet pour la topologie π -adique, muni d'une action continue de G laissant stable l'idéal (π) ;

— $(\Lambda_n)_{n \in \mathbf{N}}$ est une suite croissante de sous-anneaux fermés de $\tilde{\Lambda}^H$ contenant π et stables par Γ .

On suppose que $(\tilde{\Lambda}, (\Lambda_n)_{n \in \mathbf{N}})$ satisfait aux conditions :

(TS1) il existe $c_1 \in \mathbf{N}$ tel que, quels que soient les sous-groupes ouverts $H_1 \subset H_2$ de H , il existe $\alpha \in (\pi^{-c_1} \tilde{\Lambda})^{H_1}$ vérifiant $\sum_{\tau \in H_2/H_1} \tau(\alpha) = 1$;

(TS2) Il existe une suite d'applications $R_n : \tilde{\Lambda}^H \rightarrow \Lambda_n[\frac{1}{\pi}]$ vérifiant :

a) R_n est Λ_n -linéaire et $R_n(x) = x$ si $x \in \Lambda_n$;

b) si $x \in \tilde{\Lambda}^H$, alors $\lim_{n \rightarrow +\infty} R_n(x) = x$;

c) il existe $c_2 \in \mathbf{N}$ tel que, si $n \in \mathbf{N}$ et $x \in \pi^k \tilde{\Lambda}^H$, alors $R_n(x) \in \pi^{k-c_2} \Lambda_n$;

d) R_n commute à l'action de Γ .

(TS3) Il existe $c_3 \in \mathbf{N}$ tel que, si $n \in \mathbf{N}$, si $X_n = (1 - R_n)(\tilde{\Lambda}^H)$, si γ_n est un générateur de $\Gamma_n = \Gamma \cap (1 + p^n \mathbf{Z}_p)$, et si $x \in \pi^k X_n$, alors il existe $y \in \pi^{k-c_3} X_n$ tel que $x = (\gamma_n - 1)y$.

PROPOSITION 3.5 ([66]). — (i) Si $\tilde{\Lambda}$ vérifie la condition (TS1), et si $\sigma \rightarrow U_\sigma$ est un cocycle continu sur H à valeurs dans $GL_d(\tilde{\Lambda})$, il existe un sous-groupe ouvert H' de H tel que la restriction de $\sigma \rightarrow U_\sigma$ à H' soit cohomologue au cocycle trivial.

(ii) Si $(\tilde{\Lambda}, (\Lambda_n)_{n \in \mathbf{N}})$ vérifie les conditions (TS2) et (TS3), l'application naturelle de $\varinjlim H^1(\Gamma, GL_d(\Lambda_n))$ dans $H^1(\Gamma, GL_d(\tilde{\Lambda}^H))$ est un isomorphisme.

Remarquons que le (i), couplé avec le théorème de Hilbert 90, permet en général de démontrer que $H^1(H, GL_d(\text{Frac}(\tilde{\Lambda})))$ est trivial. La proposition 3.4 et la suite d'inflation-restriction impliquent alors que tout cocycle continu sur G à valeurs dans $GL_d(\tilde{\Lambda})$ provient d'un cocycle sur Γ à valeurs dans $GL_d(\Lambda_n)$ pour un certain n .

Les propriétés (TS1) et (TS3) sont les plus délicates à vérifier. Dans les situations arithmétiques, (TS1) se démontre en général via la théorie de la ramification (théorie des extensions \langle presque étales \rangle dans la terminologie de Faltings) et pour (TS3), il n'y a pas de recette miracle.

4. LES ANNEAUX DE FONTAINE

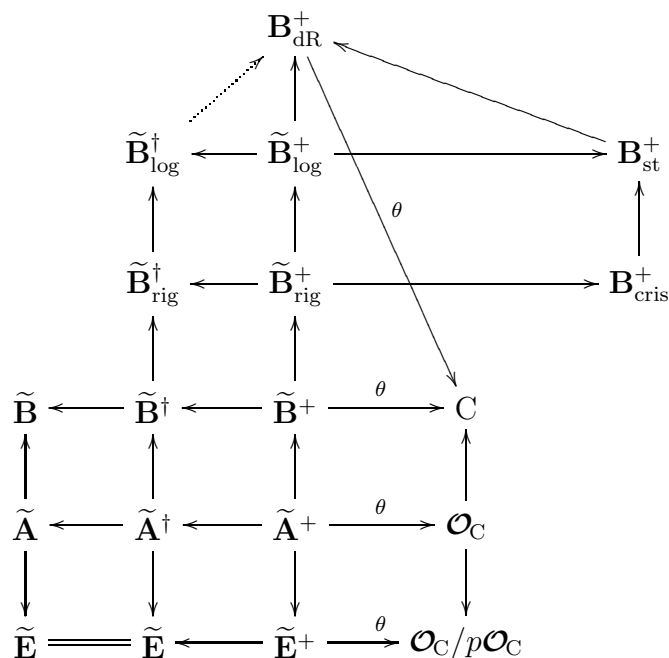
4.1. Cartographie

4.1.1. *Notations.* Soit k_F un corps parfait de caractéristique p , $\mathcal{O}_F = W(F)$ l'anneau des vecteurs de Witt à coefficients dans k_F et $F = \mathcal{O}_F[\frac{1}{p}]$ ce qui fait de F un corps complet pour la valuation p -adique v_p , de corps résiduel k_F .

Soit \overline{F} une clôture algébrique de F ; la valuation v_p s'étend de manière unique à \overline{F} . Choisissons un système $\varepsilon = (1, \varepsilon^{(1)}, \dots, \varepsilon^{(n)}, \dots)$, où les $\varepsilon^{(n)} \in \overline{F}$ vérifient $\varepsilon^{(1)} \neq 1$ et $(\varepsilon^{(n+1)})^p = \varepsilon^{(n)}$; en particulier, $\varepsilon^{(n)}$ est une racine primitive p^n -ième de l'unité.

Si $K \subset \overline{F}$ est une extension finie totalement ramifiée de F , soient $K_n = K(\varepsilon^{(n)})$ si $n \in \mathbf{N}$ et $K_\infty = \cup_{n \in \mathbf{N}} K_n$ l'extension cyclotomique de K ; on note F' l'extension maximale non ramifiée de F contenue dans K_∞ . Soient $\mathcal{G}_K = \text{Gal}(\overline{F}/K)$, $\mathcal{H}_K = \text{Gal}(\overline{F}/K_\infty)$ et $\Gamma_K = \mathcal{G}_K/\mathcal{H}_K = \text{Gal}(K_\infty/K)$. Soit aussi $\chi : \mathcal{G}_K \rightarrow \mathbf{Z}_p^*$ le caractère cyclotomique; il induit un isomorphisme de Γ_K sur un sous-groupe ouvert de \mathbf{Z}_p^* .

4.1.2. *Navigation dans le monde des anneaux.* La plupart des anneaux construits par Fontaine s'obtiennent à partir d'un anneau $\tilde{\mathbf{A}}^+$ en localisant et en complétant. Comme l'anneau $\tilde{\mathbf{A}}^+$ est de dimension 2, les anneaux que l'on obtient de cette manière forment naturellement un tableau en deux dimensions (et même en quatre si on rajoute les actions de Frobenius et Galois); il est donc quasi-impossible de les présenter de manière satisfaisante dans un texte qui, par nature, est de dimension 1. Le lecteur aura intérêt à se reporter au tableau ci-dessous pour s'orienter.



(i) Plus on monte et plus p est inversible : la ligne du bas vit en caractéristique p ; c'est la réduction modulo p de la ligne du dessus et on passe de l'avant-dernière ligne à la

précédente en rendant p -inversible. Plus on va vers la gauche (en ne tenant pas compte de la dernière colonne qui n'est là que pour mémoire), et plus $\pi = [\varepsilon] - 1$ est inversible : π est nul dans la quatrième colonne, et l'application $\theta : \mathbf{B}_{\text{dR}}^+ \rightarrow \mathbf{C}$ est la réduction modulo π .

(ii) Toutes les flèches de ce diagramme sont injectives à l'exception de la réduction modulo p et de θ .

(iii) \mathcal{G}_F opère continûment sur tout le diagramme.

(iv) φ opère sur les trois premières colonnes de manière bijective (et sur la dernière, mais pas bijectivement), et son action commute à celle de \mathcal{G}_F .

(v) Les anneaux des trois premières colonnes (sauf $\tilde{\mathbf{B}}_{\text{log}}^+$) ont un petit frère sans \sim sur lequel l'action de φ n'est qu'injective. On passe d'un anneau sans \sim à un anneau avec en rendant φ inversible et en complétant. Par exemple, $\tilde{\mathbf{E}}$ est le complété de la clôture radicielle de \mathbf{E} .

(vi) La flèche en pointillés de $\tilde{\mathbf{B}}_{\text{log}}^+$ vers \mathbf{B}_{dR}^+ est une famille de morphisme $(\iota_n)_{n \in \mathbf{N}}$, où ι_n est défini sur le sous-anneau $\tilde{\mathbf{B}}^{[0, p^{-n}]}[\log[\tilde{\rho}]]$ de $\tilde{\mathbf{B}}_{\text{log}}^+$ des éléments \llcorner convergeant sur la couronne $0 < v_p(\Gamma) \leq p^{-n} \gg$; la réunion de ces anneaux est $\tilde{\mathbf{B}}_{\text{log}}^+$ et on a $\iota_n(x) = \iota_{n-1}(\varphi^{-1}(x))$ si tout est défini.

(vii) La théorie des (φ, Γ) -modules utilise les anneaux \mathbf{E} , \mathbf{A} et \mathbf{B} de la première colonne, alors que la théorie de Hodge p -adique utilise les anneaux $\tilde{\mathbf{B}}_{\text{rig}}^+$, $\tilde{\mathbf{B}}_{\text{log}}^+$, \mathbf{B}_{dR}^+ et \mathbf{C} . Le lecteur remarquera que la plupart des flèches allant de la théorie des (φ, Γ) -modules à la théorie de Hodge p -adique sont dans le mauvais sens; il faut donc à chaque fois démontrer que l'on peut effectivement remonter.

(viii) Si K est une extension finie de F , un K en indice indique le sous-anneau fixé par \mathcal{H}_K . Les anneaux sans \sim avec un K en indice sont des anneaux de séries en une variable π_K . Dans le cas $K = F$, on a $\pi_F = \pi = [\varepsilon] - 1$; les actions de \mathcal{G}_F et φ sont données par les formules $g(\pi) = (1 + \pi)^{\chi(g)} - 1$ et $\varphi(\pi) = (1 + \pi)^p - 1$, et on obtient le diagramme suivant :

$$\begin{array}{ccccc}
 & & \mathcal{R}_F[\log \pi] & \cdots \cdots \cdots & \varinjlim F_n((\pi)) \\
 & & \uparrow & & \downarrow \\
 & & \mathcal{R}_F & \longleftarrow & \mathcal{R}_F^+ \\
 & & \uparrow & & \uparrow \\
 \mathcal{E}_F & \longleftarrow & \mathcal{E}_F^+ & \longleftarrow & \mathcal{O}_F[[\pi]]\left[\frac{1}{p}\right] \\
 \uparrow & & \uparrow & & \uparrow \\
 \mathcal{O}_{\varepsilon, F} & \longleftarrow & \mathcal{O}_{\varepsilon, F}^+ & \longleftarrow & \mathcal{O}_F[[\pi]] \\
 \downarrow & & \downarrow & & \downarrow \\
 k_F((\bar{\pi})) & \equiv & k_F((\bar{\pi})) & \longleftarrow & k_F[[\bar{\pi}]]
 \end{array}$$

(ix) $[\varepsilon] = 1 + \pi$ est un analogue p -adique de $e^{2i\pi}$, et l'application $\mu \rightarrow \int_{\mathbf{Z}_p} (1 + \pi)^x \mu(x)$ qui, à une distribution (resp. une mesure) μ sur \mathbf{Z}_p à valeurs dans F (resp. \mathcal{O}_F) associe sa transformée de Fourier, induit un isomorphisme de l'algèbre des distributions (resp. des mesures) sur $\mathbf{B}_{\text{rig},F}^+ \cong \mathcal{R}_F^+$ (resp. sur $\mathbf{A}_F^+ = \mathcal{O}_F[[\pi]]$). (On rappelle qu'une mesure est une application linéaire continue sur l'espace des fonctions continues et qu'une distribution est une application linéaire continue sur l'espace des fonctions localement analytiques.) Cette remarque n'est pas utilisée dans ce texte, mais on pourra consulter [25] pour un survol des questions dans lesquelles elle joue un rôle.

4.2. C et ses sous-corps

4.2.1. *Le corps C et l'action de \mathcal{G}_F .* Soit C le complété de \overline{F} pour la valuation v_p . C est un corps algébriquement clos muni d'une action continue de \mathcal{G}_F . Si K est une extension finie de F , soit $\widehat{K}_\infty \subset C$ le complété de K_∞ .

PROPOSITION 4.1 ([68]). — *La famille d'anneaux $(\mathcal{O}_C, (\mathcal{O}_{K_n})_{n \in \mathbf{N}})$ satisfait aux conditions de Tate-Sen relativement au couple $(\mathcal{G}_K, \mathcal{H}_K)$.*

La démonstration de Tate repose sur la théorie de la ramification qui permet en particulier de montrer que $v_p(d_{K_n/F_n})$ tend vers 0 quand n tend vers $+\infty$ (si $K \subset L$ sont deux extensions finies de F , on note $d_{L/K}$ la différentielle de l'extension K/L); dans la terminologie de Faltings cela se traduit par : \langle l'extension K_∞/F_∞ est presque étale \rangle . Les applications $R_n : \widehat{K}_\infty \rightarrow K_n$ dont on a besoin pour \langle décompléter \rangle s'obtiennent de la manière suivante : si $x \in K_\infty$, alors $\frac{1}{[K_m:K_n]} \text{Tr}_{K_m/K_n}(x)$ ne dépend pas de l'extension K_m de K_n contenant x et l'application $K_\infty \rightarrow K_n$ ainsi définie s'étend par continuité en une application $R_n : \widehat{K}_\infty \rightarrow K_n$.

On déduit de la proposition 4.1 [en fait on a juste besoin de $C^{\mathcal{H}_K} = \widehat{K}_\infty$ et de la propriété (TS2)] et de ce que Γ_K agit sur K_n à travers un quotient fini, le fait que C ne contient ni d'analogue p -adique de $\log 2i\pi$ ni d'analogue p -adique de $2i\pi$. De manière précise, on a

PROPOSITION 4.2 ([68]). — (i) C ne contient aucun élément y tel que l'on ait $g(y) = y + \log \chi(g)$ quel que soit $g \in \mathcal{G}_K$.

(ii) Si $k \in \mathbf{Z} - \{0\}$ et si $x \in C$ vérifie $g(x) = \chi(g)^k x$ pour tout $g \in \mathcal{G}_K$, alors $x = 0$.

4.2.2. *L'anneau \mathbf{B}_{HT} .* Soit $\mathbf{B}_{\text{HT}} = C[t, t^{-1}]$. On munit \mathbf{B}_{HT} d'une action continue de \mathcal{G}_F en faisant agir $g \in \mathcal{G}_F$ sur t par la formule $g(t) = \chi(g)t$. On a alors $\mathbf{B}_{\text{HT}}^{\mathcal{G}_K} = K$ d'après la proposition 4.2 et le théorème d'Ax-Sen-Tate. On munit \mathbf{B}_{HT} d'une graduation indexée par les entiers en posant $\text{Gr}^i \mathbf{B}_{\text{HT}} = C \cdot t^i$ si $i \in \mathbf{Z}$. Cette graduation est stable sous l'action de \mathcal{G}_F .

4.3. Les anneaux de caractéristique p

Soit $a = \{x, v_p(x) \geq \frac{1}{p}\}$ et soit

$$\tilde{\mathbf{E}}^+ = \{(x_n)_{n \in \mathbf{N}}, x_n \in \mathcal{O}_{\mathbf{C}}/a \text{ et } x_{n+1}^p = x_n \text{ si } n \in \mathbf{N}\}.$$

L'anneau $\mathcal{O}_{\mathbf{C}}/a$ étant de caractéristique p , l'application $x \mapsto x^p$ en est un morphisme et $\tilde{\mathbf{E}}^+$ est un anneau de caractéristique p sur lequel $\mathcal{G}_{\mathbf{F}}$ agit naturellement (composante par composante). D'autre part, si $(x_n)_{n \in \mathbf{N}} \in \tilde{\mathbf{E}}^+$ et si \hat{x}_n est un relèvement quelconque de x_n dans $\mathcal{O}_{\mathbf{C}}$, la suite de terme général $\hat{x}_{n+k}^{p^k}$ converge dans $\mathcal{O}_{\mathbf{C}}$ vers une limite $x^{(n)}$ qui ne dépend pas du choix des \hat{x}_n . Ceci permet de décrire $\tilde{\mathbf{E}}^+$ comme l'ensemble des suites $x = (x^{(0)}, \dots, x^{(n)}, \dots)$ d'éléments de $\mathcal{O}_{\mathbf{C}}$ vérifiant $(x^{(n+1)})^p = x^{(n)}$. Soit $v_{\mathbf{E}} : \tilde{\mathbf{E}} \rightarrow \mathbf{R}$ l'application définie par $v_{\mathbf{E}}(x) = v_p(x^{(0)})$.

On peut voir $\varepsilon = (1, \varepsilon^{(1)}, \dots, \varepsilon^{(n)}, \dots)$ comme un élément de $\tilde{\mathbf{E}}^+$ et, si on pose $\bar{\pi} = \varepsilon - 1$, on a $v_{\mathbf{E}}(\bar{\pi}) = \frac{p}{p-1}$.

Si \mathbf{K} est une extension finie de \mathbf{F} , soient

$$\tilde{\mathbf{E}}_{\mathbf{K}}^+ = \{(x_n)_{n \in \mathbf{N}} \in \tilde{\mathbf{E}}^+, x_n \in \mathcal{O}_{\mathbf{K}_{\infty}}/a \text{ si } n \in \mathbf{N}\},$$

$$\mathbf{E}_{\mathbf{K}}^+ = \{(x_n)_{n \in \mathbf{N}} \in \tilde{\mathbf{E}}^+, x_n \in \mathcal{O}_{\mathbf{K}_n}/a \text{ si } n \text{ est assez grand}\}.$$

$\mathbf{E}_{\mathbf{K}}^+$ contient ε et $\bar{\pi}$, ce qui nous permet de poser $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}^+[\bar{\pi}^{-1}]$, $\tilde{\mathbf{E}}_{\mathbf{K}} = \tilde{\mathbf{E}}_{\mathbf{K}}^+[\bar{\pi}^{-1}]$ et $\mathbf{E}_{\mathbf{K}} = \mathbf{E}_{\mathbf{K}}^+[\bar{\pi}^{-1}]$ si \mathbf{K} est une extension finie de \mathbf{F} .

THÉORÈME 4.3 ([43, 76]). — (i) $\tilde{\mathbf{E}}$ est un corps dont $v_{\mathbf{E}}$ est une valuation pour lequel il est complet et dont le corps résiduel est $\bar{k}_{\mathbf{F}}$. De plus l'action naturelle de $\mathcal{G}_{\mathbf{F}}$ sur $\tilde{\mathbf{E}}$ est continue.

(ii) $\mathbf{E}_{\mathbf{F}} = k_{\mathbf{F}}((\bar{\pi}))$ et, plus généralement, si \mathbf{K} est une extension finie de \mathbf{F} , alors $\mathbf{E}_{\mathbf{K}}$ muni de $v_{\mathbf{E}}$ est un corps complet pour une valuation discrète de corps résiduel $k_{\mathbf{F}}$.

(iii) Le sous-corps $\mathbf{E} = \cup_{\mathbf{K} \subset \bar{\mathbf{F}}} \mathbf{E}_{\mathbf{K}}$ est une clôture séparable de $\mathbf{E}_{\mathbf{F}}$ stable par $\mathcal{G}_{\mathbf{F}}$ et, si \mathbf{K} est une extension finie de \mathbf{F} , alors $\text{Gal}(\mathbf{E}/\mathbf{E}_{\mathbf{K}}) = \mathcal{H}_{\mathbf{K}}$; en particulier, $\mathcal{H}_{\mathbf{K}}$ agit continûment sur \mathbf{E} muni de la topologie discrète.

(iv) $\tilde{\mathbf{E}}$ (resp. $\tilde{\mathbf{E}}_{\mathbf{K}}$) est le complété de \mathbf{E} (resp. de la clôture radicielle de $\mathbf{E}_{\mathbf{K}}$) pour la valuation $v_{\mathbf{E}}$; en particulier, $\tilde{\mathbf{E}}$ est algébriquement clos et $\tilde{\mathbf{E}}_{\mathbf{K}} = \tilde{\mathbf{E}}^{\mathcal{H}_{\mathbf{K}}}$.

REMARQUE 4.4. — (i) Ce théorème permet de relier la théorie de Galois des corps locaux de caractéristique 0 à celle des corps locaux de caractéristique p ; c'est le point de départ de beaucoup des constructions de Fontaine.

(ii) Le corps $\mathbf{E}_{\mathbf{K}}$ peut naturellement être mis en bijection avec la limite projective des \mathbf{K}_n relativement aux applications $N_{\mathbf{K}_{n+1}/\mathbf{K}_n}$; c'est ce qui lui vaut l'appellation de < corps des normes >.

(iii) La théorie du corps des normes a été développée par Fontaine et Wintenberger [43, 76], dans un cadre beaucoup plus général que celui de l'extension cyclotomique (celui des extensions < arithmétiquement profinies >). En particulier, le cas où on remplace

l'extension cyclotomique par l'extension obtenue en rajoutant un système compatible de racines p^n -ièmes de p à l'air prometteur (cf. [9]).

(iv) Les (ii) et (iii) du théorème sont loin d'être évidents, mais les ingrédients utilisés pour leur démonstration se résument à

- le lemme de Hensel ;
- l'extension K_∞/F_∞ est presque étale ;
- si $x \in \mathcal{O}_{F_{n+1}}$ et $\sigma \in \mathcal{G}_{F_n}$, alors $v_p(\sigma(x) - x) \geq \frac{1}{p-1}$;

les deux derniers points permettant, en particulier, de montrer que l'on a $x^p \equiv N_{K_{n+1}/K_n}(x) \pmod{a}$, si n est assez grand et $x \in \mathcal{O}_{K_{n+1}}$.

4.4. De la caractéristique p à la caractéristique 0

4.4.1. *Notations.* On peut utiliser les foncteurs \mathcal{E} , \mathcal{E}^\dagger , \mathcal{R} , ... du n° 2.2 pour remonter les anneaux précédents en caractéristique 0. Pour cela, on doit choisir une uniformisante de \mathbf{E}_F et un Frobenius sur $\mathcal{E}(T)$ et un choix judicieux, pour des raisons qui apparaîtront bientôt, consiste à prendre $\bar{\pi}$ comme uniformisante de \mathbf{E}_F et $T \rightarrow (1+T)^p - 1$ pour φ . La grande nouveauté par rapport au n° 2.2 est que tous ces anneaux se retrouvent munis d'une action de \mathcal{G}_F plutôt que d'une action de $\mathcal{G}_{k((T))}$. On introduit donc de nouvelles notations mettant l'accent sur les extensions de F plutôt que sur celles de $k((T))$. On pose

$$\tilde{\mathbf{B}} = \mathcal{E}(\tilde{\mathbf{E}}), \quad \tilde{\mathbf{B}}_K = \mathcal{E}(\tilde{\mathbf{E}}_K), \quad \mathbf{B} = \mathcal{E}(\mathbf{E}), \quad \mathbf{B}_K = \mathcal{E}(\mathbf{E}_K)$$

$$\tilde{\mathbf{B}}^\dagger = \mathcal{E}^\dagger(\tilde{\mathbf{E}}), \quad \tilde{\mathbf{B}}_K^\dagger = \mathcal{E}^\dagger(\tilde{\mathbf{E}}_K), \quad \mathbf{B}^\dagger = \mathcal{E}^\dagger(\mathbf{E}), \quad \mathbf{B}_K^\dagger = \mathcal{E}^\dagger(\mathbf{E}_K)$$

$$\tilde{\mathbf{B}}_{\text{rig}}^\dagger = \mathcal{R}(\tilde{\mathbf{E}}), \quad \tilde{\mathbf{B}}_{\text{rig},K}^\dagger = \mathcal{R}(\tilde{\mathbf{E}}_K), \quad \mathbf{B}_{\text{rig},K}^\dagger = \mathcal{R}(\mathbf{E}_K)$$

Les anneaux ci-dessus et $\tilde{\mathbf{A}}^+ = \mathcal{O}_\varepsilon^+(\tilde{\mathbf{E}}) = W(\tilde{\mathbf{E}}^+)$, $\tilde{\mathbf{B}}^+ = \mathcal{E}^+(\tilde{\mathbf{E}})$, $\tilde{\mathbf{B}}_{\text{rig}}^+ = \mathcal{R}^+(\tilde{\mathbf{E}})$ sont plus ou moins suffisants pour les énoncés, mais pour travailler, on a aussi besoin des anneaux intermédiaires

$$\mathbf{A} = \mathcal{O}_E(\mathbf{E}) = W(\mathbf{E}), \quad \mathbf{A}_K = \mathcal{O}_E(\mathbf{E}_K), \quad \mathbf{A} = \mathcal{O}_E(\mathbf{E}), \quad \mathbf{A}_K = \mathcal{O}_E(\mathbf{E}_K)$$

$$\mathbf{A}^{(0,r]} = \mathcal{O}_E^{(0,r]}(\mathbf{E}), \quad \mathbf{A}_K^{(0,r]} = \mathcal{O}_E^{(0,r]}(\mathbf{E}_K), \quad \mathbf{A}^{(0,r]} = \mathcal{O}_E^{(0,r]}(\mathbf{E}), \quad \mathbf{A}_K^{(0,r]} = \mathcal{O}_E^{(0,r]}(\mathbf{E}_K)$$

$$\mathbf{B}^{(0,r]} = \mathcal{E}^{(0,r]}(\mathbf{E}), \quad \mathbf{B}_K^{(0,r]} = \mathcal{E}^{(0,r]}(\mathbf{E}_K), \quad \mathbf{B}^{(0,r]} = \mathcal{E}^{(0,r]}(\mathbf{E}), \quad \mathbf{B}_K^{(0,r]} = \mathcal{E}^{(0,r]}(\mathbf{E}_K)$$

$$\mathbf{B}^{]0,r]} = \mathcal{E}^{]0,r]}(\mathbf{E}), \quad \mathbf{B}_K^{]0,r]} = \mathcal{E}^{]0,r]}(\mathbf{E}_K), \quad \mathbf{B}^{]0,r]} = \mathcal{E}^{]0,r]}(\mathbf{E}), \quad \mathbf{B}_K^{]0,r]} = \mathcal{E}^{]0,r]}(\mathbf{E}_K)$$

Notons que les anneaux $\mathbf{B}_{\text{rig}}^\dagger$ et $\mathbf{B}^{[0,r]}$ sont absents des listes ci-dessus (comme \mathbf{E} est dense dans $\tilde{\mathbf{E}}$, on a $\mathcal{R}(\mathbf{E}) = \mathcal{R}(\tilde{\mathbf{E}})$ et $\mathbf{B}^{[0,r]}(\mathbf{E}) = \mathbf{B}^{[0,r]}(\tilde{\mathbf{E}})$ ce qui nous empêche de prendre la définition évidente). On les définit comme étant les sous-anneaux $\mathbf{B}_{\text{rig},K}^\dagger \otimes_{\mathbf{B}_K^\dagger} \mathbf{B}^\dagger$ (ceci ne dépend pas du choix de K) et $\mathbf{B}_F^{[0,r]} \otimes_{\mathbf{B}_F^{(0,r)}} \mathbf{B}^{(0,r]}$ de $\tilde{\mathbf{B}}_{\text{rig}}^\dagger$ et $\tilde{\mathbf{B}}^{[0,r]}$ respectivement.

On peut aussi décrire les anneaux ci-dessus de manière plus algébrique (cette description étant d'ailleurs fort utile [5] pour étudier leur lien avec \mathbf{B}_{dR}^+). Par exemple, si $\tilde{\mathbf{A}}^+\{X\}$ désigne l'ensemble des séries $\sum_{k=0}^{+\infty} a_k X^k$, où $a_k \rightarrow 0$ quand $k \rightarrow +\infty$, et si $\omega \in \tilde{\mathbf{A}}^+$ n'est pas une unité de $\tilde{\mathbf{A}}^+$ (i.e. si $v_{\mathbf{E}}(\bar{\omega}) > 0$), alors :

$$\begin{aligned} \tilde{\mathbf{B}}^\dagger &= \bigcup_{k \geq 1} \left(\tilde{\mathbf{A}}^+ \left\{ \frac{p}{\omega^k} \right\} \left[\frac{1}{p} \right] \right) \\ \tilde{\mathbf{B}}_{\text{rig}}^+ &= \bigcap_{n \geq 1} \left(\tilde{\mathbf{A}}^+ \left\{ \frac{\omega^n}{p} \right\} \left[\frac{1}{p} \right] \right) \\ \tilde{\mathbf{B}}_{\text{rig}}^\dagger &= \bigcup_{k \geq 1} \left(\bigcap_{n \geq k} \left(\tilde{\mathbf{A}}^+ \left\{ \frac{p}{\omega^k}, \frac{\omega^n}{p} \right\} \left[\frac{1}{p} \right] \right) \right) \end{aligned}$$

4.4.2. *Action de \mathcal{G}_F .* L'action de \mathcal{G}_F sur $\tilde{\mathbf{E}}$ se prolonge en une action continue (pour la topologie faible) sur $\tilde{\mathbf{A}}$ et $\tilde{\mathbf{B}}$ qui commute à celle du morphisme de Frobenius φ . De manière explicite, on a

$$\varphi \left(\sum_{k \gg -\infty}^{+\infty} p^k [x_k] \right) = \sum_{k \gg -\infty}^{+\infty} p^k [x_k^p] \quad \text{et} \quad \sigma \left(\sum_{k \gg -\infty}^{+\infty} p^k [x_k] \right) = \sum_{k \gg -\infty}^{+\infty} p^k [\sigma(x_k)] \text{ si } \sigma \in \mathcal{G}_F.$$

Ces actions s'étendent par continuité aux anneaux $\tilde{\mathbf{B}}^\dagger$, $\tilde{\mathbf{B}}_{\text{rig}}^\dagger$, $\tilde{\mathbf{B}}_{\text{rig}}^+$, $\tilde{\mathbf{B}}^{(0,r]}$ et $\tilde{\mathbf{B}}^{[0,r]}$. L'action de φ est bijective sur $\tilde{\mathbf{B}}^\dagger$, $\tilde{\mathbf{B}}_{\text{rig}}^\dagger$ et $\tilde{\mathbf{B}}_{\text{rig}}^+$ et induit un isomorphisme de $\tilde{\mathbf{B}}^{(0,r]}$ sur $\tilde{\mathbf{B}}^{(0,\frac{r}{p}]}$ et de $\tilde{\mathbf{B}}^{[0,r]}$ sur $\tilde{\mathbf{B}}^{[0,\frac{r}{p}]}$.

Soit $\pi = [\varepsilon] - 1 \in \tilde{\mathbf{A}}^+$. Les choix que l'on a faits identifient \mathbf{B}_F au sous-corps $\mathcal{E}_F(\pi)$ de $\tilde{\mathbf{B}}$. La formule $g(\pi) = (1 + \pi)^{\chi(g)} - 1$ montre que \mathbf{B}_F est stable sous l'action de \mathcal{G}_F et que l'action de \mathcal{G}_F commute à celle de φ (c'est ce qui justifie les choix en question). Cette stabilité entraîne la stabilité de \mathbf{A} , \mathbf{B} et par suite, celle de \mathbf{B}^\dagger , $\mathbf{B}_{\text{rig}}^\dagger$, $\mathbf{B}_{\text{rig}}^+$, $\mathbf{B}^{(0,r]}$ et $\mathbf{B}^{[0,r]}$ sous l'action de \mathcal{G}_F .

Comme $\mathbf{E}^{\mathcal{N}_K} = \mathbf{E}_K$ et $\tilde{\mathbf{E}}^{\mathcal{N}_K} = \tilde{\mathbf{E}}_K$, les points fixes de l'un des anneaux ci-dessus sous l'action de \mathcal{H}_K est l'anneau ayant le même nom mais avec un K en indice ; par exemple $\mathbf{B}^{\mathcal{N}_K} = \mathbf{B}_K$, $(\mathbf{B}^\dagger)^{\mathcal{N}_K} = \mathbf{B}_K^\dagger$ ou $(\mathbf{B}_{\text{rig}}^\dagger)^{\mathcal{N}_K} = \mathbf{B}_{\text{rig},K}^\dagger$. D'autre part, l'action de \mathcal{H}_K sur \mathbf{E} étant continue pour la topologie discrète, \mathcal{H}_K agit continûment sur \mathbf{A} et \mathbf{B} pour la topologie forte.

L'action de φ sur les anneaux avec \sim est bijective, mais pas sur les anneaux sans \sim , et on rajoute un n en indice pour indiquer que l'on a appliqué φ^{-n} . Par exemple,

$$\mathbf{B}_{K,n}^\dagger = \varphi^{-n}(\mathbf{B}_K^\dagger) \subset \tilde{\mathbf{B}}_K^\dagger, \quad \mathbf{A}_{K,n}^{(0,r]} = \varphi^{-n}(\mathbf{A}_K^{(0,p^{-n}r]}) \subset \tilde{\mathbf{A}}_K^{(0,r]}.$$

PROPOSITION 4.5 ([11]). — Si K est une extension finie de F , alors $(\tilde{\mathbf{A}}^{(0,r]}, (\mathbf{A}_{K,n}^{(0,r]})_{n \in \mathbf{N}})$ vérifie les conditions de Tate-Sen relativement au couple $(\mathcal{G}_K, \mathcal{H}_K)$, pour tout $r > 0$ suffisamment petit.

4.5. Le logarithme et l'anneau $\tilde{\mathbf{B}}_{\log}^{\dagger}$

Si $x \in \tilde{\mathbf{A}}^{\dagger}$ vérifie $v_{\mathbf{E}}(\bar{x} - 1) > 0$, la série $\log x = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n} (x - 1)^n$ converge dans $\tilde{\mathbf{B}}_{\log}^{\dagger}$. Par exemple, $t = \log[\varepsilon]$ est un élément de $\tilde{\mathbf{B}}_{\log}^{\dagger}$ sur lequel \mathcal{G}_F agit par multiplication par le caractère cyclotomique et qui peut être vu comme un analogue p -adique de $2i\pi$. On aimerait bien étendre cette application à $(\tilde{\mathbf{B}}^{\dagger})^*$, mais pour cela, on est forcé d'étendre un peu l'anneau $\tilde{\mathbf{B}}_{\log}^{\dagger}$ en rajoutant un analogue p -adique u de $\log p$.

Soit $\tilde{p} = (p, p^{1/p}, \dots) \in \tilde{\mathbf{E}}^+$. Si $\sigma \in \mathcal{G}_F$, il existe $c(\sigma) \in \mathbf{Z}_p$ tel que l'on ait $\sigma(\tilde{p}) = \tilde{p}\varepsilon^{c(\sigma)}$ ($\sigma \rightarrow c(\sigma)$ est le cocycle à valeurs dans $\mathbf{Z}_p(1)$ associé à p par la théorie de Kummer).

Soit $\tilde{\mathbf{B}}_{\log}^{\dagger} = \tilde{\mathbf{B}}_{\log}^{\dagger}[u]$. On munit $\tilde{\mathbf{B}}_{\log}^{\dagger}$ d'un opérateur \langle de monodromie $\rangle N = -\frac{d}{du}$ et d'une action de φ (resp. \mathcal{G}_F), compatible avec celle existant sur $\tilde{\mathbf{B}}_{\log}^{\dagger}$, en posant $\varphi(u) = pu$ (resp. $\sigma(u) = u + c(\sigma)t$). Les actions de φ et N commutent à celle de \mathcal{G}_F et on a $N\varphi = p\varphi N$.

PROPOSITION 4.6. — Il existe une unique application $\log : (\tilde{\mathbf{B}}^{\dagger})^* \rightarrow \tilde{\mathbf{B}}_{\log}^{\dagger}$ vérifiant les propriétés suivantes :

- (i) $\log xy = \log x + \log y$;
- (ii) $\log x = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n} (x - 1)^n$ si la série converge ;
- (iii) $\log[a] = 0$ si $a \in \overline{k}_F$;
- (iv) $\log p = 0$ et $\log[\tilde{p}] = u$.

De plus, on a $\varphi(\log x) = \log \varphi(x)$ et $\sigma(\log x) = \log \sigma(x)$ si $\sigma \in \mathcal{G}_F$.

On note $\tilde{\mathbf{B}}_{\log}^+$ le sous-anneau $\tilde{\mathbf{B}}_{\log}^+[u]$ de $\tilde{\mathbf{B}}_{\log}^{\dagger}$; il est stable par N , φ et \mathcal{G}_F . Si K est une extension finie de F , on note $\mathbf{B}_{\log,K}^{\dagger}$ le sous-anneau $\mathbf{B}_{\log,K}^{\dagger}[\log \pi_K]$ de $\tilde{\mathbf{B}}_{\log}^{\dagger}$; il est stable par N , φ et \mathcal{G}_K (qui agit à travers Γ_K).

4.6. L'anneau \mathbf{B}_{dR}^+

4.6.1. Construction de \mathbf{B}_{dR}^+ .

PROPOSITION 4.7 ([35]). — L'application $\theta : \tilde{\mathbf{A}}^+ \rightarrow \mathcal{O}_C$, donnée par $\sum_{n=0}^{+\infty} p^n [x_n] \mapsto \sum_{n=0}^{+\infty} p^n x_n^{(0)}$, est un morphisme surjectif d'anneaux dont le noyau est un idéal principal engendré par $\omega = \frac{\pi}{\varphi^{-1}(\pi)} = 1 + [\varepsilon] + \dots + [\varepsilon^{p-1}]$.

REMARQUE 4.8. — $(\tilde{\mathbf{A}}^+, \theta)$ est l'épaississement p -adique universel de \mathcal{O}_C (cf. [38]).

On prolonge θ en un morphisme de $\tilde{\mathbf{B}}^+$ sur C , on note \mathbf{B}_{dR}^+ l'anneau $\varprojlim \tilde{\mathbf{B}}^+ / (\ker \theta)^n$. Ceci fait de \mathbf{B}_{dR}^+ un anneau de valuation discrète d'idéal maximal $\ker \theta$ et de corps résiduel C . En particulier, la clôture séparable de F dans \mathbf{B}_{dR}^+ s'identifie à \overline{F} .

On munit \mathbf{B}_{dR}^+ de la topologie pour laquelle les $p^k \tilde{\mathbf{A}}^+ + (\ker \theta)^n$, avec $n, k \in \mathbf{N}$, forment une base de voisinages de 0. L'action de \mathcal{G}_F sur $\tilde{\mathbf{B}}^+$ s'étend par continuité en une action continue de \mathcal{G}_F sur \mathbf{B}_{dR}^+ . La série définissant $t = \log[\varepsilon] = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n} \pi^n$ converge dans \mathbf{B}_{dR}^+ et t est un générateur de $\ker \theta$ (ce qui explique pourquoi on a mis autant de temps à comprendre ce qu'était l'analogie p -adique de $2i\pi$).

On pose $\mathbf{B}_{\text{dR}} = \mathbf{B}_{\text{dR}}^+[t^{-1}]$, ce qui fait de \mathbf{B}_{dR} un corps et on munit \mathbf{B}_{dR} de la filtration décroissante définie par $\text{Fil}^i \mathbf{B}_{\text{dR}} = t^i \mathbf{B}_{\text{dR}}^+$. Cette filtration est stable par l'action de \mathcal{G}_K .

4.6.2. \mathbf{B}_{dR}^+ et le reste du monde. Il n'existe pas de morphisme continu de $\tilde{\mathbf{B}}_{\log}^+$ dans \mathbf{B}_{dR} car, bien que ces anneaux soient obtenus en localisant et complétant $\tilde{\mathbf{A}}^+$, les topologies sont trop différentes. Toutefois, $\tilde{\mathbf{B}}_{\log}^+$ est une limite inductive d'anneaux qui se plongent dans \mathbf{B}_{dR}^+ , mais il faut changer de plongement pour chacun de ces sous-anneaux. De manière précise,

(i) $\tilde{\mathbf{B}}_{\text{rig}}^+$ et \mathbf{B}_{dR}^+ sont deux complétés de $\tilde{\mathbf{B}}^+$ et l'identité $\tilde{\mathbf{B}}^+ \rightarrow \tilde{\mathbf{B}}^+$ se prolonge en une injection continue de $\tilde{\mathbf{B}}_{\text{rig}}^+$ dans \mathbf{B}_{dR}^+ .

(ii) La série $\log \frac{[p]}{p} = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n} \left(\frac{[p]}{p} - 1\right)^n$ converge dans \mathbf{B}_{dR}^+ , ce qui nous fournit une injection naturelle de $\tilde{\mathbf{B}}_{\log}^+$ dans \mathbf{B}_{dR}^+ qui commute à l'action de \mathcal{G}_F . Les anneaux $\tilde{\mathbf{B}}_{\text{rig}}^+$, $\tilde{\mathbf{B}}_{\log}^+$ et \mathbf{B}_{dR} sont alors reliés par les *suites exactes fondamentales* [38, 26]

$$\begin{aligned} 0 &\longrightarrow \tilde{\mathbf{B}}_{\text{rig}}^+[\tfrac{1}{t}] \longrightarrow \tilde{\mathbf{B}}_{\log}^+[\tfrac{1}{t}] \xrightarrow{\text{N}} \tilde{\mathbf{B}}_{\log}^+[\tfrac{1}{t}] \longrightarrow 0 \\ 0 &\longrightarrow (\tilde{\mathbf{B}}_{\text{rig}}^+[\tfrac{1}{t}])^{\varphi=1} \longrightarrow \tilde{\mathbf{B}}_{\text{rig}}^+[\tfrac{1}{t}] \xrightarrow{1-\varphi} \tilde{\mathbf{B}}_{\text{rig}}^+[\tfrac{1}{t}] \longrightarrow 0 \\ 0 &\longrightarrow \mathbf{Q}_p \longrightarrow (\tilde{\mathbf{B}}_{\text{rig}}^+[\tfrac{1}{t}])^{\varphi=1} \longrightarrow \mathbf{B}_{\text{dR}}/\text{Fil}^0 \longrightarrow 0 \end{aligned}$$

D'autre part, on a $\mathbf{B}_{\text{dR}}^{\mathcal{G}_K} = K$ et $(\tilde{\mathbf{B}}_{\log}^+[\tfrac{1}{t}])^{\mathcal{G}_K} = F$, et l'application naturelle $K \otimes_F \tilde{\mathbf{B}}_{\log}^+[\tfrac{1}{t}] \rightarrow \mathbf{B}_{\text{dR}}$ est injective [35, 38, 26].

(iii) Si $x = \sum_{k=0}^{+\infty} p^k [x_k] \in \tilde{\mathbf{A}}$, la série converge dans \mathbf{B}_{dR}^+ si et seulement si la série $\theta(x) = \sum_{k=0}^{+\infty} p^k x_k^{(0)}$ converge dans \mathbf{C} , c'est-à-dire si et seulement si $k + v_{\mathbf{E}}(x_k)$ tend vers $+\infty$ quand k tend vers $+\infty$. On en déduit une application naturelle $\iota_0 : \tilde{\mathbf{B}}^{(0,1]} \rightarrow \mathbf{B}_{\text{dR}}^+$ qui s'avère être injective.

(iv) Comme $\tilde{\mathbf{B}}^{[0,1]} = \tilde{\mathbf{B}}_{\text{rig}}^+ + \tilde{\mathbf{B}}^{(0,1]}$ et comme φ^{-n} induit une bijection continue de $\tilde{\mathbf{B}}^{[0,p^{-n}]}[u]$ sur $\tilde{\mathbf{B}}^{[0,1]}[u]$, on obtient [5], quel que soit $n \in \mathbf{N}$, une injection $\iota_n = \iota_0 \circ \varphi^{-n}$ de $\tilde{\mathbf{B}}^{[0,p^{-n}]}[u]$ dans \mathbf{B}_{dR}^+ . Cette injection peut être vue comme la \langle localisation en $\varepsilon^{(n)} - 1 \rangle$. Si K est une extension finie de F , il existe $n(K)$ tel que, pour tout $n \geq n(K)$, on ait

$$\iota_n \left(\mathbf{B}_K^{[0,p^{-n}]}[\log \pi_K] \right) \subset K_n[[t]].$$

5. CLASSIFICATION DES REPRÉSENTATIONS p -ADIQUES

5.1. Le (φ, Γ) -module associé à une représentation de \mathcal{G}_K

Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , le \mathbf{B}_K -espace vectoriel $\mathbf{D}(V) = (\mathbf{B} \otimes_{\mathbf{Q}_p} V)^{\mathcal{H}_K}$ est muni d'une action semi-linéaire de φ provenant de l'action de φ sur \mathbf{B} (et qui est donc *étale*, c'est-à-dire de φ -pente 0) et d'une action résiduelle de $\Gamma_K = \mathcal{G}_K/\mathcal{H}_K$. Pour des raisons évidentes, un tel objet s'appelle *un (φ, Γ) -module étale sur \mathbf{B}_K* .

Si D est un (φ, Γ) -module sur \mathbf{B}_K , alors $\mathbf{V}(D) = (\mathbf{B} \otimes_{\mathbf{B}_K} D)^{\varphi=1}$ est un \mathbf{Q}_p -espace vectoriel muni d'une action continue de \mathcal{G}_K .

THÉORÈME 5.1 ([37]). — *La catégorie des \mathbf{Q}_p -représentations de \mathcal{G}_K est équivalente à celle des (φ, Γ) -modules étales sur \mathbf{B}_K . Plus précisément, si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , alors $\mathbf{D}(V)$ est un (φ, Γ) -module étale sur \mathbf{B}_K et $\mathbf{V}(\mathbf{D}(V)) = V$ et, réciproquement, si D est un (φ, Γ) -module étale sur \mathbf{B}_K , alors $\mathbf{V}(D)$ est une \mathbf{Q}_p -représentation de \mathcal{G}_K et $\mathbf{D}(\mathbf{V}(D)) = D$.*

La démonstration de ce théorème repose sur les faits suivants :

— si on munit \mathbf{E} de la topologie discrète, alors $H^1(\mathcal{H}_K, \mathrm{GL}_d(\mathbf{E})) = 1$ d'après le théorème de Hilbert 90 et donc, si on munit \mathbf{B} de la topologie forte, alors $H^1(\mathcal{H}_K, \mathrm{GL}_d(\mathbf{B})) = 1$ (cf. prop. 3.3 et 3.4). Autrement dit, toute représentation de \mathcal{H}_K est \mathbf{B} -admissible et $V \rightarrow \mathbf{D}(V)$ a de bonnes propriétés.

— $\mathbf{B}^{\varphi=1} = \mathbf{Q}_p$, et donc $\mathbf{V}(\mathbf{D}(V)) = V$ si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K .

— si $(a_{i,j})_{1 \leq i,j \leq d} \in \mathrm{GL}_d(\mathbf{E})$, alors le système d'équations $x_i^p = \sum_{j=1}^d a_{i,j} x_j$, $1 \leq j \leq d$ admet p^d solutions dans \mathbf{E}^d et ces solutions forment un \mathbf{F}_p -espace vectoriel de dimension d engendrant \mathbf{E}^d . Un (φ, Γ) -module est étale si et seulement si il existe une base dans laquelle la matrice de φ appartient à $\mathrm{GL}_d(\mathbf{A}_K)$ et le résultat précédent permet de montrer que, si Δ est le \mathbf{A}_K réseau de D engendré par cette base, alors $(\mathbf{E} \otimes_{\mathbf{A}_K} \Delta)^{\varphi=1}$ est un \mathbf{F}_p -espace vectoriel de dimension d engendrant $\mathbf{E} \otimes_{\mathbf{A}_K} \Delta$; autrement dit, $\langle \mathbf{V}(D) \bmod p \rangle$ a la bonne dimension. Le résultat cherché s'en déduit \langle par dévissage et passage à la limite \rangle .

REMARQUE 5.2. — (i) Nous n'avons pas utilisé Γ_K pour montrer que les foncteurs $V \rightarrow \mathbf{D}(V)$ et $D \rightarrow \mathbf{V}(D)$ sont inverses l'un de l'autre. Ceci permet de montrer que la catégorie des \mathbf{Q}_p -représentations de $\mathcal{H}_K \cong \mathcal{G}_{k_F((T))}$ est équivalente à celle des φ -modules étales sur $\mathbf{B}_K \cong \mathcal{E}_F$. Ce résultat est une version locale de résultats généraux de Katz [51, chap. 4].

(ii) Nous avons privilégié dans ce texte les \mathbf{Q}_p -représentations, mais la théorie des (φ, Γ) -modules donne d'aussi bons résultats, en tensorisant par \mathbf{A} au lieu de \mathbf{B} , si on considère des \mathbf{Z}_p -modules de type fini (pas nécessairement libres) munis d'une action continue de \mathcal{G}_K (ou \mathcal{H}_K).

Comme application, on a par exemple le résultat suivant qui entre dans la démonstration du théorème 2.9 : si D est un φ -module étale sur \mathbf{B}_K , il existe une extension finie L de K

et une base de $\mathbf{B}_L \otimes_{\mathbf{B}_K} D$ sur \mathbf{B}_L dans laquelle la matrice de φ appartient à $1 + p^n M_d(\mathbf{A}_L)$; en effet, ce n'est qu'une traduction de ce que, si V est une \mathbf{Q}_p -représentation de \mathcal{H}_K , il existe un \mathbf{Z}_p -réseau T de V stable par \mathcal{H}_K et une extension finie L de K telle que \mathcal{H}_L agisse trivialement sur $T/p^n T$.

(iii) On peut utiliser l'équivalence de catégories ci-dessus pour étudier la cohomologie galoisienne et la théorie d'Iwasawa des représentations de \mathcal{G}_K ; nous renvoyons à [23] pour un résumé des résultats et à [46, 47, 4, 12] pour les détails.

(iv) Le théorème ci-dessus n'est pas assez fin pour vraiment étudier les problèmes de classification des représentations p -adiques ; le problème est que le corps \mathbf{B}_K est un objet un peu grossier et aller plus loin demande de descendre les coefficients à \mathbf{B}_K^\dagger .

La méthode de Sen permet, utilisant la proposition 4.5, de prouver le résultat suivant :

PROPOSITION 5.3 ([11]). — *Si K est une extension finie de F et d un entier ≥ 1 , les applications naturelles*

$$\lim_{\longrightarrow} H^1(\Gamma_K, GL_d(\mathbf{B}_{K,n}^\dagger)) \longrightarrow H^1(\Gamma_K, GL_d(\tilde{\mathbf{B}}_K^\dagger)) \longrightarrow H^1(\mathcal{G}_K, GL_d(\tilde{\mathbf{B}}^\dagger)),$$

induites par l'inflation de Γ_K à \mathcal{G}_K et les inclusions $\mathbf{B}_{K,n}^\dagger \subset \tilde{\mathbf{B}}_K^\dagger \subset \tilde{\mathbf{B}}^\dagger$, sont des bijections.

En particulier, comme le cocycle correspondant à une \mathbf{Q}_p -représentation de \mathcal{G}_K est fixe par φ^n pour tout n , il est dans l'image de $H^1(\Gamma_K, GL_d(\mathbf{B}_K^\dagger))$; on en déduit la \mathbf{B}_K^\dagger -admissibilité des représentations de \mathcal{H}_K obtenues par restriction d'une représentation de \mathcal{G}_K . Ceci permet, en définissant $\mathbf{D}^\dagger(V) = (\mathbf{B}^\dagger \otimes_{\mathbf{Q}_p} V)^{\mathcal{H}_K}$ pour une \mathbf{Q}_p -représentation de \mathcal{G}_K et $\mathbf{V}^\dagger(D) = (\mathbf{B}^\dagger \otimes_{\mathbf{B}_K^\dagger} D)^{\varphi=1}$ pour un (φ, Γ) -module sur \mathbf{B}_K^\dagger , de raffiner le théorème 5.1 sous la forme

THÉORÈME 5.4 ([11]). — *La catégorie des \mathbf{Q}_p -représentations de \mathcal{G}_K est équivalente à celle des (φ, Γ) -modules étales sur \mathbf{B}_K^\dagger . Plus précisément, si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , alors $\mathbf{D}^\dagger(V)$ est un (φ, Γ) -module étale sur \mathbf{B}_K et $\mathbf{V}^\dagger(\mathbf{D}^\dagger(V)) = V$ et, réciproquement, si D est un (φ, Γ) -module étale sur \mathbf{B}_K^\dagger , alors $\mathbf{V}^\dagger(D)$ est une \mathbf{Q}_p -représentation de \mathcal{G}_K et $\mathbf{D}^\dagger(\mathbf{V}^\dagger(D)) = D$.*

D'autre part, on montre que tout sous- \mathbf{B}_K^\dagger -espace vectoriel de dimension finie de $\mathbf{D}(V)$ stable par φ est inclus dans $\mathbf{D}^\dagger(V)$; ceci nous fournit le < résultat de descente > suivant dans lequel les représentations galoisiennes ont disparu.

COROLLAIRE 5.5. — *Si D est un (φ, Γ) -module étale sur \mathbf{B}_K , alors l'ensemble des sous- \mathbf{B}_K^\dagger -espaces vectoriels de dimension finie stables par φ admet un plus grand élément D^\dagger et on a $D = \mathbf{B}_K \otimes_{\mathbf{B}_K^\dagger} D^\dagger$. [On a $D^\dagger = \mathbf{D}^\dagger(\mathbf{V}(D))$.]*

5.2. La hiérarchie des représentations galoisiennes [39]

5.2.1. *Représentations de Hodge-Tate.* Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , le K -espace vectoriel $\mathbf{D}_{\text{HT}}(V) = (\mathbf{B}_{\text{HT}} \otimes V)^{\mathcal{G}_K}$ est muni d'une graduation indexée par les entiers et un

entier i tel que $\text{Gr}^{-i}\mathbf{D}_{\text{HT}}(V) \neq 0$ est un *poinds de Hodge-Tate* de V . Une représentation \mathbf{B}_{HT} -admissible est dite *de Hodge-Tate*. Une représentation C-admissible est clairement de Hodge-Tate et, réciproquement, une représentation de Hodge-Tate est C-admissible si 0 est son seul poinds de Hodge-Tate.

5.2.2. *Représentations de de Rham.* Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , le K -espace vectoriel $\mathbf{D}_{\text{dR}}(V) = (\mathbf{B}_{\text{dR}} \otimes V)^{\mathcal{G}_K}$ est muni d'une filtration par des sous- K -espaces vectoriels $\mathbf{D}_{\text{dR}}^i(V)$ pour $i \in \mathbf{Z}$ qui est décroissante (i.e. $\mathbf{D}_{\text{dR}}^{i+1}(V) \subset \mathbf{D}_{\text{dR}}^i(V)$) exhaustive (i.e. $\mathbf{D}_{\text{dR}}^i(V) = \mathbf{D}_{\text{dR}}(V)$ si $i \ll 0$) et séparée (i.e. $\mathbf{D}_{\text{dR}}^i(V) = \{0\}$ si $i \gg 0$). Une représentation \mathbf{B}_{dR} -admissible est dite *de de Rham*. Si V est de de Rham, alors V est de Hodge-Tate et $\mathbf{D}_{\text{HT}}(V)$ est le gradué associé à $\mathbf{D}_{\text{dR}}(V)$.

Utilisant l'injection $\iota_n : \widetilde{\mathbf{B}}^{(0,p^{-n})} \rightarrow \mathbf{B}_{\text{dR}}^+$, on peut retrouver $\mathbf{D}_{\text{dR}}(V)$ et sa filtration à partir de $\mathbf{D}^\dagger(V)$. (Si $r > 0$, on note $\mathbf{D}^{(0,r]}(V)$ le $\mathbf{B}_K^{(0,r]}$ -module $(\mathbf{B}^{(0,r]} \otimes V)^{\mathcal{H}_K}$; c'est [10] le plus grand sous- $\mathbf{B}_K^{(0,r]}$ -module D de type fini de $\mathbf{D}(V)$ tel que l'on ait $\varphi(D) \subset \mathbf{B}_K^{(0, \frac{r}{p}]} \otimes_{\mathbf{B}_K^{(0,r]}} D$)

PROPOSITION 5.6 ([5]+[41]). — Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , il existe $n(V)$ tel que si $n \geq n(V)$ et $i \in \mathbf{Z}$, alors

$$\mathbf{D}_{\text{dR}}^i(V) = \left(t^i K_n[[t]] \otimes_{\mathbf{B}_K^{(0,p^{-n})}} \mathbf{D}^{(0,p^{-n})}(V) \right)^{\Gamma_K},$$

où $t^i K_n[[t]]$ est considéré comme un $\mathbf{B}_K^{(0,p^{-n})}$ -module via l'injection $\iota_n : \mathbf{B}_K^{(0,p^{-n})} \rightarrow K_n[[t]]$.

5.2.3. *Représentations semi-stables.* Une \mathbf{Q}_p -représentation de \mathcal{G}_K qui est $\mathbf{B}_{\log}^+[\frac{1}{t}]$ -admissible (resp. $\mathbf{B}_{\text{rig}}^+[\frac{1}{t}]$ -admissible) est dite *semi-stable* (resp. *cristalline*). Une \mathbf{Q}_p -représentation V de \mathcal{G}_K est dite *potentiellement semi-stable* s'il existe une extension finie L de K telle que la restriction de V à \mathcal{G}_L soit semi-stable (en tant que représentation de \mathcal{G}_L). On a les implications suivantes :

- (i) $\langle V \text{ cristalline} \rangle \Rightarrow \langle V \text{ semi-stable} \rangle$;
- (ii) $\langle V \text{ semi-stable} \rangle \Rightarrow \langle V \text{ potentiellement semi-stable} \rangle$;
- (iii) $\langle V \text{ potentiellement semi-stable} \rangle \Rightarrow \langle V \text{ de de Rham} \rangle$;
- (iv) $\langle V \text{ de de Rham} \rangle \Rightarrow \langle V \text{ de Hodge-Tate} \rangle$.

La première implication vient de l'inclusion $\mathbf{B}_{\text{rig}}^+[\frac{1}{t}] \subset \mathbf{B}_{\log}^+[\frac{1}{t}]$, la seconde est une évidence, la troisième vient de ce que \mathbf{B}_{dR} contient $\mathbf{B}_{\log}^+[\frac{1}{t}]$ et $\overline{\mathbf{F}}$, et la dernière de ce que \mathbf{B}_{HT} est l'algèbre graduée associée à \mathbf{B}_{dR} .

En ce qui concerne les implications en sens inverse, on a

(o) il existe des représentations cristallines non triviales; par exemple, si $k \in \mathbf{Z}$ le \mathbf{Q}_p -espace vectoriel $\mathbf{Q}_p(k)$ de dimension 1 sur lequel $\sigma \in \mathcal{G}_K$ agit par multiplication par la puissance k -ième $\chi(\sigma)^k$ du caractère cyclotomique; plus généralement, et ça a été la motivation [35] pour l'introduction de toutes ces notions, si X est une variété propre et lisse

sur K possédant un modèle sur \mathcal{O}_K ayant bonne réduction, alors les \mathbf{Q}_p -représentations $H_{\text{ét}}^i(X_{\overline{K}}, \mathbf{Q}_p)$ de \mathcal{G}_K fournies par la cohomologie étale de X , sont cristallines ;

(i) la représentation $\sigma \rightarrow \begin{pmatrix} \chi(\sigma) & c(\sigma) \\ 0 & 1 \end{pmatrix}$ associée à $\log p$ est semi-stable mais pas cristalline ;

(ii) une représentation sur laquelle l'inertie de \mathcal{G}_K agit à travers un quotient fini non trivial est potentiellement semi-stable (et même potentiellement cristalline), mais pas semi-stable ;

(iii) l'implication $\langle V \text{ de Rham} \rangle \Rightarrow \langle V \text{ potentiellement semi-stable} \rangle$ est la conjecture de monodromie p -adique de Fontaine ;

(iv) une extension non triviale $0 \rightarrow \mathbf{Q}_p \rightarrow V \rightarrow \mathbf{Q}_p(k) \rightarrow 0$ est de Hodge-Tate mais pas de de Rham si $k > 0$;

(v) la représentation $\sigma \rightarrow \begin{pmatrix} 1 & \log \chi(\sigma) \\ 0 & 1 \end{pmatrix}$ associée à $\log 2i\pi$ n'est pas de Hodge-Tate.

5.2.4. (φ, N) -modules filtrés. Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , le F -espace vectoriel $\mathbf{D}_{\text{st}}(V) = (\mathbf{B}_{\log}^+[\frac{1}{t}] \otimes V)^{\mathcal{G}_K}$ est muni d'une action semi-linéaire de φ et d'une action linéaire de N vérifiant la relation $N\varphi = p\varphi N$. D'autre part, $K \otimes_F \mathbf{D}_{\text{st}}(V)$ s'injecte dans $\mathbf{D}_{\text{dR}}(V)$ et est muni d'une filtration par des sous- K -espaces vectoriels qui est décroissante, exhaustive et séparée. Un tel objet est appelé *un (φ, N) -module filtré sur K* . Le sous- F -espace vectoriel $\mathbf{D}_{\text{cris}}(V) = \mathbf{D}_{\text{st}}(V)^{N=0}$ de $\mathbf{D}_{\text{st}}(V)$ s'identifie à $(\mathbf{B}_{\text{rig}}^+[\frac{1}{t}] \otimes V)^{\mathcal{G}_K}$; c'est *un φ -module filtré sur K* .

Si D est un (φ, N) -module filtré sur K et $D_K = K \otimes_F D$, l'injection de $\mathbf{B}_{\log}^+[\frac{1}{t}]$ dans \mathbf{B}_{dR} induit une application

$$\left(\mathbf{B}_{\log}^+[\frac{1}{t}] \otimes_F D \right)^{N=0, \varphi=1} \longrightarrow \left(\mathbf{B}_{\text{dR}} \otimes_K D_K \right) / \left(\mathbf{B}_{\text{dR}} \otimes_K D_K \right)^0$$

dont le noyau $\mathbf{V}_{\text{st}}(D)$ est un \mathbf{Q}_p -espace vectoriel (pas forcément de dimension finie) muni d'une action de \mathcal{G}_K . (Les actions de φ , N et la filtration sur un produit tensoriel étant données par les formules naturelles $N(a \otimes d) = N(a) \otimes d + a \otimes N(d)$, $\varphi(a \otimes d) = \varphi(a) \otimes \varphi(d)$ $(\mathbf{B}_{\text{dR}} \otimes_K D_K)^i = \sum_{j \in \mathbf{Z}} \mathbf{B}_{\text{dR}}^{i-j} \otimes D_K^j$.)

Si D est un (φ, N) -module filtré sur K , on peut lui associer deux invariants numériques $t_N(D)$ et $t_H(D)$ définis par

$$t_N(D) = v_p(\det \varphi) \quad \text{et} \quad t_H(D) = \sum_{i \in \mathbf{Z}} i \cdot \dim_K(D_K^i / D_K^{i+1}).$$

(Comme φ n'est que semi-linéaire, son déterminant dépend de la base dans laquelle il est calculé, mais sa valuation p -adique n'en dépend pas.) On dit que D est *admissible* si on a $t_H(D) = t_N(D)$ et $t_H(D') \leq t_N(D')$ pour tout sous- F -espace vectoriel D' de D stable par φ et N (D'_K étant muni de la filtration induite).

5.2.5. *Construction des représentations semi-stables.* Le théorème suivant fournit une description concrète des représentations semi-stables et donc (modulo la conjecture de monodromie de Fontaine) des représentations de de Rham.

THÉORÈME 5.7 ([27, 26]). — *La catégorie des représentations semi-stables de \mathcal{G}_K est équivalente à celle des (φ, N) -modules filtrés admissibles sur K . Plus précisément, si V est une représentation semi-stable, alors $\mathbf{D}_{\text{st}}(V)$ est un (φ, N) -module filtré admissible sur K et $\mathbf{V}_{\text{st}}(\mathbf{D}_{\text{st}}(V)) = V$ et, réciproquement, si D est un (φ, N) -module filtré admissible sur K , alors $\mathbf{V}_{\text{st}}(D)$ est une représentation semi-stable de \mathcal{G}_K et $\mathbf{D}_{\text{st}}(\mathbf{V}_{\text{st}}(D)) = D$.*

De plus, cette équivalence de catégories induit une équivalence entre la catégorie des représentations cristallines et celle des φ -modules filtrés admissibles sur K .

On dispose de deux descriptions des représentations semi-stables : l'une, donnée par le théorème ci-dessus, en termes de (φ, N) -modules filtrés admissibles sur K et l'autre en termes de (φ, Γ) -modules étales sur \mathbf{B}_K^\dagger . Une question naturelle qui se pose est : < comment passe-t-on de l'une à l'autre ? >. Dans un sens, en utilisant le fait que $(\tilde{\mathbf{B}}_{\text{rig}, K}^\dagger, (\mathbf{B}_{\text{rig}, K, n}^\dagger)_{n \in \mathbf{N}})$ vérifie la propriété de Tate-Sen (TS2), on obtient le résultat suivant qui montre comment retrouver les actions de φ et N sur $\mathbf{D}_{\text{st}}(V)$ à partir de $\mathbf{D}^\dagger(V)$; couplé avec la proposition 5.6, il fournit une description du (φ, N) -module filtré $\mathbf{D}_{\text{st}}(\mathbf{V}^\dagger(D))$ si D est un (φ, Γ) -module étale sur \mathbf{B}_K^\dagger .

PROPOSITION 5.8 ([5]). — *Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , alors*

$$\mathbf{D}_{\text{cris}}(V) = \left(\mathbf{B}_{\text{rig}, K}^\dagger \left[\frac{1}{t} \right] \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V) \right)^{\Gamma_K} \quad \text{et} \quad \mathbf{D}_{\text{st}}(V) = \left(\mathbf{B}_{\text{log}, K}^\dagger \left[\frac{1}{t} \right] \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V) \right)^{\Gamma_K}.$$

Voir [74] pour des résultats de nature similaire. Le vrai problème est d'aller dans l'autre sens car il est nettement plus facile de construire à la main un (φ, N) -module filtré admissible qu'un (φ, Γ) -module. En d'autres termes, comment décrire les matrices de φ et γ sur $\mathbf{D}^\dagger(\mathbf{V}_{\text{st}}(D))$ à partir de D , si D est un (φ, N) -module filtré admissible sur K . Ce n'est pas une question complètement gratuite car on peut lire sur un (φ, Γ) -module des propriétés de la représentation modulo p qui sont fort mystérieuses sur le (φ, N) -module filtré. Une réponse < raisonnable > à cette question aurait des applications à la théorie des déformations des représentations galoisiennes.

Le seul cas où l'on ait une réponse satisfaisante est le cas [75] où $K = F$ est non ramifié, et où $\mathbf{V}_{\text{st}}(D)$ est une représentation cristalline à poids de Hodge-Tate compris entre 0 et $p - 1$. (Il s'agit d'une démonstration des résultats de Fontaine et Laffaille [42] via la théorie des (φ, Γ) -modules.) D'autre part, dans le cas où $K = F$ est non ramifié, et où $\mathbf{V}_{\text{st}}(D)$ est une représentation cristalline (sans hypothèse sur les poids de Hodge-Tate), on sait [24] que l'on peut se débrouiller pour que les matrices de φ et γ soient à coefficients dans $\mathcal{O}_F[[\pi]]$.

5.3. Représentations galoisiennes et équations différentielles

5.3.1. *L'opérateur de Sen.* Si on utilise la méthode de Sen et la proposition 4.1, on obtient le résultat suivant qui est précisément le cas qu'avait considéré Sen [66].

THÉORÈME 5.9 ([66]). — *Si $d \geq 1$, les applications naturelles*

$$\lim_{\rightarrow} H^1(\Gamma_K, GL_d(K_n)) \longrightarrow H^1(\Gamma_K, GL_d(\widehat{K}_\infty)) \longrightarrow H^1(\mathfrak{G}_K, GL_d(\mathbb{C})),$$

induites par l'inflation de Γ_K à \mathfrak{G}_K et les inclusions $K_n \subset \widehat{K}_\infty \subset \mathbb{C}$ sont des bijections.

D'après cette proposition, si V est une représentation p -adique de \mathfrak{G}_K et n est assez grand, alors $\mathbb{C} \otimes_{\mathbb{Q}_p} V$ possède un unique sous- K_n -espace vectoriel $\mathbf{D}_{\text{Sen},n}(V)$ fixe par \mathcal{H}_K , stable par \mathfrak{G}_K , de dimension $\dim_{\mathbb{Q}_p} V$ et tel que $\mathbb{C} \otimes_{K_n} \mathbf{D}_{\text{Sen},n}(V) = \mathbb{C} \otimes_{\mathbb{Q}_p} V$. (L'unicité ne sort pas directement de la proposition, mais se démontre avec les mêmes techniques.) On montre alors facilement que $\frac{\gamma-1}{\chi(\gamma)-1}$ vu comme opérateur \mathbb{Q}_p -linéaire de $\mathbf{D}_{\text{Sen},n}(V)$ tend vers une limite quand $\gamma \in \Gamma_K$ tend vers 1. On obtient de la sorte [66] un opérateur Θ_V qui est K_n -linéaire et peut être vu comme un < poids de Hodge-Tate généralisé > :

PROPOSITION 5.10. — *Si V est une représentation de Hodge-Tate de \mathfrak{G}_K , alors Θ_V est diagonalisable et ses valeurs propres sont les poids de Hodge-Tate de V (avec multiplicité). Réciproquement, si Θ_V est diagonalisable et ses valeurs propres sont des entiers, alors V est de Hodge-Tate. En particulier, V est \mathbb{C} -admissible si et seulement si $\Theta_V = 0$.*

D'autre part, Θ_V peut être vu comme un élément de $\mathbb{C} \otimes_{\mathbb{Q}_p} \text{End}(V)$ et Sen a démontré le résultat suivant :

THÉORÈME 5.11 ([65, 66]). — *La sous-algèbre de Lie de $\text{End}(V)$ engendrée par les logarithmes des éléments du sous-groupe d'inertie de \mathfrak{G}_K est la plus petite sous-algèbre de Lie de $\text{End}(V)$ définie sur \mathbb{Q}_p dont les \mathbb{C} -points contiennent Θ_V . En particulier, $\Theta_V = 0$ (i.e. V est \mathbb{C} -admissible) si et seulement si le sous-groupe d'inertie de \mathfrak{G}_K agit à travers un quotient fini.*

5.3.2. *Le (φ, ∇) -module associé à une représentation galoisienne.* On note ∂ la dérivation $(1 + \pi) \frac{d}{d\pi}$ de \mathbf{B}_F^\dagger . Cette dérivation s'étend de manière unique à \mathbf{B}_K^\dagger pour toute extension finie K de F et on a $\partial \circ \varphi = p \varphi \circ \partial$.

PROPOSITION 5.12 ([5]). — *Si $x \in \mathbf{B}_{\text{rig},K}^\dagger$, alors $\lim_{\gamma \rightarrow 1, \gamma \in \Gamma_K} \frac{\gamma-1}{\chi(\gamma)-1} x = \nabla x$, où $\nabla = t\partial = t \frac{d}{dt}$ est une dérivation de $\mathbf{B}_{\text{rig},K}^\dagger$.*

L'existence de la limite résulte d'une étude directe de l'action de Γ_K ; le reste est un calcul immédiat.

Ce qui précède correspond au cas de la représentation triviale; dans le cas général, on a le résultat suivant :

PROPOSITION 5.13 ([5]). — Si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K , la famille d'opérateurs $\frac{\gamma-1}{\chi(\gamma)-1}$ tend, quand $\gamma \in \Gamma_K$ tend vers 1, vers une connexion ∇_V au-dessus de ∇ sur $\mathbf{D}_{\text{rig}}^\dagger(V)$. De plus, si n est assez grand, cette connexion laisse stable le sous- $\mathbf{B}_K^{[0,p^{-n}]}$ -module $\mathbf{D}^{[0,p^{-n}]}(V) = \mathbf{B}_K^{[0,p^{-n}]} \otimes_{\mathbf{B}_K^{(0,p^{-n})}} \mathbf{D}^{(0,p^{-n})}(V)$ de $\mathbf{B}_{\text{rig},K}^\dagger \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V)$.

On a donc associé à toute \mathbf{Q}_p -représentation V de \mathcal{G}_K un module avec connexion et structure de Frobenius sur l'anneau de Robba $\mathbf{B}_{\text{rig},K}^\dagger$. Ce n'est malheureusement pas suffisant pour pouvoir utiliser les résultats généraux sur les équations différentielles p -adiques car t n'est pas inversible dans $\mathbf{B}_{\text{rig},K}^\dagger$, ce qui se traduit par l'existence potentielle d'une infinité de singularités dans toute couronne du type $0 < v_p(T) \leq r$. Comme t n'a que des pôles simples, ces singularités sont régulières et on peut localiser en $\varepsilon^{(n)} - 1$ pour les étudier.

5.3.3. *Localisation.* La proposition 3.3 permet de déduire de la proposition 5.3 le résultat suivant :

PROPOSITION 5.14. — Si $d \geq 1$, les applications naturelles

$$\lim_{\rightarrow} H^1(\Gamma_K, \text{GL}_d(K_n[[t]])) \longrightarrow H^1(\Gamma_K, \text{GL}_d((\mathbf{B}_{\text{dR}}^+)^{\mathcal{H}_K})) \longrightarrow H^1(\mathcal{G}_K, \text{GL}_d(\mathbf{B}_{\text{dR}}^+)),$$

induites par l'inflation de Γ_K à \mathcal{G}_K et les inclusions $K_n[[t]] \subset (\mathbf{B}_{\text{dR}}^+)^{\mathcal{H}_K} \subset \mathbf{B}_{\text{dR}}^+$ sont des bijections.

Comme dans le cas de C , cette proposition permet de montrer que, si V est une \mathbf{Q}_p -représentation de \mathcal{G}_K et si n est assez grand, alors $\mathbf{B}_{\text{dR}}^+ \otimes_{\mathbf{Q}_p} V$ possède un unique sous- $K_n[[t]]$ -module libre $\mathbf{D}_{\text{Dif},n}^+(V)$ fixe par \mathcal{H}_K , stable par \mathcal{G}_K , de rang $\dim_{\mathbf{Q}_p} V$ et tel que $\mathbf{B}_{\text{dR}}^+ \otimes_{K_n} \mathbf{D}_{\text{Dif},n}^+(V) = \mathbf{B}_{\text{dR}}^+ \otimes_{\mathbf{Q}_p} V$. La famille d'opérateurs $\frac{\gamma-1}{\chi(\gamma)-1}$ tend, quand $\gamma \in \Gamma_K$ tend vers 1, vers une connexion ∇_V au-dessus de $\nabla = t \frac{d}{dt}$ sur $\mathbf{D}_{\text{Dif},n}^+(V)$, et le résultat suivant est immédiat :

PROPOSITION 5.15 ([5]). — (i) On obtient $(\mathbf{D}_{\text{Sen},n}(V), \Theta_V)$ à partir de $(\mathbf{D}_{\text{Dif},n}^+(V), \nabla_V)$ en regardant modulo t .

(ii) L'isomorphisme $\mathbf{D}_{\text{Dif},n}^+(V) = K_n[[t]] \otimes_{\mathbf{B}_K^{[0,p^{-n}]}} \mathbf{D}^{[0,p^{-n}]}(V)$ commute à ∇_V .

Le (ii) montre que $\mathbf{D}_{\text{Dif},n}^+(V)$ est le localisé du (φ, ∇) -module $\mathbf{D}^{[0,p^{-n}]}(V)$ en $\varepsilon^{(n)} - 1$ et le (i) montre que le résidu de la connexion ∇_V en ce point est Θ_V . Pour pouvoir se débarrasser de ces singularités, il faut au moins que Θ_V soit diagonalisable à valeurs propres entières (i.e. que V soit de Hodge-Tate), mais ce n'est pas suffisant (voir ci-dessous) ; en tout cas, ce résultat montre que le (φ, ∇) -module associé à une représentation galoisienne admet en général une infinité de singularités < non enlevables >.

PROPOSITION 5.16 ([5, 41]). — (i) $\mathbf{D}_{\text{Dif},n}^+(V)$ est sans singularité si et seulement si V est C -admissible.

(ii) $\mathbf{D}_{\text{Dif},n}^+(V)$ est à singularités apparentes [la connexion a une base de sections horizontales dans $\mathbf{D}_{\text{Dif},n}(V) = K_n((t)) \otimes_{K_n[[t]]} \mathbf{D}_{\text{Dif},n}^+(V)$] si et seulement si V est de de Rham.

Le (i) est une évidence et le (ii) se démontre en utilisant la proposition 5.6. Remarquons que si V est de de Rham, alors $\mathbf{D}_{\text{Dif},n}(V) = K_n((t)) \otimes_K \mathbf{D}_{\text{dR}}(V)$ et donc $\mathbf{D}_{\text{Dif},n}(V)$ possède un sous- $K_n[[t]]$ -réseau sur lequel on peut diviser ∇_V par t , à savoir $K_n[[t]] \otimes_K \mathbf{D}_{\text{dR}}(V)$.

5.3.4. *Le (φ, ∂) -module attaché à une représentation de de Rham.* D'après le paragraphe précédent, si V est de de Rham, les singularités de ∇_V ne sont qu'apparentes d'un point de vue local; le problème est donc d'arriver à les supprimer simultanément ou, ce qui revient au même, de trouver un réseau sur lequel on puisse diviser ∇_V par t . Le lemme suivant fournit la clé de ce que l'on doit faire.

LEMME 5.17. — x est divisible par t dans $\mathbf{B}_{\text{rig},K}^\dagger$ si et seulement si $(\iota_n(x)) \in t\mathbf{B}_{\text{dR}}^+$, pour tout n assez grand.

Soit alors $\mathbf{D}_{\text{rig}}^\dagger(V) = \mathbf{B}_{\text{rig},K}^\dagger \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V)$, et soit $\mathbf{N}_{\text{dR}}(V)$ l'ensemble des $x \in \mathbf{D}_{\text{rig}}^\dagger(V)[\frac{1}{t}]$, tels que $\iota_n(x) \in K_n[[t]] \otimes_K \mathbf{D}_{\text{dR}}(V)$ si n est assez grand (dépendant de x). Ce $\mathbf{B}_{\text{rig},K}^\dagger$ -module est fermé et, si a et b sont respectivement le plus grand et le plus petit poids de Hodge-Tate de V , alors $\mathbf{N}_{\text{dR}}(V)$ contient $t^{-b}\mathbf{D}_{\text{rig}}^\dagger(V)$ et est contenu dans $t^{-a}\mathbf{D}_{\text{rig}}^\dagger(V)$; c'est donc un $\mathbf{B}_{\text{rig},K}^\dagger$ -module libre de rang d (cf. prop. 1.1). Un peu plus de travail permet d'obtenir :

THÉORÈME 5.18 ([5]). — Si V est une \mathbf{Q}_p -représentation de de Rham de \mathcal{G}_K de dimension d , alors $\mathbf{D}_{\text{rig}}^\dagger(V)[\frac{1}{t}]$ contient un unique sous- $\mathbf{B}_{\text{rig},K}^\dagger$ -module $\mathbf{N}_{\text{dR}}(V)$ libre de rang d stable par $\partial_V = t^{-1}\nabla_V$.

De plus, ce module est stable par φ et Γ_K et on a

$$\mathbf{B}_{\text{rig},K}^\dagger[\frac{1}{t}] \otimes_{\mathbf{B}_{\text{rig},K}^\dagger} \mathbf{N}_{\text{dR}}(V) = \mathbf{B}_{\text{rig},K}^\dagger[\frac{1}{t}] \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V).$$

Ce théorème admet comme corollaire la conjecture de monodromie de Fontaine (modulo la conjecture de monodromie de Crew) : si V est de de Rham, $\mathbf{N}_{\text{dR}}(V)$ est un (φ, ∂) -module sur l'anneau de Robba $\mathbf{B}_{\text{rig},K}^\dagger$. Un tel module étant quasi-unipotent d'après la conjecture de Crew, il existe une extension finie L de K telle que $\mathbf{B}_{\text{log},L}^\dagger[\frac{1}{t}] \otimes_{\mathbf{B}_K^\dagger} \mathbf{D}^\dagger(V)$ admette une base constituée de sections horizontales pour ∇_V . Vu le lien entre Γ_L et ∇_V , cela implique que Γ_L agit à travers un quotient fini sur l'espace vectoriel des sections horizontales, et la proposition 5.8 permet de montrer que V est semi-stable en tant que représentation de \mathcal{G}_{L_n} si n est assez grand.

REMARQUE 5.19 ([5]). — Si V est C-admissible, alors $\mathbf{N}_{\text{dR}}(V) = \mathbf{D}^\dagger(V)$. Ceci permet de déduire du théorème de Tsuzuki (cor. 2.10) l'équivalence (cf. th. 5.11)

$$\langle V \text{ est C-admissible} \rangle \Leftrightarrow \langle \text{l'inertie de } \mathcal{G}_K \text{ agit à travers un quotient fini} \rangle.$$

RÉFÉRENCES

- [1] Y. André, Représentations galoisiennes et opérateurs de Bessel p -adiques, *Ann. Inst. Fourier* **52** (2002).
- [2] Y. ANDRÉ, Filtrations de type Hasse-Arf et monodromie p -adique, *Inv. Math.* **148** (2002), 285–317.
- [3] J. AX, Zeroes of polynomials over local fields. The Galois action, *J. Algebra* **15** (1970), 417–428.
- [4] D. BENOIS, On Iwasawa theory of crystalline representations, *Duke Math. J.* **104** (2000) 211–267.
- [5] L. BERGER, Représentations p -adiques et équations différentielles, *Inv. Math.* **148** (2002), 219–284.
- [6] P. BERTHELOT, Finitude et pureté cohomologique en cohomologie rigide, *Inv. Math.* **128** (1997), 329–377.
- [7] D. BERTRAND, Groupes algébriques et équations différentielles linéaires, *Séminaire Bourbaki* 1991/92, exp. 750, *Astérisque* **206** (1992), 183–204.
- [8] N. BOURBAKI, *Algèbre commutative, IX-X*, Masson, Paris, 1983.
- [9] C. BREUIL, Une application du corps des normes, *Comp. Math.* **117** (1999), 189–203.
- [10] F. CHERBONNIER, Représentations p -adiques surconvergentes, thèse de l’université d’Orsay (1996).
- [11] F. CHERBONNIER et P. COLMEZ, Représentations p -adiques surconvergentes, *Inv. Math.* **133** (1998), 581–611.
- [12] F. CHERBONNIER et P. COLMEZ, Théorie d’Iwasawa des représentations p -adiques d’un corps local, *J. Amer. Math. Soc.* **12** (1999), 241–268.
- [13] B. CHIARELLOTTO et G. CHRISTOL, Overconvergent isocrystals and F-isocrystals, *Comp. Math.* **100** (1996), 77–99.
- [14] G. CHRISTOL, About a Tsuzuki theorem, *p -adic functional analysis (Ioannina, 2000)*, 63–74, *Lecture Notes in Pure and Appl. Math.* **222**, Dekker, New York, 2001.
- [15] G. CHRISTOL et B. DWORK, Modules différentiels sur des couronnes, *Ann. Inst. Fourier* **44** (1994), 689–720.
- [16] G. CHRISTOL et Z. MEBKHOUT, Sur le théorème de l’indice des équations différentielles p -adiques I, *Ann. Inst. Fourier* **43** (1993), 1545–1574.
- [17] G. CHRISTOL et Z. MEBKHOUT, Sur le théorème de l’indice des équations différentielles p -adiques II, *Ann. of Maths* **146** (1997), 345–410.

- [18] G. CHRISTOL et Z. MEBKHOUT, Sur le théorème de l'indice des équations différentielles p -adiques III, *Ann. of Maths* **151** (2000), 385–457.
- [19] G. CHRISTOL et Z. MEBKHOUT, Sur le théorème de l'indice des équations différentielles p -adiques IV, *Inv. Math.* **143** (2001), 629–672.
- [20] G. CHRISTOL et Z. MEBKHOUT, Équations différentielles p -adiques et coefficients p -adiques sur les courbes, *Astérisque* **279** (2002), 125–183.
- [21] G. CHRISTOL et P. ROBBA, *Équations différentielles p -adiques. Applications aux sommes exponentielles*, Actualités Mathématiques, Hermann, Paris, 1994.
- [22] P. COLMEZ, Sur un résultat de Shankar Sen, *C.R.A.S.* **318** (1994), 983–985.
- [23] P. COLMEZ, Représentations p -adiques d'un corps local, *Proceedings of the International Congress of Mathematicians II (Berlin 1998)*, Doc. Mat. Extra vol. II (1998), 153–162.
- [24] P. COLMEZ, Représentations cristallines et représentations de hauteur finie, *J. Reine Angew. Math.* **514** (1999), 119–143.
- [25] P. COLMEZ, Fonctions L p -adiques, *Séminaire Bourbaki 1998/99*, exp. 851, *Astérisque* **266** (2000), 21–58.
- [26] P. COLMEZ, Espaces de Banach de dimension finie, *J. Inst. Math. Jussieu* (à paraître).
- [27] P. COLMEZ et J-M. FONTAINE, Construction des représentations p -adiques semi-stables, *Inv. Math.* **140** (2000), 1–43.
- [28] R. CREW, F-isocrystals and p -adic representations, *Algebraic Geometry - Bowdoin 1985*, Proc. Symp. Pure Math. **XLVI** (2) (1987), 111–138.
- [29] R. CREW, Finiteness theorems for the cohomology of an overconvergent isocrystal on a curve, *Ann. scient. E.N.S.* **31** (1998), 717–763.
- [30] R. CREW, Canonical extensions, irregularities, and the Swan conductor, *Math. Ann.* **316** (2000), 19–37.
- [31] B. DWORK, On exponents of p -adic differential modules, *J. Reine Angew. math.* **484** (1997), 85–126.
- [32] B. DWORK, N. GEROTTO et F. SULLIVAN, *An introduction to G-functions*, Ann. of Math. Studies **133**, Princeton Univ. Press (1994).
- [33] G. FALTINGS, Almost étale extensions, *Astérisque* **279** (2002), 185–270.
- [34] J-M. FONTAINE, Modules galoisiens, modules filtrés et anneaux de Barsotti-Tate, *Journées arithmétiques de Rennes III*, *Astérisque* **65** (1979), 3–80.
- [35] J-M. FONTAINE, Sur certains types de représentations p -adiques du groupe de Galois d'un corps local ; construction d'un anneau de Barsotti-Tate. *Ann. Math.* **115** (1982), 529–577.

- [36] J-M. FONTAINE, Cohomologie de de Rham, cohomologie cristalline et représentations p -adiques, *Algebraic geometry (Tokyo/Kyoto,1982)*, SLN **1016** (1983), 86–108.
- [37] J-M. FONTAINE, Représentations p -adiques des corps locaux, *The Grothendieck Festschrift*, vol II, Birkhäuser, Boston 249–309, 1991.
- [38] J-M. FONTAINE, Le corps des périodes p -adiques, *Périodes p -adiques* exp. II, Astérisque **223** (1994), 59–102.
- [39] J-M. FONTAINE, Représentations p -adiques semi-stables. *Périodes p -adiques*, exp. III, Astérisque **223** (1994) 113–184.
- [40] J-M. FONTAINE, exposé à Orsay, 1998.
- [41] J-M. FONTAINE, Arithmétique des représentations galoisiennes p -adiques, Astérisque (à paraître).
- [42] J-M. FONTAINE et G. LAFFAILLE, Construction de représentations p -adiques, *Ann. Sci. E.N.S.* **15** (1982), 547–608.
- [43] J-M. FONTAINE et J-P. WINTENBERGER, Le corps des normes de certaines extensions algébriques de corps locaux, *C.R.A.S.* **288** (1979) 367–370.
- [44] A. GROTHENDIECK, lettres des 24/9/64 p. 183 et 3-5/10/64 p. 204, *Correspondance Grothendieck-Serre*, Documents Mathématiques **2**, S.M.F. (2001).
- [45] G. HENNIART, La conjecture de Langlands locale numérique pour $GL(n)$, *An. Sci. E.N.S.* **21** (1988), 497–544.
- [46] L. HERR, Sur la cohomologie galoisienne des corps p -adiques, *Bull. S.M.F.* **126** (1998) 563–600.
- [47] L. HERR, Une approche nouvelle de la dualité de Tate, *Math. Ann.* **320** (2001), 307–337.
- [48] O. HYODO, $H_g^1(K, V) = H_{st}^1(K, V)$, *proceedings of a symposium on arithmetic geometry*, ed. K. Kato, M. Kurihara, T. Saito, Univ. Tokyo (1991).
- [49] L. ILLUSIE, Autour du thémème de monodromie locale, *Périodes p -adiques*, exp. I, Astérisque **223** (1994), 9–57.
- [50] A. J. DE JONG, Smoothness, semi-stability and alterations, *Publ. Math. I.H.E.S.* **83** (1996), 51–93.
- [51] N. KATZ, p -adic properties of modular schemes and modular forms, *Modular functions of one variable III*, SLN **350** (1973), 69–190.
- [52] K. KEDLAYA, A p -adic local monodromy theorem, Preprint disponible sur xxx.arxiv.org : math.AG/0110124.
- [53] K. KEDLAYA, Full faithfulness for overconvergent F-isocrystals, preprint 2001.

- [54] K. KEDLAYA, Semi-stable reduction for overconvergent F-isocrystals on a curve, preprint 2001.
- [55] M. LAZARD, Les zéros des fonctions analytiques d'une variable sur un corps valué complet, Publ. Math. I.H.E.S. **14** (1962), 47–75.
- [56] F. LOESER, Exposants p -adiques et théorèmes d'indice pour les équations différentielles p -adiques (d'après G. Christol et Z. Mebkhout), *Séminaire Bourbaki* 1996/97, exp. 822, Astérisque **245** (1997), 57–81.
- [57] S. MATSUDA, Local indices of p -adic differential operators corresponding to Artin-Schreier-Witt coverings, Duke Math. J. **77** (1995), 607–625.
- [58] S. MATSUDA, Katz correspondence for quasi-unipotent overconvergent isocrystals, Comp. Math. (à paraître).
- [59] Z. MEBKHOUT, Sur le théorème de finitude de la cohomologie p -adique d'une variété affine non singulière, Amer. J. Math. **119** (1997), 1027–1081.
- [60] Z. MEBKHOUT, Analogie p -adique du théorème de Turrutin et le théorème de la monodromie p -adique, Inv. Math. **148** (2002), 319–351.
- [61] P. ROBBA, On the index of p -adic differential operators I, Ann. of Math. **101** (1975), 280–316.
- [62] P. ROBBA, On the index of p -adic differential operators II, Duke Math. J. **43** (1976), 19–31.
- [63] P. ROBBA, On the index of p -adic differential operators III, applications to twisted exponential sums, Astérisque **119-120** (1984), 191–266.
- [64] P. ROBBA, Indice d'un opérateur différentiel p -adique IV. Cas des systèmes. Mesure de l'irrégularité dans un disque, Ann. Inst. Fourier **35** (1985), 13–55.
- [65] S. SEN, Lie algebras of Galois groups arising from Hodge-Tate modules, Ann. of Math. **97** (1973), 160–170.
- [66] S. SEN, Continuous cohomology and p -adic Galois representations, Inv. Math. **62** (1980/81), 89–116.
- [67] J-P. SERRE, *Corps locaux*, Deuxième édition. Publications de l'Université de Nancago, No. VIII. Hermann, Paris, 1968. 245 pp.
- [68] J. TATE, p -divisible groups, *Proc. Conf. Local Fields (Driebergen, 1966)* 158–183 Springer, Berlin.
- [69] T. TSUJI, p -adic étale cohomology and crystalline cohomology in the semi-stable reduction case. Inv. Math. **137** (1999), 233–411.
- [70] T. TSUJI, Semi-stable conjecture of Fontaine-Jannsen : a survey, Astérisque **279** (2002), 323–370.

- [71] N. TSUZUKI, The local index and the Swan conductor, *Comp. Math.* **111** (1998), 245–288.
- [72] N. TSUZUKI, Slope filtration of quasi-unipotent overconvergent F-isocrystals, *Ann. Inst. Fourier* **48** (1998), 379–412.
- [73] N. TSUZUKI, Finite local monodromy of overconvergent unit-root F-isocrystal on a curve, *Amer. J. Math.* **120** (1998), 1165–1190.
- [74] N. WACH, Représentations p -adiques potentiellement cristallines, *Bull. S.M.F.* **124** (1996), 375–400.
- [75] N. WACH, Représentations cristallines de torsion, *Comp. Math.* **108** (1997), 185–240.
- [76] J-P. WINTENBERGER, Le corps des normes de certaines extensions infinies des corps locaux ; applications, *Ann. Sci. E.N.S.* **16** (1983), 59–89.
- [77] P. YOUNG, Radii of convergence and index for p -adic differential operators, *Trans. A.M.S.* **333** (1992), 769–785.

Pierre COLMEZ

Institut de mathématiques de Jussieu

4 place Jussieu

F-75005 PARIS

E-mail : colmez@math.jussieu.fr

ON THE $n!$ -CONJECTURE

by **Claudio PROCESI**

1. MACDONALD POLYNOMIALS, THE POSITIVITY CONJECTURE

In 1988 Macdonald defined 2-parameter symmetric functions unifying the theory of Hall-Littlewood and Jack polynomials (cf. [M],[M1]).

We recall a variant of his construction (cf. [H]). For given positive integer n we want to define symmetric functions $\tilde{H}_\lambda(X)$ indexed by partitions of n and with coefficients in $\mathbb{Q}(q, t)$. We use the fact that symmetric functions (in infinitely many variables) are polynomials in the Newton functions $\psi_k := \sum_i x_i^k$.

The $\tilde{H}_\lambda(X)$ are implicitly defined using dominance order and plethystic transformations.

For a symmetric function $f(X)$ the plethystic transformation $f(X) \rightarrow f(X[1-q])$ is the unique morphism $\mathbb{Q}[\psi_1, \dots, \psi_m, \dots]$ to $\mathbb{Q}(q)[\psi_1, \dots, \psi_m, \dots]$ sending $\psi_k \rightarrow \psi_k(1-q^k)$. The dominance order of partitions, is

$$(p_1, p_2, \dots, p_n) \geq (q_1, q_2, \dots, q_n), \iff \forall k \ p_1 + p_2 + \dots + p_k \geq q_1 + q_2 + \dots + q_k.$$

Finally for a partition λ its dual λ' is obtained exchanging rows and columns in its Young diagram.

1.1. THEOREM (Macdonald⁽¹⁾). — *There exist unique symmetric functions $\tilde{H}_\lambda(X)$ satisfying:*

1) $\tilde{H}_\lambda(X[1-q])$ lies in the vector space over $\mathbb{Q}(q, t)$ generated by the Schur functions $S_\mu(X)$, $\mu \geq \lambda$.

2) $\tilde{H}_\lambda(X[1-t])$ lies in the vector space over $\mathbb{Q}(q, t)$ generated by the Schur functions $S_\mu(X)$, $\mu \geq \lambda'$.

3) In the expansion of $\tilde{H}_\lambda(X)$ through Schur functions, the coefficient of S_λ is 1.

⁽¹⁾I follow here Haiman's approach [H].

Positivity conjecture

The coefficients of the expansion $\tilde{H}_\lambda(X)$ through Schur functions are polynomials in q, t with coefficients positive integers.

For further discussion and a guide through the literature we refer to [H].

2. $n!$ -CONJECTURE

Since the work of Frobenius, the connection between symmetric functions and representations of the symmetric group has been well understood. In particular it is useful to associate to the irreducible representation indexed by a partition λ the Schur function $S_\lambda(X)$. Extending this by linearity one has a linear isomorphism $\chi \rightarrow F(\chi)$, (called Frobenius character) between the space of characters of the symmetric group on n letters and the space of symmetric functions of degree n .

With this convention suppose we have a bigraded representation of the symmetric group $V_{i,j}$ and let $\chi_{i,j}$ be the corresponding bigraded character. Then we construct the 2-parameter symmetric function, called its bigraded Frobenius character:

$$\sum_{i,j} q^i t^j F(\chi_{i,j}).$$

So, to prove the positivity conjecture one should construct, for each partition λ , a bigraded representation whose bigraded Frobenius character is the Macdonald polynomial $\tilde{H}_\lambda(X)$.

In 1991 Adriano Garsia and Mark Haiman, inspired by similar constructions for the simpler case of q -Kostka polynomials (cf. [GP]), proposed such a construction.

Let $R := \mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]$ be the polynomial ring in $2n$ variables.

A partition λ of n will be always identified to a set of n points in the integral lattice. The n pairs $\lambda := \{(i_h, j_h)\}$, $h = 1, \dots, n$ of numbers give (up to sign) the polynomial:

$$D_\lambda := \det(x_k^{i_h} y_k^{j_h}) \in R, \quad (h, k = 1, \dots, n).$$

Consider the R module structure on R setting:

$$p.f := p\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}; \frac{\partial}{\partial y_1}, \dots, \frac{\partial}{\partial y_n}\right)f$$

set:

$$I_\lambda := \{p \in R \mid p.D_\lambda = 0\}, \quad V_\lambda := \{p.D_\lambda \mid p \in R\},$$

$V_\lambda \cong R/I_\lambda$ is the space spanned by all the derivatives of the polynomial D_λ .

$n!$ -conjecture [GH]: $\dim R/I_\lambda = \dim V_\lambda = n!$

We let the symmetric group S_n act on $R := \mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]$ by the diagonal action (or simultaneously on the x and y). Since D_λ is bihomogeneous and skew symmetric, it is clear that $V_\lambda = R/I_\lambda$ is a bigraded representation of the symmetric group so:

Second conjecture [GH]: The bigraded Frobenius character of V_λ is the Macdonald polynomial $\tilde{H}_\lambda(X)$.

Both conjectures have now been proved by Mark Haiman, his final results are in [H2]. It has turned out that the most difficult part of the project has been to establish the $n!$ -conjecture.

The proof of the $n!$ -conjecture is based on a deep property of the Hilbert scheme of n -tuples of points in the plane. This connection has also allowed Haiman to solve other conjectures on diagonal harmonics as we shall explain at the end (cf. [H1], [H3]).

In order to see how the Hilbert scheme enters, let us first make an elementary remark. Define a linear form T_λ on R as:

$$(2.1) \quad T_\lambda(p) := (p.D_\lambda)(0).$$

2.2. LEMMA. — $I_\lambda := \{p \in R \mid T_\lambda(pq) = 0, \forall q \in R\}$.

Proof. — A polynomial $p \in R$ is 0 if and only if $(q.p)(0) = 0$ for every $q \in R$. Thus $p.D_\lambda = 0$ if and only if $q.(p.D_\lambda)(0) = 0$, for every q . Since $q.(p.D_\lambda) = (qp).D_\lambda$ we have the claim. \square

Remark that, if h, k is a pair of natural integers, external to the partition λ we have, for every i :

$$\frac{\partial}{\partial x_i} \frac{\partial}{\partial y_i} D_\lambda = 0$$

It follows that I_λ contains the monomials $x_i^h y_i^k$ for every such pair. In other words, set J_λ to be the ideal of $\mathbb{C}[x, y]$ generated by the monomials $x^h y^k$, $(h, k) \notin \lambda$ and $A_\lambda := \mathbb{C}[x, y]/J_\lambda$. A_λ has dimension n and, as basis, the monomials $x^i y^j$, $(i, j) \in \lambda$.

Finally, identifying $R = \mathbb{C}[x, y]^{\otimes n}$ we have that R/I_λ is a quotient of $A_\lambda^{\otimes n}$.

The linear form T_λ factors through $A_\lambda^{\otimes n}$ and it is antisymmetric. Any antisymmetric linear form factors through antisymmetrization $A_\lambda^{\otimes n} \rightarrow \wedge^n A_\lambda$. We have $\dim \wedge^n A_\lambda = 1$, since $\dim A_\lambda = n$, hence such a form up to scalar is unique.

Thus the $n!$ -conjecture is equivalent to:

Rank conjecture The form $T_\lambda(pq)$ on $A_\lambda^{\otimes n}$ has rank $n!$.

It has turned out to be too difficult to analyze directly the ideal J_λ , but rather one must work more globally on H_n , the Hilbert scheme of all ideals I of codimension n in $\mathbb{C}[x, y]$. H_n comes together with the universal family

$$(2.3) \quad \mathcal{F} := \{(I, p) \in H_n \times \mathbb{C}^2 \mid I(p) = 0\}.$$

The projection map $p : \mathcal{F} \rightarrow H_n$ is flat and $F := p_* \mathcal{O}_{\mathcal{F}}$ is the tautological vector bundle $F = \{(I, \mathbb{C}[x, y]/I)\}$, a bundle of algebras of dimension n over H_n .

A fundamental theorem of Fogarty [F] states that the Hilbert scheme is smooth of dimension $2n$, from which it follows easily that it gives a (crepant) resolution of singularities:

$$H_n \xrightarrow{\rho} \mathbb{C}^{2n}/S_n.$$

In the Hilbert scheme the ideals J_λ play a special role. In fact on \mathbb{C}^2 and hence on H_n , acts a two dimensional torus $\mathcal{T} := \{(\alpha, \beta)\}$ through $(\alpha, \beta)(x, y) := (\alpha x, \beta y)$. A fixed point is just a bihomogeneous ideal and one easily sees that these are exactly the ideals J_λ (indexed by partitions of n).

One uses always the fact that any \mathcal{T} stable closed subset of H_n contains a fixed point.

Now we can globalize the rank conjecture. Consider the antisymmetrization $T : F^{\otimes n} \rightarrow \wedge^n F$ which induces a form $T(ab)$ on the bundle $F^{\otimes n}$.

2.4. LEMMA. — *T has generically rank $n!$. The $n!$ -conjecture is equivalent to the statement that the form T has always rank $n!$. In this case $F^{\otimes n}/\text{Ker}(T)$ is a bundle of algebras each carrying the regular representation of S_n .*

Proof. — In a generic point of the Hilbert scheme, the ideal I defines n -distinct points and the algebra $\mathbb{C}[x, y]/I = \bigoplus_{i=1}^n \mathbb{C}e_i$ with the e_i orthogonal idempotents.

It is easily seen that the kernel of T has, as basis, the elements $e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n}$ where the indices are not all distinct. Its complement is the regular representation with basis $e_{\sigma_1} \otimes e_{\sigma_2} \otimes \cdots \otimes e_{\sigma_n}$, $\sigma \in S_n$.

If the rank is constant we must have a bundle of regular representations. The set of points with rank $n!$ is open dense and \mathcal{T} stable, hence if the complement, where rank $< n!$, is non empty it must contain a fixed point J_λ contradicting the $n!$ conjecture. \square

Let now \mathcal{B}_n be the sheaf of sections of $F^{\otimes n}$ (a bundle of algebras) and \mathcal{I}_n the sheaf of ideals kernel of T . We have:

$$(2.5) \quad \text{Spec}(\mathcal{B}_n) := \{(I, p_1, p_2, \dots, p_n) \in H_n \times (\mathbb{C}^2)^n \mid I(p_i) = 0, \forall i = 1, \dots, n\}.$$

Now we can define the subvariety X_n of $\text{Spec}(\mathcal{B}_n)$ which is the closure of its open subset where all p_i are distinct.

2.6. LEMMA. — X_n is defined by the sheaf of ideals \mathcal{I}_n . The commutative diagram:

$$(2.7) \quad \begin{array}{ccc} X_n & \longrightarrow & \mathbb{C}^{2n} \\ p \downarrow & & \downarrow \\ H_n & \longrightarrow & \mathbb{C}^{2n}/S_n \end{array}$$

identifies X_n to the reduced fiber product.

Sketch of proof. The sheaf of ideals \mathcal{I}_n restricted to the part of H_n consisting of reduced subschemes, defines the set where all the p_i are distinct. One obtains immediately the first statement. In the commutative diagram, by construction, X_n is a subvariety of the reduced fiber product \overline{X}_n . On the regular part the fiber product is reduced, so it is enough to show that \overline{X}_n is irreducible. One sees this by induction on n using the fact that the preimage under p , of a subscheme supported in a unique point, is a point.

2.8. THEOREM. — The kernel of T has constant rank if and only if X_n is Cohen-Macaulay and Gorenstein. In this case $\mathcal{B}_n/\mathcal{I}_n = p_*(\mathcal{O}(X_n))$.

X_n is Cohen-Macaulay if and only if the morphism $p : X_n \rightarrow H_n$ is flat. In this case $\mathcal{B}_n/\mathcal{I}_n = p_*(\mathcal{O}(X_n))$.

Proof. — Some parts are fairly straightforward. Let us see how the Gorenstein property plays a role. Assume X_n is Cohen-Macaulay and Gorenstein. Take a point $I_\lambda \in H_n$ we have seen that there is a unique point in $p^{-1}(I_\lambda)$. The coordinate ring B_λ of the scheme theoretic fiber $p^{-1}(I_\lambda)$ is the local ring in this point modulo a regular sequence, hence by the Gorenstein assumption it has a 1-dimensional socle (a unique minimal ideal). It must necessarily be S_n stable and hence carry the sign representation, then the kernel of T on B_λ is an ideal which, if non 0, must contain the sign representation. This is clearly absurd so the form $T(ab)$ is non degenerate on the $n!$ dimensional algebra B_λ . The converse follows a similar line. \square

3. THE G -HILBERT SCHEME

It is quite interesting (and useful) to reinterpret the previous discussion as follows.

If we have that the map $p : X_n \rightarrow H_n$ is flat, we also have that each fiber of p is a subscheme of length $n!$ in \mathbb{C}^{2n} . From the theory of the Hilbert scheme we have then a classifying map $i : H_n \rightarrow H_{n!,2n}$ where $H_{n!,2n}$ is the Hilbert scheme parameterizing subschemes of length $n!$ in \mathbb{C}^{2n} . On the other hand, the open part of H_n corresponding to subschemes with n -distinct points parametrizes the generic orbits of S_n in \mathbb{C}^{2n} .

In general, given a finite group G acting faithfully on an irreducible quasi-projective variety X we have the following construction of Ito and Nakamura [I-N]. Consider the open set X^0 (the union of the generic orbits) over which G acts freely. The set of such orbits X^0/G can be identified with a locally closed subset of the Hilbert scheme $H_{|G|,X}$ of finite subschemes of length $|G|$ in X . One sees easily that this is in fact open in the subscheme of G fixed points of $H_{|G|,X}$.

The closure $\tilde{H}_{G,X}$ of X^0/G in $H_{|G|,X}$ is an irreducible component of the subscheme of G -stable finite schemes for which the coordinate ring carries the regular representation.⁽²⁾

By continuity the universal family restricted to $\tilde{H}_{G,X}$ is a flat family of G -stable finite schemes for which the coordinate ring carries the regular representation.

For a G -variety X this closure of X^0/G will be called the G -Hilbert scheme and denoted $\tilde{H}_{G,X}$. It comes equipped with a proper birational morphism $\rho : \tilde{H}_{G,X} \rightarrow X/G$. It seems to be quite interesting to determine when $\tilde{H}_{G,X}$ is smooth, and hence ρ a canonical resolution of singularities (cf. [BKR]). It is not hard to prove that, in our setting, from the flatness of $p : X_n \rightarrow H_n$ follows that H_n is identified to the G -equivariant Hilbert scheme of \mathbb{C}^{2n} and X_n to its universal family.

4. X_n IS COHEN-MACAULAY AND GORENSTEIN

The main theorem proved by Haiman in [H2], from which he deduces both the $n!$ and the Macdonald positivity conjectures, is:

4.1. THEOREM. — X_n is Cohen-Macaulay and Gorenstein.

The idea of the proof is to use induction on n and the fact that, on the points H_n^0 of H_n which do not define n coincident points, we have a local structure, analytically a product

$$\begin{array}{ccc} X_h \times X_k & \xrightarrow{g} & \mathbb{C}^{2h} \times \mathbb{C}^{2k} \\ p_h \times p_k \downarrow & & \downarrow \\ H_h \times H_k & \longrightarrow & \mathbb{C}^{2h} \times \mathbb{C}^{2k} / S_h \times S_k \end{array}$$

By induction if X_n^0 is the open set of X_n lying over H_n^0 , then X_n^0 is Cohen-Macaulay and Gorenstein, moreover $Rg_*^i \mathcal{O}(X_n^0) = 0$, $\forall i > 0$.

More precisely one has to use the *flag* Hilbert scheme $H_{n,n-1}$ and the corresponding variety $X_{n,n-1}$ and exploit the birational map $X_{n,n-1} \rightarrow X_n$.

⁽²⁾These subschemes are called G -clusters in [BKR].

Set theoretically $H_{n,n-1}$ is made of pairs of ideals

$$I \subset J \subset \mathbb{C}[x, y] \mid \dim \mathbb{C}[x, y]/I = n, \dim \mathbb{C}[x, y]/J = n - 1.$$

We have 3 natural morphisms:

$$H_{n,n-1} \xrightarrow{q} H_n, \quad H_{n,n-1} \xrightarrow{r} H_{n-1}, \quad H_{n,n-1} \xrightarrow{f} \mathbb{C}^2.$$

From a result of Tikhomirov and also Cheah (cf. [C]) we have:

4.2. THEOREM. — $H_{n,n-1}$ is smooth and $H_{n,n-1} \xrightarrow{q \times f} H_n \times \mathbb{C}^2$ has as image the universal family \mathcal{F} and it is a resolution of the singularities of \mathcal{F} .

Composing r with the morphism $H_{n-1} \xrightarrow{s} \mathbb{C}^{2(n-1)}/S_{n-1}$ we obtain

$$H_{n,n-1} \xrightarrow{sr \times f} \mathbb{C}^{2(n-1)}/S_{n-1} \times \mathbb{C}^2 = \mathbb{C}^{2n}/S_{n-1}.$$

At this point define $X_{n,n-1}$ as reduced fiber product:

$$(4.3) \quad \begin{array}{ccc} X_{n,n-1} & \longrightarrow & X_{n-1} \\ & \downarrow & \downarrow p \\ H_{n,n-1} & \longrightarrow & H_{n-1} \end{array}$$

On the varieties $X_{n,n-1}, X_n, H_{n,n-1}, H_n$ we have standard line bundles. On H_n define $\mathcal{O}_{H_n}(1)$ as the maximal exterior power of the tautological bundle and $\mathcal{O}_{H_n}(k) = \mathcal{O}_{H_n}(1)^k$. On $H_{n,n-1}$ we have the line bundles obtained by pull-back through the 2 projections, $H_{n,n-1} \xrightarrow{q} H_n, H_{n,n-1} \xrightarrow{r} H_{n-1}$ we set:

$$\mathcal{O}(h, k) := r^* \mathcal{O}_{H_{n-1}}(h) \otimes q^* \mathcal{O}_{H_n}(k).$$

On $X_{n,n-1}, X_n$ define also sheaves $\mathcal{O}(h, k), \mathcal{O}(k)$ by pull-back from $H_{n,n-1}, H_n$.

We can now reformulate Theorem 4.1 in a more precise form:

4.4. THEOREM. — $T(n) \quad X_n$ is Cohen-Macaulay, Gorenstein with dualizing sheaf $\mathcal{O}(-1)$.

$U(n) \quad X_{n,n-1}$ is Cohen-Macaulay and Gorenstein with dualizing sheaf $\mathcal{O}(0, -1)$.

The technique of the proof will be to follow the sequence of implications

$$T(n-1) \implies U(n) \implies T(n).$$

To proceed we must compute some canonical sheaves (cf. [H2], §3.6):

4.5. THEOREM. — For $H_{n,n-1}, H_n$ we have as canonical sheaves:

$$\omega_{H_{n,n-1}} = \mathcal{O}(1, -1), \quad \omega_{H_n} = \mathcal{O}(0) \quad \text{structure sheaf.}$$

It is necessary first of all to prove:

4.6. LEMMA. — *Suppose by induction that $X_{n-1} \rightarrow H_{n-1}$ is flat, the diagram*

$$\begin{array}{ccc} X_{n,n-1} & \longrightarrow & X_{n-1} \\ \downarrow & & \downarrow p \\ H_{n,n-1} & \longrightarrow & H_{n-1} \end{array}$$

is a fiber product.

Proof. — The statement claims that, forming the fiber product

$$\begin{array}{ccc} Y & \longrightarrow & X_{n-1} \times \mathbb{C}^2 \\ p' \downarrow & & \downarrow p \\ H_{n,n-1} & \longrightarrow & H_{n-1} \times \mathbb{C}^2 \end{array}$$

we have that Y is reduced and thus it coincides with $X_{n,n-1}$. In order to prove it we use the fact that p is flat and finite and so also p' is flat and finite, thus Y is Cohen-Macaulay and it suffices to prove that it is reduced in codimension 0, which one obtains restricting to the regular part of the diagram. \square

Now prove that $T(n-1) \implies U(n)$.

From the previous Lemma it follows that $X_{n,n-1}$ is Gorenstein, moreover the dualizing sheaf relative to the morphism $X_{n,n-1} \rightarrow X_{n-1}$ is the pull-back of the dualizing sheaf relative to the morphism $H_{n,n-1} \rightarrow H_{n-1}$ which is $\mathcal{O}(1, -1)$ while the dualizing sheaf of X_{n-1} is by induction $\mathcal{O}(-1, 0)$ thus the tensor product is $\mathcal{O}(0, -1)$.

Now the implication $U(n) \implies T(n)$.

One has to analyze the morphism $g : X_{n,n-1} \rightarrow X_n$. The main Theorem follows from general principles from the proposition:

4.7. PROPOSITION. —

$$(4.8) \quad R^i g_* \mathcal{O}(X_{n,n-1}) = 0, \forall i > 0, \quad g_* \mathcal{O}(X_{n,n-1}) = \mathcal{O}(X_n).$$

This proposition is based on a basic Lemma and some geometric considerations.

4.9. LEMMA. — *Given a proper morphism $g : Y \rightarrow X$ between algebraic varieties over \mathbb{C} . Suppose we have given m global functions z_1, \dots, z_m on X and let Z be the subvariety of X where they vanish and $U := X - Z$ the complement.*

Assume the following conditions:

- (1) *The z_i form a regular sequence in every local ring $\mathcal{O}_{X,P}$, $P \in Z$.*
- (2) *The z_i form a regular sequence in every local ring $\mathcal{O}_{Y,Q}$, $Q \in g^{-1}(Z)$.*
- (3) *Every fiber of g has dimension $< m - 1$.*

(4) On the open set U the canonical morphism $\mathcal{O}_X \rightarrow Rg_*\mathcal{O}_Y$ is an isomorphism, then the canonical morphism $\mathcal{O}_X \rightarrow Rg_*\mathcal{O}_Y$ is an isomorphism (everywhere).

This Lemma has a fairly simple cohomological proof ([H2], Lemma 3.8.5). The hard point is to apply this Lemma to the morphism g . We choose as sequence z_i the $n - 1$ functions $y_1 - y_2, \dots, y_1 - y_n$ and we need to verify the hypotheses (1)-(4).

The most difficult is (1).

In any case both for (1) and (2) one proves the stronger statement that y_1, y_2, \dots, y_n is a regular sequence.

One observes:

For (2) by induction all the local rings are Cohen-Macaulay, one must verify that the codimension of the variety given by the equations $y_1 = y_2 = \dots = y_n = 0$ is n . It is enough to do it for $H_{n,n-1}$, since $X_{n,n-1} \rightarrow H_{n,n-1}$ is finite.

For (3) it is enough to analyze the morphism $H_{n,n-1} \rightarrow H_n$.

The fiber has maximal dimension on the fixed points, i.e. the ideals J_λ , and consists of the ideals of dimension 1 of $\mathbb{C}[x, y]/J_\lambda$. By direct inspection one sees that this is a projective space of dimension $d - 1$ where d is the number boundary cases of the diagram. For $n > 3$ the inequality is easy while for $n \leq 3$ one must prove the Lemma directly.

To prove (4) we see that on U the morphism is locally isomorphic to a product of two morphisms $g : X_{k,k-1} \times X_{h,h-1} \rightarrow X_k \times X_h$. Thus we can proceed by induction.

(1) is the difficult part.

First a reduction: Presenting H_n and X_n as blow-ups.

First H_n . Let $R := \mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]$ be the polynomial ring in $2n$ variables:

$$S := \mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]^{S_n}, \quad A := \{f \in R \mid \sigma(f) = \epsilon_\sigma f, \sigma \in S_n\}$$

the invariants and the alternating elements under the diagonal action of S_n (ϵ_σ denotes the sign of the permutation). Finally let $I := AR$ be the ideal of R generated by A . We add an indeterminate t and consider the two graded algebras:

$$U_n := S \bigoplus_{i=1}^{\infty} A^i t^i, \quad V_n := R \bigoplus_{i=1}^{\infty} I^i t^i.$$

One acts with S_n on $R[t]$ by the diagonal action on R and acting on t with the sign representation.

4.10. THEOREM. — a) $U = V^{S_n}$.

b) $H_n = Proj(U_n)$, $X_n = Proj(V_n)$.

Sketch of proof. It is well known that H_n can be described as follows (cf. [N]).

Consider the variety Z of triples (X, Y, v) where X, Y are two $n \times n$ matrices with $XY = YX$, $v \in \mathbb{C}^n$ a vector. Define Z_0 to be the open set of Z where the vectors $X^i Y^j v$ generate the space \mathbb{C}^n .

Z_0 is smooth, the group $GL(n, \mathbb{C})$ acts freely on Z_0 and finally $H_n = Z_0/GL(n, \mathbb{C})$.

On Z many conjectures are open but we know that it is irreducible and that pairs of diagonalizable matrices are dense in Z .

Construct the quotient in two steps: $\tilde{H}_n = Z_0/SL(n, \mathbb{C})$, $H_n = \tilde{H}_n/\mathbb{C}^*$ we see that $Z_0/SL(n, \mathbb{C})$ is an open set of the variety $Z//SL(n, \mathbb{C})$.

By definition $Z//SL(n, \mathbb{C})$ is the spectrum of the ring of invariants and, by classical invariant theory such invariants are generated by:

$$\text{tr}(M), [M_1 v, M_2 v, \dots, M_n v]$$

where M, M_i denote monomials in the matrices X, Y and $[M_1 v, M_2 v, \dots, M_n v]$ denotes the determinant of these vectors.

By the previous remarks we can compute this ring by restricting it to pairs of diagonal matrices and then it is easy to see that it is identified to U_n . Moreover the open set $Z_0/SL(n, \mathbb{C})$ is the part where at least one of the determinants $[M_1 v, M_2 v, \dots, M_n v]$ is non 0 from which the statement easily follows.

As for X_n one has clearly a commutative diagram:

$$(4.11) \quad \begin{array}{ccc} \text{Proj}(V_n) & \longrightarrow & \mathbb{C}^{2n} \\ p \downarrow & & \downarrow \\ \text{Proj}(U_n) & \longrightarrow & \mathbb{C}^{2n}/S_n \end{array}$$

comparing it to 2.7 one has the claim.

At this point one has the next reduction:

In order to prove condition (1) it suffices to prove that, for every d the ideal I^d is a free module over the polynomial ring $\mathbb{C}[y_1, \dots, y_n]$.

This last statement will be further reduced to a more combinatorial statement. In order to explain it we must introduce some new objects, the *polygraphs*.

Thus, given a positive integer l let us denote by $[l]$ the segment $[1, 2, \dots, l]$.

Given a function $f : [l] \rightarrow [n]$ consider the induced linear map $\pi_f : (\mathbb{C}^2)^n \rightarrow (\mathbb{C}^2)^l$, and its graph $W_f \subset (\mathbb{C}^2)^n \times (\mathbb{C}^2)^l$.

The union $Z(n, l) := \cup_f W_f$ as f varies on the set of all functions $f : [l] \rightarrow [n]$ is a *polygraph*. Clearly on this polygraph operate various groups, in particular we will use the group S_l , and its subgroups, which permutes functions and graphs.

Call x_i, y_i the coordinates on $(\mathbb{C}^2)^n$ and a_i, b_i those on $(\mathbb{C}^2)^l$.

The coordinate ring of $Z(n, l)$ is a quotient $R(n, l) := \mathbb{C}[x_i, y_i, a_j, b_j]/I(n, l)$.

We consider next $l = nd$ and decompose

$$(\mathbb{C}^2)^n \times (\mathbb{C}^2)^n \times \cdots \times (\mathbb{C}^2)^n = (\mathbb{C}^2)^l$$

the group S_n^d operates permuting separately the coordinates of the factors and on the ring $R(n, l)$. Consider the subspace $R(n, l)^\epsilon$ of antisymmetric elements (with respect to all factors S_n).

4.12. LEMMA. — *There is a canonical isomorphism as R modules of $R(n, l)^\epsilon$ and I^d .*

Proof. — Let $f_0 : [l] \rightarrow [n]$ be defined as $f_0(kn + i) = i, \forall 1 \leq i \leq n, 0 \leq k < d$ call $W_{f_0} = (\mathbb{C}^2)^n$ the associated graph. The restriction of functions of $R(n, l)$ to such a space is a morphism of R modules and sends $a_{kn+i} \rightarrow x_i, b_{kn+i} \rightarrow y_i$, it is easily seen that it maps $R(n, l)^\epsilon$ surjectively to J^d .

It is enough to prove that it is also injective. In fact an antisymmetric function vanishes on W_f if $f(kn + i) = f(kn + j), 1 \leq i \leq n, 0 \leq k < d, i \neq j$. Any other function is in the orbit of f_0 relatively to S_n^d . By antisymmetry a function in $R(n, l)^\epsilon$ is determined completely by its values on W_{f_0} . \square

So finally let us show that everything will follow from the final key statement:

4.13. THEOREM. — *$R(n, l)$ is a free module on the polynomial ring $\mathbb{C}[y_1, \dots, y_n]$.*

Assume the last statement. Decomposing $R(n, l)$ into isotypic components with respect to S_n^d we see that $R(n, l)^\epsilon$ is a direct summand as module and so from the freeness of $R(n, l)$ follows also that of I^d isomorphic to $R(n, l)^\epsilon$.

At the end of this long sequence of reductions we have to face Theorem 4.12. This is proved by Haiman with a very long and complex induction only using a very careful bookkeeping and commutative algebra which occupies more than 30 pages of his paper. Rather than try to discuss this highly technical point we prefer to discuss some further developments. Nevertheless one should point out that, from the considerations that we will see in the next paragraph (Theorem 5.1), the appearance of polygraphs and their properties are geometrically natural.

5. DIAGONAL HARMONICS

The space of diagonal harmonics D_n can be defined as the subspace of polynomials in $\mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]$ solutions of the system of differential equations $P(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_i})f = 0$ where P runs over all polynomials without constant term which are symmetric with respect to the diagonal action of S_n .

D_n can also be identified with polynomial ring $\mathbb{C}[x_1, \dots, x_n; y_1, \dots, y_n]$ modulo the ideal generated by all S_n invariant polynomials without constant term.

Garsia and Haiman discovered a series of interesting conjectures mixing algebra, combinatorics and geometry on the space D_n , which is a bigraded, finite dimensional representation of S_n . The simplest of which describe its dimension as a vector space and its structure as representation.

- (1) $\dim D_n = (n + 1)^{n-1}$.
- (2) As a representation D_n is isomorphic to the permutation representation on the set of *parking functions* tensored by the sign representation.⁽³⁾

More precise conjectures on the bigraded character can be found in [GH1], in particular a rather remarkable expression for its bigraded Frobenius character.

One can attack these conjectures using the Lefschetz fixed point formula of Atiyah-Bott (cf. [AB]).

The idea is to prove that the space of diagonal harmonics can be identified with the global sections of the vector bundle $p_*\mathcal{O}_{X_n}$ restricted to the subvariety of H_n consisting of subschemes supported at 0 and then compute its character by localization principles.

This vector bundle is acted upon by the torus T and the Lefschetz fixed point formula of Atiyah-Bott can be applied provided one knows the vanishing of suitable cohomology groups. Finally these vanishing theorems can be deduced by applying the theory of Bridgeland King Reid to X_n (this can be done because of the solution of the $n!$ -conjecture). Their theory in our case establishes an equivalence of derived categories *the BKR correspondence* (generalized McKay correspondence) between the derived category of coherent sheaves on H_n and that of S_n -equivariant modules over R .

The announced geometric interpretation of polygraphs and their property is:

5.1. THEOREM. — *Under the BKR correspondence the polygraph $R(n, l)$ corresponds to the bundle F^l .*

⁽³⁾ n -people choose a spot out of n linearly ordered parking spots. They must reach their chosen spot and, if free park in it, otherwise reach the first free spot and park. The choice made is a *parking function* if, no matter in which order the people arrive they always park.

The locus $V(y_1, \dots, y_n)$ in X_n is a complete intersection.

This provides the requested vanishing theorems.

Once all of this is done one obtains finally an explicit formula, in term of Macdonald polynomials of the bigraded character of D_n (cf. [H3], 3.10 and 3.11). In particular the two previous conjectures follow from this formula.

Final comments

The theory we have sketched applies to the sum of two copies of the standard reflection representation of S_n . It thus suggests possible generalizations to other reflection groups.

At the moment it is not very clear what can be generalized and in which form, as the question of understanding which G -Hilbert schemes are smooth is completely mysterious. Haiman has made some computations for type B_n .

References

- [AB] Atiyah M, Bott R. – *A Lefschetz fixed point formula for elliptic differential operators*, Bull. Amer. Math. Soc. **72** (1966), 245–250.
- [Br] Briançon, J. – *Description de $\text{Hilb}^n \mathbb{C}[x, y]$* , Invent. Math. **41** (1977), n° 1, 45–89.
- [BKR] Bridgeland, T., King A., Reid M. – *Mukai implies McKay: the McKay correspondence as an equivalence of derived categories*, Electronic preprint, arXiv:math.AG/9908027, 1999.
- [Ch] Cheah J. – *Cellular decompositions for nested Hilbert schemes of points*, Pacific J. Math. **183** (1998), 39–90.
- [F] Fogarty J. – *Algebraic families on an algebraic surface*, Amer. J. Math. **90** (1968), 511–521.
- [GH] Garsia A., Haiman M. – *A graded representation model for Macdonald's polynomials*, Proc. Nat. Acad. Sci. U.S.A. **90**, n° 8 (1993), 3607–3610.
- [GH1] Garsia A., Haiman M. – *A remarkable $q-t$ -Catalan sequence and q -Lagrange inversion*, J. Algebraic Comb. **5**, n° 3 (1996), 191–244.
- [GP] Garsia A., Procesi C. – *On certain graded S_n -modules and the q -Kostka polynomials*, Adv. Math. **94**, n° 1 (1992), 82–138.
- [H] Haiman M. – *Macdonald polynomials and geometry*, New perspectives in geometric combinatorics (Billera, Björner, Greene, Simion and Stanley ed.), M.S.R.I. Publications **38** (1999), 207–254.
- [H1] Haiman M. – *Conjectures on the quotient ring by diagonal invariants*, J. Algebraic Combin. **5**, n° 1 (1994), 17–76.
- [H2] Haiman M. – *Hilbert schemes, polygraphs, and the Macdonald positivity conjecture*, Journal of the A.M.S. (2001) (to appear).
- [H3] Haiman M. – *Vanishing theorems and character formulas for the Hilbert scheme of points in the plane*, preprint (2001), 1–33.
- [IN] Ito Y., Nakamura I. – *McKay correspondence and Hilbert schemes*, Proc. Japan Acad. Ser. A Math. Sci. **72**, n° 7 (1996), 135–138.
- [M] Macdonald I.G. – *A new class of symmetric functions*, Actes du 20° Séminaire lotharingien, Publ. I.R.M.A. Strasbourg **372/S-20** (1988), 131–171.

- [M1] Macdonald I.G. – *Symmetric functions and Hall polynomials*, second edition, The Clarendon Press, Oxford University Press, New York (1995).
- [N] Nakajima H. – *Lectures on Hilbert schemes of points on surfaces*, American Math. Society, Providence RI (1999).

Claudio PROCESI

Dipartimento di Matematica

G. Castelnuovo

Università di Roma La Sapienza

Piazzale A. Moro

I-00185 ROMA (Italia)

E-mail: `claudio@mat.uniroma1.it`

DUALITÉS DE CHAMPS ET DE CORDES
[d'après 't Hooft, Polyakov, Witten et al.]

par **Daniel BENNEQUIN**

L'auteur souhaite seulement présenter, pour des mathématiciens, le cadre géométrique des dualités en physique. En particulier, faute de temps et de place, l'exposé écrit ne développe pas les principales applications en mathématique : symétrie miroir et théorie de Seiberg-Witten.

1. DUALITÉS CLASSIQUES

1.1. Une dualité s'installe lorsque deux points de vue opposés sont nécessaires pour décrire un même ensemble de phénomènes. La dualité est l'échange de points de vue complémentaires. Elle est partout présente en physique.

Une première manifestation de dualité est expliquée dans le cours de Feynman : la description d'un champ doit passer par la description du mouvement d'une particule test ou d'un appareil amplificateur (par exemple de la limaille de fer pour un champ magnétique ou des antennes pour les ondes gravitationnelles). À un niveau plus fin, la particule crée des champs ; plus fondamentalement, particule et appareil sont eux-mêmes des champs ; comment décrire cette interaction entre champs sans choisir un point de vue extérieur aux champs ?

La perception visuelle fait aussi intervenir une dualité, que B. Julia m'avait signalée : les photons qui renseignent sur une position de la matière ne communiquent que des impulsions ; à partir d'une fonction des impulsions et des fréquences (p, ω) , notre système nerveux extrait une fonction des positions et du temps (q, t) . La transformation de Fourier est une expression mathématique de cette dualité. Elle

traduit en Analyse la plus ancienne dualité développée en mathématique : la dualité projective de Poncelet.

En mécanique quantique, la transformation de Fourier devient un changement de repère (W. Heisenberg, [38]), exprimant la *dualité onde-corpuscule*. Le principe d'incertitude limite la compatibilité des points de vue mis en dualité, mais la mécanique quantique rétablit aussi un équilibre entre la matière et les champs d'interaction, en introduisant la fonction d'onde de Schrödinger.

Le domaine classique impose des équations différentielles à ses champs, par exemple l'équation de Maxwell pour le champ électromagnétique, l'équation d'Einstein pour le champ gravitationnel. La Mécanique quantique relativiste ajoute les équations de Dirac pour les fonctions d'ondes. Dualelement, le champ de Maxwell peut être vu comme fonction d'onde du photon.

Dans le vide sur \mathbf{R}^4 , les équations de Maxwell pour les champs électrique et magnétique E et B s'écrivent

$$\begin{aligned}\nabla \times B &= \frac{\partial E}{\partial t}, & \nabla \times E &= -\frac{\partial B}{\partial t} \\ \nabla \cdot B &= 0, & \nabla \cdot E &= 0\end{aligned}$$

elles sont symétriques pour la transformation

$$E \longmapsto B \quad , \quad B \longmapsto -E.$$

C'est surtout cette sorte de symétrie qui s'appelle *dualité* dans les développements récents (Misner et Wheeler, cf. [56]).

Pour que cette symétrie reste valable en présence de charges et de courants, Dirac a proposé, en 1931, d'ajouter des charges et des courants magnétiques aux côtés des charges et des courants électriques usuels ([14], [15]). Ce fut l'invention du *monopôle magnétique*, activement recherché mais toujours non détecté.

Cependant, le potentiel magnétique est un vecteur et le potentiel électrique, lui, est un scalaire, ce qui fait penser qu'à un niveau plus profond, notamment en mécanique quantique des difficultés s'opposent à la dualité électrique–magnétique (cf. Witten, 1997, [93]).

Dans la théorie des interactions faibles (principales responsables des effets radio-actifs), on rencontre une généralisation des équations de Maxwell, les équations de Yang-Mills-Higgs (cf. §1.2). Pour des équations analogues, en 1974, 't Hooft et Polyakov ont découvert des *monopôles magnétiques non abéliens*, et établi une forme de dualité échangeant magnétisme et électricité (cf. §2.3). Pourtant, là aussi, les degrés de liberté des potentiels sont différents.

C. Montonen et D. Olive (1977) ont conjecturé une vraie dualité électrique–magnétique, mais dans une *théorie de jauge* super-Yang-Mills-Higgs, avec 4 *supersymétries*, en dimension 4. Après des progrès notables de D. Olive, E. Witten (1978) et N. Seiberg (1988), la conjecture a été établie en 1994 par A. Sen (cf. § 2.4).

Une propriété remarquable de cette dualité est qu'elle échange deux *échelles* différentes de la théorie : les propriétés à courte distance avec celles aux grandes distances.

La *dualité* se révèle être un principe fondamental en théorie quantique des champs. Elle transforme le principe de correspondance de Bohr : une théorie quantique est associée à une théorie de champ classique, mais, une fois la théorie quantique développée, il apparaît d'autres théories classiques représentant certaines limites, certains points de vue différents, sur la même théorie quantique. Comme une variété peut avoir besoin de plusieurs cartes.

De plus, la plupart des dualités connues proviennent de dualités en théorie des cordes. Le creuset des dualités semble être la M-théorie (§ 3.3).

La fin du § 1 expose brièvement le cadre des champs classiques, avec des fermions à partir de 1.3 et des super-champs à partir de 1.4. Le § 2 donnera une idée de la renormalisation, ce qui est nécessaire pour préciser la définition d'une dualité et il exposera quelques exemples de dualité de champs. Le § 3 fera une introduction aux dualités de cordes qui semblent détenir le secret des dualités de champs.

1.2. En théorie de Yang-Mills (cf. [40], [55]), la variable est une connection ∇ sur un fibré principal P en groupe de Lie G au-dessus d'une variété W de dimension D . En choisissant une trivialisatoin locale de P , on écrit $\nabla = d + A$, avec une 1-forme différentielle A à valeurs dans l'algèbre de Lie \mathfrak{g} de G . La courbure F , ou F_A , de ∇ est la 2-forme définie par $F_A = dA + A \wedge A$; à l'origine classique, c'est F qui s'appelait le champ. Soit d_A la dérivée extérieure covariante, étendant ∇ en une dérivation graduée du module des formes différentielles à valeurs dans \mathfrak{g} (sur l'algèbre des formes différentielles scalaires), l'identité de Bianchi, toujours satisfaite, est $d_A F_A = 0$. Pour écrire l'équation de Yang-Mills, on choisit une métrique sur W , lorentzienne, de signature $(D - 1, 1)$ si on souhaite un temps réel, riemannienne si on préfère le temps imaginaire pur, et l'on note $*$ l'opérateur de Hodge-de Rham associé, échangeant les formes extérieures de degré p et $D - p$. Dans le vide, l'équation est

$$d_A * F_A = 0.$$

S'il y a des courants j , l'équation devient

$$d_A * F_A = *j.$$

Les équations de Maxwell sont obtenues dans le cas particulier $D = 4$, $G = U_1$; alors F est une 2-forme différentielle ordinaire. En signature lorentzienne, et en choisissant une coordonnée t pour le temps, la courbure F se décompose en un champ de (co)vecteurs E (les composantes en $dt \wedge dx^i$) et un champ de (bi)vecteurs B (les composantes en $dx^i \wedge dx^j$). Alors l'action de l'opérateur $*$ se traduit par $E \mapsto B$, $B \mapsto -E$.

Dans ce formalisme, avec $W = M \times \mathbf{R}$, M de dimension 3, pour toute surface fermée $\Sigma \subset M$, on définit les charges électriques et magnétiques par

$$Q_e(\Sigma) = \int_{\Sigma} \frac{*F}{2\pi} \quad , \quad Q_m(\Sigma) = \int_{\Sigma} \frac{F}{2\pi}.$$

Le fait que F est la courbure d'un fibré impose à $Q_m(\Sigma)$ d'être un nombre entier. En l'absence de mouvement des courants, $Q_e(\Sigma)$ est indépendant de t en vertu des équations de Maxwell. On dit que Q_e est une charge de Noether et que Q_m est une charge topologique.

Les équations de Yang-Mills sont les équations d'Euler-Lagrange de l'action dans le vide

$$S = \frac{1}{4g^2} \int_W \text{Tr}(F_A \wedge *F_A).$$

On dit que $\frac{1}{4g^2} \text{Tr}(F_A \wedge *F_A)$ est le lagrangien de la théorie.

La théorie est invariante par *changement de jauge*, c'est-à-dire par automorphisme de P induisant l'identité de W .

La variable de la gravitation d'Einstein est une métrique \mathbf{g} (lorentzienne pour un temps réel) sur W et le lagrangien est celui de Hilbert, c'est-à-dire $R \cdot \text{Vol}_{\mathbf{g}}$, où $\text{Vol}_{\mathbf{g}}$ est l'élément de volume de \mathbf{g} et R la courbure scalaire de \mathbf{g} . Einstein a aussi pensé au lagrangien plus général $(R - 2\Lambda) \cdot \text{Vol}_{\mathbf{g}}$, pour une constante Λ , la *constante cosmologique* (cf. [37], [89]). La théorie est invariante par difféomorphisme de W . Afin d'obtenir les équations du mouvement en présence d'autres champs, comme ceux de Maxwell ou de Yang et Mills, on ajoute les actions entre elles et on prend les équations variationnelles d'Euler-Lagrange. Les théories d'Einstein et de Maxwell ou Yang-Mills ensemble deviennent invariantes par tout automorphisme de P . (Cf. [9], [37], [42].)

Un des exemples les plus étonnants de dualité classique est celui d'Ehlers et Geroch pour la gravitation en dimension 4 ([23], [28]) :

Lorsqu'existe un champ de vecteurs K préservant \mathbf{g} , c'est-à-dire un champ de Killing, on peut voir la théorie d'Einstein comme une théorie de dimension 3 sur le quotient par K ; les champs étant la métrique quotient \mathbf{h} , un champ scalaire λ ,

venant de la norme de K , et une 1-forme \mathbf{A} exprimant le produit scalaire par K des vecteurs transverses. Cette forme \mathbf{A} hérite d'une invariance de jauge U_1 , et l'équation d'Einstein en dimension 4 impose l'équation de Maxwell : $d * d\mathbf{A} = 0$. J. Ehlers a donc pu introduire le *champ scalaire* dual, au sens de Poincaré-Hodge : $d\omega = *d\mathbf{A}$. Pour écrire les équations différentielles sur $\mathbf{h}, \lambda, \omega$, le mieux selon R. Geroch est de poser $\tau = \omega + i\lambda$ et $\tilde{\mathbf{h}} = \lambda\mathbf{h}$; dès lors, en notant $\tilde{\nabla}$ la dérivée covariante associée à $\tilde{\mathbf{h}}$, on a

$$\tilde{R} = \frac{1}{2} (\text{Im } \tau)^{-2} (\tilde{\nabla}\tau) \circ (\tilde{\nabla}\bar{\tau}), \quad d^{\tilde{\nabla}} * d^{\tilde{\nabla}} \tau = (\text{Im } \tau)^{-1} |\tilde{\nabla}\tau|^2.$$

Le miracle de dualité est l'invariance par $SL_2(\mathbf{R})$ de ces équations : en posant $\tilde{\mathbf{h}}' = \tilde{\mathbf{h}}$, et

$$\tau' = \frac{a\tau + b}{c\tau + d},$$

pour tout ensemble de nombres réels a, b, c, d , tels que $ad - bc = 1$, on déduit d'une solution \mathbf{h}, τ une autre solution \mathbf{h}', τ' .

C'était le rêve d'Einstein de trouver une théorie unitaire mettant tous les champs sur le même pied que la métrique \mathbf{g} . Une proposition dans ce sens fut celle de Th. Kaluza (1921), précisée par O. Klein (1926). L'idée est présente dans l'exemple précédent : la réduction dimensionnelle (cf. [7], [9], [21]).

Partant du lagrangien d'Einstein en dimension D , on impose à W d'être le produit d'une variété X par une variété M de dimension k , et on impose à la métrique le long de M de varier dans une famille de dimension finie, par exemple $M = G/H$ et $\mathbf{g}_{|M}$ G -invariante, ou encore (M, \mathbf{g}) , variété riemannienne d'Einstein-Kähler. Alors le champ \mathbf{g} sur W se réinterprète comme un ensemble de champs sur X . Dans le cas le plus simple, $M = S^1$, $X = \mathbf{R}^4$, on trouve le couplage sur X des équations d'Einstein et de Maxwell, mais accompagné d'un champ scalaire supplémentaire qui ne peut pas être pris constant dans les solutions. Lorsque $M = G$, groupe de Lie compact, on trouve à côté de \mathbf{g} sur X un champ de Yang-Mills ∇ , et comme champ supplémentaire, une métrique invariante (des deux côtés) sur G . Lorsque $M = G/H$, on trouve une métrique, une théorie de jauge (pour le normalisateur N de H dans G) et des champs de Higgs φ_1 , encore accompagnés de champs scalaires φ_2 (cf. [9]). Ces *champs scalaires ajoutés* φ , et leurs généralisations en présence d'autres champs que \mathbf{g} sur W , par exemple en présence d'un champ anti-symétrique \mathbf{b} sur W vont fournir le principal réservoir de dualité, par un mécanisme analogue à celui de l'exemple d'Ehlers ([10], [11], [46], [48]).

1.3. En face des tenseurs comme \mathbf{g} , F , φ véhiculant des interactions entre particules et se manifestant quantiquement comme des bosons, il y a, pour la matière ordinaire, des champs de spineurs ψ qui ont le comportement quantique des fermions.

Lorsque \mathbf{g} est une métrique définie positive, ou de signature lorentzienne $(D-1, 1)$, sur un espace vectoriel de dimension $D \geq 3$, la composante connexe de l'identité dans le groupe des rotations, $SO(\mathbf{g})$, possède un groupe fondamental à deux éléments. Le groupe $Spin(\mathbf{g})$ est le revêtement universel de $SO(\mathbf{g})$; c'est donc un revêtement à deux feuillets. Une *représentation spin* est une représentation linéaire complexe de $Spin(\mathbf{g})$ ayant la plus petite dimension possible. En dimension D impaire, il existe une représentation spin unique à isomorphisme près S de dimension $2^{(D-1)/2}$ sur \mathbf{C} ; en dimension D paire, il y en a deux S^+ , S^- de dimension $2^{D/2}$, dites de Weyl ([13], [33]). La somme $S = S^+ \oplus S^-$ s'appelle espace des spineurs de Dirac. Quand S est la complexifiée d'une représentation réelle $S_{\mathbb{R}}$, on dit que $S_{\mathbb{R}}$ est un espace de spineurs de Majorana. S'il existe $S_{\mathbb{R}}^+$ et $S_{\mathbb{R}}^-$, on parle de Majorana-Weyl.

Une structure spin sur une variété riemannienne orientée W est un revêtement double du fibré des repères orthonormés directs qui induit le revêtement $Spin(\mathbf{g})$ de $SO(\mathbf{g})$ en chaque point de W . Lorsqu'une structure spin existe, on peut définir des fibrés de spineurs encore notés S et un opérateur de Dirac $D : S \rightarrow S$; l'équation de Dirac sans masse $D\psi = 0$ dérive du lagrangien $\tilde{\psi} \cdot D\psi$ ([13]).

En présence d'un fibré principal $P \rightarrow W$ en groupe G muni d'une connexion ∇ et en choisissant un fibré vectoriel E associé à P , on a un opérateur de Dirac tordu $D_A : S \otimes E \rightarrow S \otimes E$, $D_A = D \otimes 1 + 1 \otimes \nabla$; le lagrangien associé est $\tilde{\psi} \cdot D_A\psi$.

Si l'on ajoute à ∇ et ψ les *champs de Higgs*, qui sont des sections de certains fibrés associés à P , on obtient tous les champs du « modèle standard ».

La théorie des représentations unitaires du groupe de Poincaré-Lorentz selon E. Wigner associe à toute particule une composante irréductible de la représentation du groupe des spineurs sur une puissance tensorielle $S \otimes S \otimes \dots \otimes S$ avec n facteurs ; dans ce cas, on parle d'*hélicité* $\frac{n}{2}$, ou encore de *spin* $\frac{n}{2}$. Par exemple, l'électron est de spin $\frac{1}{2}$; le photon, un champ de vecteurs, un champ de covecteurs, une connexion ∇ est de spin 1 ; une métrique est de spin 2, la particule quantique associée est le *graviton*. Les particules de spin demi-entier sont les *fermions*, celle de spin entier, les *bosons*.

Élie Cartan a montré que le couplage des équations d'Einstein avec celles de Dirac exige un changement de point de vue : il ne suffit plus de considérer \mathbf{g} , il faut ajouter une torsion des connexions métriques, car le tenseur énergie-moment d'un spineur

n'est pas symétrique. La méthode retenue pour résoudre le problème (Sciama-Kibble, 1961) consiste à introduire un sous-fibré réel V de $S \otimes S^*$, isomorphe au fibré tangent $T(W)$, et à faire varier l'isomorphisme $\varepsilon : T(W) \rightarrow V$, nommé *vielbein* et noté e_μ^α . Un choix de ε détermine une métrique. Les variables de la théorie sont alors ε , ψ et une connexion lorentzienne auxiliaire ω sur V (cf. [33]).

1.4. Toutes ces interactions, tous ces lagrangiens qui s'ajoutent font intervenir des échelles de grandeurs extrêmement différentes (des rapports de 10^{20}). La *super-symétrie*, introduite au début des années 70 (pour les besoins des cordes), se propose d'échanger les bosons avec les fermions ; si elle n'était qu'approximativement réalisée, cela pourrait expliquer les différences d'échelles.

Salam et Strathdee ont introduit la notion de *super-espace* dans ce contexte ; elle a depuis envahi toute la théorie des champs et des cordes (cf. [84], [55], [33], [13]).

Remarque : Il y a là un point qui gêne souvent le dialogue entre mathématiciens et physiciens, car les définitions adoptées couramment en mathématique et en physique sont différentes. Mais A.S. Schwartz [73] a montré qu'elles sont équivalentes :

Pour $d \in \mathbb{N} \cup \{\infty\}$, on note L_d l'algèbre extérieure $\Lambda(\mathbb{R}^d)$, $L_d^0 = \mathbb{R}$ l'ensemble de ses éléments de degré 0, L_d^+ et L_d^- les facteurs de degré pair > 0 et de degré impair respectivement.

Le modèle local d'une *super-variété* au sens de De Witt de dimension $(m|n)$ et de dimension cachée $d \in \mathbb{N} \cup \{\infty\}$ est un produit $U \times (\mathbb{R}^m \otimes L_d^+) \times (\mathbb{R}^n \otimes L_d^-)$, pour U ouvert de \mathbb{R}^m , que l'on note $U^{m|n}$. (Cf. [72], [84].)

Pour toute fonction C^∞ , $f : U \rightarrow \mathbb{R}$, on note $Z[f]$ la fonction de $U \times (\mathbb{R}^m \otimes L_d^+)$ dans L_d définie par la formule

$$Z[f](u_1, \dots, u_m) = \sum_{\alpha \in \mathbb{N}^m} \frac{D^\alpha f(a)}{\alpha!} (u^+)^{\alpha},$$

en décomposant les coordonnées : $u_i = a_i + u_i^+$ avec $a_i \in L_d^0$ et $u_i^+ \in L_d^+$.

Une *super-fonction* Φ de $U^{m|n}$ dans L_d est une fonction de $U^{m|n}$ dans L_d de la forme suivante, pour $u \in U \times (\mathbb{R}^m \otimes L_d^+)$ et $v \in \mathbb{R}^n \otimes L_d^-$:

$$\Phi(u, v) = \sum_{\beta \in \mathbb{N}^n} Z[f_\beta](u) \cdot v^\beta,$$

où les f_β sont des fonctions C^∞ de U dans \mathbb{R} .

On dit que Φ est paire (resp. impaire) si son image appartient à $L_d^+ \oplus L_d^0$ (resp. L_d^-).

À partir de là, on définit les *morphismes* $F = (\Phi_1, \dots, \Phi_p, \Psi_1, \dots, \Psi_q)$ de $U^{m|n}$ dans $V^{p|q}$ par leurs composantes super-fonctions, les Φ_i étant paires et les Ψ_j impaires. Ces morphismes s'appellent applications de classe H^∞ dans [72].

En recollant les modèles, on obtient une catégorie de super-variétés. Cette catégorie pour $d = \infty$ est équivalente à la catégorie des super-variétés de Berezin, Kostant, Manin ([13], [55], [72], [73]), c'est-à-dire celle des espaces annelés avec le modèle local de faisceaux $\Lambda^{m|n}(U) = C^\infty(U, L_n)$, pour U ouvert dans \mathbf{R}^m , les morphismes étant les morphismes pairs d'algèbres graduées (cf. [13], p. 66, lemme de Leites et Manin).

Le super-espace-temps est le produit de l'espace-temps usuel W de dimension D avec le produit $(\mathbf{R}^D \otimes L_d^+) \times (S^N \otimes L_d^-)$, pour d assez grand. Le nombre N se nomme *nombre de super-symétrie*. L'espace qui sert pour la dimension impaire est l'espace S d'une représentation spinorielle. La super-algèbre de Lie des *super-translations* est construite à partir d'une application bilinéaire symétrique $\Gamma : S \otimes S \rightarrow W$ envoyant la diagonale dans le cône du futur. L'espace total des super-translations est le super-espace temps et le super-crochet est donné par $\{s_1, s_2\} = -2\Gamma(s_1, s_2)$ sur chaque composante S . La super-algèbre de Lie dite de super-Poincaré est l'extension de l'algèbre de Lie $so(\mathfrak{g})$ du groupe de Lorentz, ou plutôt de $so(\mathfrak{g}) \otimes (L_d^0 \oplus L_d^+)$, par l'algèbre des super-translations. Elle agit naturellement sur le super-espace-temps.

Les théories de jauge s'étendent en théories de «super-Yang-Mills-Higgs», mais aucune de ces théories n'est évidente à construire. Il était encore plus difficile de construire des théories de *supergravité*. C'est pourtant ce qu'ont fait, pour $D = 4$, $N = 1$, d'abord, Freedman et Van Nieuwenhuizen, Deser et Zumino en 1976 (cf. [24]). Puis la «théorie maximale» $D = 4$, $N = 8$, fut construite par E. Cremmer et B. Julia (1979, [10]) en compactifiant, selon une recette Kaluza-Klein super, l'extraordinaire théorie $D = 11$, $N = 1$ de Cremmer, Julia, Scherk (1978, [12]).

Un argument de W. Nahm, joint au parathéorème affirmant que les champs de spin > 2 n'ont pas d'évolution causale, entraîne que, pour la physique, $N = 8$ est un maximum en $D = 4$ et $N = 1$ est un maximum en $D = 11$. D'où l'importance de la supergravité en dimension 11.

En plus de la métrique \mathfrak{g} et de son super-partenaire, qui est un vecteur de spineurs appelé *gravitino*, il faut faire appel à une 3-forme $A_{\mu\nu\rho}$ pour construire cette théorie.

Les théories de supergravité contiennent souvent des champs antisymétriques $F_{\mu\nu}, A_{\mu\nu\rho}, \dots$. Les compactifications à la Kaluza-Klein font apparaître des champs scalaires φ couplés à des formes. Une conséquence des symétries de la fibre compacte et de la dualité de Hodge sur les espaces de solutions est que l'ensemble de ces

champs F, A, φ, \dots forme des sections de fibrés en *espaces riemanniens symétriques* de Cartan G/H de type non compact (voir [7], [21]).

Par exemple, en compactifiant sur un tore \mathbf{T}^k la supergravité $D = 11$, Cremmer et Julia ont détecté comme espace G/H : pour $k = 1$, \mathbf{R}_+^\times ; $k = 2$, $(SL_2(\mathbf{R}) \times \mathbf{R}_+^\times)/U_1$; $k = 3$, $(SL_3(\mathbf{R}) \times SL_2(\mathbf{R}))/SO_3 \times U_1$; $k = 4$, SL_5/SO_5 ; $k = 5$, $SO_{5,5}/SO_5 \times SO_5$; $k = 6$, $E_{6(6)}/Sp_4$; $k = 7$, $E_{7(7)}/SU_8$; $k = 8$, $E_{8(8)}/SO_{16}$, à l'aide d'Ehlers pour le dernier cas.

Pour la supergravité $D = 4$, $N = 4$, couplée avec 22 champs de vecteurs A_μ , on trouve $G = SL_2(\mathbf{R}) \times O(6, 22)$, $H = U_1 \times O_6 \times O_{22}$. Pour $D = 4$, $N = 8$, on a vu $G = E_{7(7)}$, $H = SU_8$. Pour $D = 4$, $N = 2$, il y a plusieurs théories, l'une d'entre elles donne $G = E_{7(-25)}$, $H = E_6 \times U_1$. On rencontre les espaces AI, BDI, EI, EV, EVII, EVIII en formes réelles normales (Helgason).

On trouvera les dualités quantiques de champs et de cordes à l'intérieur d'un sous-groupe discret $G(\mathbf{Z})$ du groupe de symétrie G (cf. [11], [19], [30], [46], [58], [90]).

2. DUALITÉS QUANTIQUES

2.1. En vertu du principe de correspondance de Bohr, une théorie quantique de champs, abrégée en TQC, est associée à un ensemble \mathcal{A} de champs classiques A_1, A_2, \dots , comme des connexions, des métriques ou des spineurs, sur une variété (ou une super-variété) W de dimension D , et à un lagrangien $\mathcal{L}(A_1, A_2, \dots)$ qui est une fonction sur W , dont la valeur en $x \in W$ ne dépend des A_i qu'à travers leurs jets en x .

En mécanique quantique, la densité de probabilité de présence d'une particule est le carré du module d'une fonction d'onde ; en TQC, on suppose que la probabilité de trouver les particules correspondant aux champs A_1, A_2, \dots avec des impulsions-énergies données p_1, p_2, \dots à l'entrée ou à la sortie d'une collision s'obtient à partir de nombres complexes notés $\langle A_1(p_1)A_2(p_2)\dots \rangle$, appelés *amplitudes*. Il est pratique de rassembler les amplitudes pour toutes les valeurs p_1, p_2, \dots possibles en prenant les transformées de Fourier inverses de ces fonctions des p_i : d'où une distribution notée (abusivement) $\langle A_1(x_1)A_2(x_2)\dots \rangle$, et interprétée comme amplitude de probabilité de présence. (Cf. [31], [47], [62], [78].)

On se donne en plus une famille de fonctions des champs, certaines locales, d'autres non $F(A_1, A_2, \dots)$, les quantités observables, et l'on souhaite calculer leurs

amplitudes aussi $\langle F(A) \rangle$. Pour simplifier, nous noterons momentanément A la collection A_1, A_2, \dots . Richard Feynman a proposé de faire comme s'il existait une mesure de volume $\mathcal{D}A$ pour laquelle les amplitudes s'écriraient

$$\langle F(A) \rangle = \frac{1}{Z} \int \mathcal{D}A e^{i \int_W \mathcal{L}(A)} F(A),$$

où $Z = \int \mathcal{D}(A) \exp\left(\frac{i}{\hbar} \int_W \mathcal{L}(A)\right)$ s'appelle la *fonction de partition*. L'intégrale sur W de $\mathcal{L}(A)$ est l'*action classique* $S(A)$. Les intégrales suivant $\mathcal{D}A$ portent le nom d'*intégrales de Feynman*.

A priori, cette écriture n'est qu'un guide formel, un programme à réaliser. La théorie de la *renormalisation* a pour but de réaliser ce programme quand c'est possible (cf. [8], [43], [47], [62], [71]).

La plupart des résultats sont *perturbatifs* : c'est-à-dire que les physiciens ont dû choisir des paramètres, considérés comme petits, comme la valeur du g dans le $\frac{1}{g^2}$ devant le \mathcal{L} de Yang-Mills, ou $\tilde{}$, et procéder ordre par ordre en les puissances de g ou de $\tilde{}$ (voir l'exposé de L. Boutet de Monvel à ce séminaire sur les travaux de Connes et Kreimer). Mais quelques constructions rigoureuses, comme celles des champs conformes, où les calculs numériques impressionnants de E. Wilson sur la discrétisation de la théorie de Yang-Mills justifient l'espoir d'une théorie non-perturbative.

Remarque : G. Segal ([78]) donne une définition plus précise des TQC, inspirée de son axiomatisation (1987) des théories de champs conformes et de celle donnée par Atiyah (1988) pour les théories topologiques des champs. Dans ces derniers cas, la théorie toute renormalisée peut être définie directement. Le point de vue de Segal fait le lien entre le point de vue, dit constructif, d'une vraie mesure euclidienne $\mathcal{D}A \exp(-S(A))$ et la quantification géométrique à la Souriau-Kostant-Weinstein.

Bien qu'on ne puisse pas entrer ici dans les détails, il me semble indispensable d'indiquer en peu de mots l'idée géométrique de la renormalisation ou plutôt comment elle devrait se passer en suivant l'approche de Wilson au début des années 70. En effet, la définition même d'une dualité en TQC et, comme on le verra, en théorie des cordes repose sur les concepts de la renormalisation. Et ne rien savoir de la renormalisation en TQC est à peu près comme ne rien savoir de la notion de limite quand on parle de vitesse et d'équation différentielle ordinaire.

D'abord les pionniers Feynman, Schrödinger, Tomonoga ont découvert qu'il n'y a pas de définition univoque des amplitudes, leur valeur dépend de l'*échelle d'observation*. (Aujourd'hui, l'un des principaux tests de la théorie des interactions fortes

(QCD) est l'évolution de la charge renormalisée en fonction de l'échelle de l'énergie mise en jeu.)

Pour les théories de jauge, les principales recettes sont la «troncature», élimination des A de trop grandes et trop petites impulsions, la «régularisation dimensionnelle», déplacement en dimension non entière par prolongement analytique ou la «discrétisation sur réseaux». Mais, dans un deuxième temps, il faut s'affranchir de ces régularisations arbitraires : des années 40 aux années 70 furent élaborés des procédés pour définir la «partie finie» des amplitudes en adaptant arbitrairement la limite d'un nombre fini de quantités (cf. [47], [62], [13], [71]). Au début des années 70, l'approche de Wilson par le groupe de renormalisation a bien éclairci la situation ([85], [86], [87], [8], [63]). Voici la démarche :

Soit \mathcal{L} un lagrangien de champ classique ; par exemple

$$\mathcal{L}(\varphi) = \frac{a}{2} |\nabla\varphi|^2 - \frac{m^2}{2} \varphi^2 + \frac{\lambda}{24} \varphi^4$$

pour la théorie dite φ^4 d'un champ scalaire (juste une fonction à valeur réelle sur l'espace-temps, modèle simplifié du magnétisme si l'on rend le temps et la masse imaginaires purs). Et soit U_0 la variété des paramètres dont \mathcal{L} dépend ; dans l'exemple, U_0 est l'espace de coordonnées (a, m^2, λ) , a se nomme l'«intensité de champ», m la «masse» et λ la «constante de couplage» ; dans un \mathcal{L} plus général, on trouve des coordonnées naturelles pour U_0 , avec des coefficients de monômes de dérivées des champs ; on les appelle toutes «constantes de couplages» pour simplifier.

L'échelle d'observation des particules (que A_1, A_2, \dots sont censés décrire) est donnée par l'inverse d'une distance ; elle est mesurée par un nombre Λ réel strictement positif, i.e. $\Lambda \in \mathbf{R}_+^\times$. Dans une théorie relativiste quantique, avec des unités où $c = \hbar = 1$, ce nombre Λ fixe aussi l'ordre de grandeur d'une fréquence, d'une énergie, d'une masse ou d'une impulsion.

La théorie achevée de $\mathcal{L}(A)$ doit prévoir des amplitudes $\langle A_1(x_1) A_2(x_2) \cdots A_n(x_n) \rangle$ bien définies à une échelle donnée, pour des valeurs données des constantes de couplages :

une théorie quantique relativiste est décrite par une variété V munie d'un isomorphisme π sur un ouvert de $U_0 \times \mathbf{R}_+^\times$ et d'un feuilletage ρ de dimension 1.

Un point de V représente une collection cohérente d'amplitudes et ρ représente l'évolution des champs quantiques en fonction de Λ . L'application π correspond aux *conditions de renormalisation*, interprétation des constantes de couplages physiques par certaines amplitudes bien choisies.

Par exemple dans la théorie φ^4 , la masse (fois i) donne le pôle des fonctions à deux points $\langle \varphi(p_1)\varphi(p_2) \rangle$ en fonction de l'impulsion $p = p_1 - p_2$, et la constante λ donne la valeur d'une amplitude (convenablement normalisée) à quatre points $\langle \varphi(p_1)\varphi(p_2)\varphi(p_3)\varphi(p_4) \rangle$.

Comme le choix d'une unité d'échelle est indifférent, les feuilles de ρ se projettent sur les trajectoires d'un champ de vecteurs β sur U_0 ; si bien que β sur un ouvert de $U_0 \times \mathbf{R}_+^\times$ peut servir à paramétrer les amplitudes de V . Si l'on maintient (V, π, ρ) , c'est pour indiquer qu'un point de V n'est pas un lagrangien de U_0 , mais un lagrangien en général bien plus compliqué, non polynomial.

Le champ β , ainsi que son action (simple) sur les amplitudes (Gell-Mann et Low), porte le nom de *flot de renormalisation* ou *groupe de renormalisation*. (Selon la convention la plus répandue, le champ β est orienté dans le sens des énergies décroissantes.)

Donnons une idée de la construction de V , ρ et π , ou plutôt racontons l'histoire comme elle devrait se passer en général. Pour fixer les idées, travaillons avec la régularisation tronquant les grandes impulsions ([63]) ; les autres cas se laissent décrire de manière analogue ([43], [86]). On commence par plonger U_0 dans un espace U assez grand de lagrangiens (en général U est de dimension infinie), en supposant que chaque point de U donne une valeur finie de l'intégrale fonctionnelle tronquée au-delà d'une certaine fréquence. En outre, pour tout couple (Λ_0, Λ_1) dans $\mathbf{R}_+^\times \times \mathbf{R}_+^\times$ tel que $\Lambda_0 < \Lambda_1$, on se donne une fibration $\pi_{\Lambda_1, \Lambda_0}$ de U sur U_0 qui impose la valeur de certaines amplitudes tronquées par Λ_0 à l'échelle d'observation Λ_1 . (Dans l'exemple φ^4 , les valeurs de a, m^2, λ sont données par le développement en impulsions de l'intégrale du lagrangien jusqu'aux monômes de degré ≤ 4 en φ .) On suppose en plus que, partant de \mathcal{L} dans U , tronquant l'intégrale de Feynman au-dessus de Λ_0 et calculant la moyenne sur les Λ dans l'intervalle $[\Lambda_1, \Lambda_0]$, on obtient les mêmes amplitudes qu'en tronquant simplement à Λ_1 un autre lagrangien $\mathcal{L}' = R_{\Lambda_1, \Lambda_0}(\mathcal{L})$, quitte à faire un changement de variables $x' = f(x)$ et normaliser autrement les champs $A'_i = Z_i A_i$. Ce \mathcal{L}' est le *potentiel effectif* selon Wilson, associé à \mathcal{L} , à l'échelle Λ_1 et à la troncature au-dessus de Λ_0 .

Le (*semi-*)*groupe de renormalisation* est l'action multiplicative de $]0, 1[$ sur $U \times \mathbf{R}_+^\times$ définie par la formule suivante : $R_\lambda(\mathcal{L}, \Lambda) = (R_{\lambda\Lambda, \Lambda}(\mathcal{L}), \lambda\Lambda)$, pour $\lambda \in]0, 1[$, $\mathcal{L} \in U$ et $\Lambda \in \mathbf{R}_+^\times$. On a $R_{\lambda_2, \lambda_1} \circ R_{\lambda_1, \lambda_0} = R_{\lambda_2, \lambda_0}$ donc $R_{\lambda\mu} = R_\lambda \circ R_\mu$, $\lambda < 1$, $\mu < 1$. La variété $U_{\Lambda_1, \Lambda_0} = R_{\Lambda_1, \Lambda_0}(U_0)$ est censée couper transversalement les fibres de $\pi_{\Lambda_1, \Lambda_0}$. Tout cela est espéré pour que la théorie soit renormalisable mais c'est surtout le point suivant qui constitue un « théorème de renormalisation » : lorsque Λ_0 tend vers

$+\infty$ la variété U_{Λ_1, Λ_0} converge vers une sous-variété U_{Λ_1} de U et la restriction de $\pi_{\Lambda_1, \Lambda_0}$ à U_{Λ_1, Λ_0} converge vers une application π_{Λ_1} de U_{Λ_1} sur U_0 . Pour tout couple $\Lambda_1 < \Lambda_2$, on a $R_{\Lambda_1, \Lambda_2}(U_{\Lambda_2}) = U_{\Lambda_1}$. Les $U_{\Lambda} \times \{\Lambda\}$ forment une sous-variété V de $U \times \mathbf{R}_+^{\times}$ que la collection des π_{Λ} envoie dans $U_0 \times \mathbf{R}_+^{\times}$; le semi-groupe R induit un flot sur V ; c'est l'invariance d'échelle qui assure que ce flot provient d'un champ de vecteurs β sur U_0 . On peut voir V et son feuilletage ρ comme la « vraie théorie » (ou l'ensemble des vraies théories), l'identification à $U_0 \times \mathbf{R}_+^{\times}$ et le champ β sont des souvenirs du lagrangien classique ; ce sont eux qui expriment qu'une théorie classique a été quantifiée.

Attention : Très peu de théories sont renormalisables au-dessus de $U_0 \times \mathbf{R}_+^{\times}$ tout entier ; en particulier, les rôles des Λ petits et des Λ grands sont en général dissymétriques dès le début de la construction. Dire que π projette bien V sur un ouvert où Λ tend vers ∞ signifie que la théorie s'analyse bien en ultraviolet, c'est-à-dire pour des distances de plus en plus petites ; au contraire, au-dessus d'un ouvert où Λ tend vers 0, on décrit le comportement infrarouge, pour les grandes distances.

L'adjectif « renormalisable » est souvent réservé aux théories qui sont renormalisables perturbativement en ultraviolet, cependant en mécanique statistique, qui est une version de la théorie euclidienne des champs quantiques ([68]), la renormalisation se fait surtout en infrarouge comme l'a expliqué K.G. Wilson ([86], [87], [8]). Qu'il existe aussi une théorie perturbative infrarouge en TQC est signalé entre autres par E. Witten (cf. [13], p. 1141).

La renormalisabilité perturbative UV dépend des poids des constantes de couplage g vis-à-vis des dilatations d'espace $x \mapsto t^{-1}x$: la théorie est formellement renormalisable si tous les poids sont positifs ou nuls. Exemple : φ^4 pour $D \leq 4$, mais pas $D > 4$; φ^3 pour $D = 6$; φ^6 pour $D = 3$; φ^k quel que soit k pour $D = 2$; Einstein, $D = 2$; Yang-Mills-Higgs, $D = 4$.

La dynamique β sur U_0 a des conséquences importantes pour la dynamique des champs dans l'espace-temps. *A priori* β pourrait engendrer un flot compliqué, cycles limites, attracteurs étranges, etc., mais une conjecture qui tient bon prétend que β est le gradient d'une fonction. D'ailleurs ce sont surtout les points d'équilibre qui ont retenu l'attention.

Les points de U_0 qui sont des attracteurs de β lorsque $\Lambda \rightarrow \infty$ sont dits *stables dans l'ultraviolet* (ou UV-stables), ceux qui sont attracteurs quand $\Lambda \rightarrow 0$ sont dits *stables dans l'infrarouge* (ou IR-stables). Les limites 0 et ∞ sont les deux racines de l'arc-en-ciel. Lorsqu'un point correspondant à une théorie découplée, sans interaction, est UV-stable, la théorie est dite *asymptotiquement libre* ; c'est alors que la théorie

des perturbations est pleinement justifiée. D'après 't Hooft, c'est le cas de la théorie de Yang-Mills non abélienne qui décrit les interactions fortes (QCD), interactions entre quarks, au contraire de l'électrodynamique quantique (QED) où l'absence d'interaction est plutôt IR-stable (le propre des théories de jauge abéliennes). Donc, aux grands moments de transfert (c'est-à-dire Λ grand), le couplage entre quarks et gluons tend vers zéro ; la théorie perturbative peut faire des prédictions. Le problème est à grande échelle (c'est-à-dire Λ petit). Là, logiquement, la théorie des perturbations ne peut plus dire grand chose.

C'est la volonté de comprendre le comportement infrarouge de ces théories qui mène à la dualité.

Il est temps d'indiquer ce qui s'appelle *dualité* dans une théorie quantique des champs :

Soit $\mathcal{C}, \mathcal{C}'$ deux théories de champs renormalisées avec des algèbres d'observables $\mathcal{A}, \mathcal{A}'$ et des ensembles U_0, U'_0 de lagrangiens classiques (ou de constantes de couplages) ; une *dualité* entre \mathcal{C} et \mathcal{C}' est une application α d'une partie de \mathcal{A} dans une partie de \mathcal{A}' et une application φ d'une partie de $U_0 \times \mathbf{R}_+^\times$ dans une partie de $U'_0 \times \mathbf{R}_+^\times$ envoyant le flot de β sur celui de β' , de telle sorte que les amplitudes se correspondent :

$$\langle \alpha(F(A)) \rangle_{\varphi(u, \Lambda)} = \langle F(A) \rangle_{u, \Lambda}.$$

(Pour une parfaite dualité, on considère la collection complète des amplitudes, mais il peut arriver que la collection se réduise à une fonction de partition.)

Une dualité est un changement de coordonnées ; le plus proche analogue en Analyse classique est la transformation de Fourier $f \mapsto \widehat{f}$. Un exemple d'identité d'amplitude est la formule de Plancherel $\|f\|_2 = C\|\widehat{f}\|_2$. On peut aussi penser aux généralisations en Analyse harmonique non commutative.

Les exemples les plus intéressants échangent une théorie descriptible en UV avec une théorie descriptible en IR .

Les principales dualités sont des « dualités faible/fort », c'est-à-dire qu'elles échangent les couplages faibles de \mathcal{C} avec les couplages forts de \mathcal{C}' et réciproquement.

Par exemple, pour la théorie φ^4 en dimension $D = 2$, le groupe de renormalisation β agissant sur le cadran $\{\lambda \in \mathbf{R}_+, m^2 \in \mathbf{R}_+\}$ possède deux sources $(0, 0)$, $(\infty, 0)$ et un col $(\lambda_c, 0)$. Il existe une dualité de la théorie φ^4 avec elle-même qui préserve le col et échange les sources.

Une discrétisation remarquable de φ^4 en dimension 2 (euclidienne) est le système d'Ising décrivant un ensemble de spins classiques $\sigma_{i,j} \in \{\pm 1\}$ aux sommets du réseau

$\Gamma = \mathbf{Z}^2$, avec une statistique de Boltzmann $\exp(-\frac{1}{kT} H(\sigma))$ (cf. [27], [68]). La théorie duale provient du réseau dual Γ^* dont les sommets sont les centres des faces de Γ , avec les variables de désordre $\mu_{i^*,j^*} \in \mathbf{R}$, qu'on peut définir par des formules (non locales) à partir des corrélations des spins (cf. [68]). À la température exactement critique $T = T_c$, les corrélations des σ sur Γ coïncident avec les corrélations des μ sur Γ^* . On peut prendre la *limite continue critique* des systèmes d'Ising au voisinage de la température critique : en partant d'une boîte finie de côté n , on fait tendre T vers T_c et n vers ∞ de sorte que $n(T - T_c)$ tende vers une limite finie $\lambda_c^1 - \lambda^1$, et le plan continu est rapporté aux coordonnées $(x, y) = \lim \frac{1}{n}(i, j)$. L'ensemble des corrélations tend vers l'ensemble des amplitudes d'une théorie quantique des champs qui peut être identifiée à la théorie φ^4 (cf. Polyakov, Zamolodchikov).

Les formules de la théorie des champs holonomes de Jimbo, Miwa, Sato offrent ainsi l'un des rares cas où il est possible d'explicitier les applications $\alpha : \sigma \mapsto \mu$ et $\varphi : \frac{\lambda}{\lambda_c} \mapsto \frac{\lambda_c}{\lambda}$ de la définition d'une dualité.

Un autre exemple de dualité en dimension 2 a l'intérêt d'échanger magnétisme et électricité : le modèle de Thirring est dual de la théorie de Sine-Gordon. Comme dans le cas d'Ising, il s'agit de théories *complètement intégrables*, c'est-à-dire qu'on connaît des formules fermées pour des amplitudes renormalisées.

2.2. Dualités abéliennes

Quelques exemples simples de dualités quantiques faibles/fortes en petites dimensions exploitent la dualité de Poincaré-Hodge (cf. Witten *in* [13]). On peut traiter ces exemples rigoureusement à partir d'intégrales de Feynman gaussiennes, qui se ramènent à des mesures de Wiener ; leur exposé se fait donc dans le cadre des variétés riemanniennes. Les équations de champs classiques associées sont linéaires.

a) Le plus simple est en dimension $D = 2$, sur une surface de Riemann Σ . (Il est connu comme «équivalence $R \longleftrightarrow 1/R$ ».) On considère un champ scalaire unitaire, c'est-à-dire une fonction $\varphi : \Sigma \rightarrow S^1 = \mathbf{R}/2\pi\mathbf{Z}$. (On note aussi φ le relèvement aux revêtements universels $\tilde{\Sigma} \rightarrow \mathbf{R}$.) Le lagrangien est $\mathcal{L}(\varphi) = \frac{R^2}{4\pi} d\varphi \wedge *d\varphi$. L'équation classique est $\Delta\varphi = 0$, dont les solutions sont les fonctions harmoniques. Localement, si φ est harmonique, il existe une fonction σ à valeurs réelles telle que $*d\sigma = d\varphi$; on a aussi $\Delta\sigma = 0$. Et, à une constante additive près, σ et φ se déterminent l'une l'autre, $*d\varphi = -d\sigma$.

Afin de réaliser au mieux la dualité sous la forme d'une transformation de Fourier, on introduit un espace de champs φ plus grand ; celui de toutes les sections de tous les fibrés en cercle triviaux L au-dessus de Σ . Et à côté de φ et σ , on considère aussi

les champs ∇ , connexions unitaires sur L ; leur courbure est notée F . Le lagrangien complet est

$$\mathcal{L}(\varphi, \nabla, \sigma) = \frac{R^2}{4\pi} \nabla\varphi \wedge *\nabla\varphi - \frac{i}{2\pi} \sigma F.$$

Son équation d'Euler-Lagrange donne $F = 0$. Localement l'annulation de F permet d'utiliser ∇ pour trivialisier L et on retrouve dans cette trivialisiation l'équation $\Delta\varphi = 0$. L'intégrale fonctionnelle de Feynman justifie cette trivialisiation globalement : en intégrant d'abord sur σ , on trouve $F = 0$, mais en intégrant ensuite sur les classes d'équivalence de jauge de ∇ , on trouve que ∇ ne peut avoir d'holonomie non triviale.

Surtout, en intégrant d'abord sur φ et ∇ sans toucher à σ (ce qui se fait avec la méthode des fantômes de Faddeev-Popov, cf. [33]), on obtient la formule suivante :

$$\int \mathcal{D}\varphi e^{-\frac{R^2}{4\pi} \int d\varphi \wedge *d\varphi} = \left(\frac{R}{\pi}\right)^{\chi(\Sigma)} \int \mathcal{D}\sigma e^{-\frac{1}{4\pi R^2} \int d\sigma \wedge *d\sigma},$$

qui généralise le calcul de la transformée de Fourier des gaussiennes en dimension finie.

Cette formule recouvre la propriété modulaire des fonctions thêta, cf. le cours de K. Gawędzki dans [13].

Mais les calculs ne s'arrêtent pas aux fonctions de partitions : par exemple, (mêmes références) l'observable $e^{i(\varphi(P)-\varphi(Q))}$ pour deux points P et Q de Σ , correspond par dualité à la fonction de partition des champs σ , tels que la forme $d\sigma$ s'étend à $\Sigma \setminus (P \cup Q)$ en une forme fermée avec résidus 1 en P et -1 en Q . On voit que le dual d'un observable local peut ne pas être local. Il peut arriver aussi qu'il le soit : par exemple le dual de $d\varphi$ est $\frac{1}{iR^2} \star d\sigma$.

b) En dimension 3, la dualité échange les champs scalaires $\varphi : \Sigma \rightarrow S^1$ comme ci-dessus, avec les 1-formes A modulo équivalence de jauge $A \sim A + df$. La théorie duale de celle des fonctions harmoniques est la théorie magnétostatique de Maxwell. Le champ qui couple les deux, comme ∇ au a), est une connexion ∇_B sur un fibré L_B dont φ est une section :

$$\mathcal{L}(\varphi, \nabla_B, A) = \frac{R^2}{4\pi} \nabla_B\varphi \wedge *\nabla_B\varphi - \frac{i}{2\pi} A \wedge F_B.$$

Pour les fonctions de partition, on a

$$\int \mathcal{D}\varphi e^{-\frac{R^2}{4\pi} \int d\varphi \wedge *d\varphi} = C \sum_{L_A} \int \mathcal{D}A e^{-\frac{1}{4\pi R^2} \int F_A \wedge *F_A}.$$

L'observable « boucle de Wilson » $e^{i\mu} \oint_C A$, donnée par l'holonomie de A sur un lacet C dans M^3 , correspond aux sections φ d'un fibré L_B singulières le long de C , avec une monodromie prescrite $e^{2\pi i\mu}$.

c) Sur une variété W de dimension 4, la dualité abélienne, ou gaussienne, échange une connexion ∇_A sur un fibré en cercle L et une connexion ∇_C sur un fibré N , toutes les deux en théorie électromagnétique de Maxwell. C'est l'étoile de Hodge du § 1, $F_C = *F_A$, qui fournit la dualité. Le champ intermédiaire est une 2-forme ordinaire G .

't Hooft et Polyakov ont montré qu'on gagne beaucoup en ajoutant le terme topologique $\frac{i\theta}{4\pi^2} F_A \wedge F_A$ au lagrangien, qui devient :

$$\mathcal{L}(A) = \frac{1}{2e^2} F_A \wedge *F_A + \frac{i\theta}{4\pi^2} F_A \wedge F_A.$$

L'action associée, si W est compacte, est modifiée par un élément de $2\pi i\mathbf{Z}$. En posant $\tau = \frac{\theta}{\pi} + i\frac{2\pi}{e^2}$, et en introduisant les composantes autoduales et antiautoduales $F_{\pm} = \frac{1}{2}(F_A \pm *F_A)$, on a

$$\mathcal{L}(A) = \left(\frac{i\bar{\tau}}{4\pi} \|F_+\|^2 - \frac{i\tau}{4\pi} \|F_-\|^2 \right) \text{Vol}_3.$$

Mais pour établir une dualité, on introduit à côté de A une 2-forme G et une connexion C sur un fibré en droite N , et on considère le lagrangien

$$\mathcal{L}(A, G, C) = \left(\frac{i\bar{\tau}}{4\pi} \|\mathcal{F}_+\|^2 - \frac{i\tau}{4\pi} \|\mathcal{F}_-\|^2 \right) \text{Vol}_3 - \frac{i}{2\pi} F_C \wedge G,$$

où $\mathcal{F} = F_A - G$. La dualité s'énonce ainsi : la fonction de partition du côté A avec la constante τ est égale (à une constante multiplicative près) à celle du côté C , mais avec la constante $-1/\tau$. D'après E. Witten :

$$Z(\tau) = \tau^{-\frac{(\chi+\sigma)}{4}} \bar{\tau}^{-\frac{(\chi-\sigma)}{4}} Z\left(-\frac{1}{\tau}\right),$$

où χ et σ sont les nombres d'Euler et la signature de W .

Comme tout est manifestement invariant (sur W compacte) par le changement $\theta \mapsto \theta + 2\pi$ (et invariant par $\theta \mapsto \theta + \pi$ si W est spinorielle), rien ne change par $\tau \mapsto \tau + 2$ (resp. $\tau \mapsto \tau + 1$), et Z est une forme modulaire.

Lorsque $W = M \times \mathbf{R}$, on peut écrire la théorie sous forme hamiltonienne, le moment conjugué π_A de A est l'étoile de Hodge de

$$F_A^\vee = 2\pi i \frac{\partial \mathcal{L}(A)}{\partial F_A} = \frac{2\pi i}{e^2} *F_A - \frac{\theta}{\pi} F_A.$$

La dualité est la transformation de Legendre $F_A^\vee = F_C$, $F_C^\vee = -F_A$. Elle échange les charges électriques et magnétiques ; pour $\Sigma \subset M^3$:

$$Q_e(\Sigma) = \int_{\Sigma} \frac{F_A^\vee}{2\pi} = \int_{\Sigma} \frac{F_C}{2\pi} \quad , \quad Q_m(\Sigma) = \int_{\Sigma} \frac{F_A}{2\pi} = - \int_{\Sigma} \frac{F_C^\vee}{2\pi}.$$

Le groupe $SL_2(\mathbf{Z})$ agit alors sur l'espace de Hilbert des états associé à M .

2.3. Dualités non abéliennes

Le premier exemple de dualité non linéaire est dû à A.M. Polyakov (1975) : il s'agissait d'une théorie de jauge en dimension 3, avec une connexion ∇ sur un fibré en groupe SO_3 et un champ vectoriel \vec{s} à valeur dans la représentation standard \mathbf{R}^3 . Un mécanisme de «brisure de symétrie» (cf. ci-dessous) fait apparaître un champ de jauge abélien A ; le champ principal de la théorie duale φ provient de solutions particulières (\vec{s}, ∇) , les «monopôles» de la théorie initiale, à travers l'équation $*d\varphi = F_A$. Ce champ acquiert un potentiel transcendant $\cos(2\varphi)$, renormalisable en infrarouge mais pas en ultraviolet (cf. [13], tome 2).

Les monopôles de 't Hooft et Polyakov sont des solutions particulières des équations de Yang-Mills-Higgs, version classique du «modèle standard» pour les particules en «interaction faible» ([42]). La théorie standard n'est pas purement celle de Yang et Mills : il y a bien un groupe de jauge G (ici $SU_2 \times U_1$ ou $SU_3 \times SU_2 \times U_1$ si l'on veut tenir compte des quarks et des gluons) et une connexion ∇ , mais il y a aussi au moins un champ φ , appelé *champ de Higgs*, section du fibré adjoint en algèbre de Lie \mathfrak{g} . Le lagrangien de Yang-Mills-Higgs est

$$\mathcal{L}(A, \varphi) = \frac{1}{4g^2} (\text{Tr}(F_A \wedge *F_A) + |d_A\varphi|^2 + \lambda(|\varphi|^2 - a^2)^2).$$

Cependant, les monopôles les plus étudiés (nous ne parlerons que de ceux-là) sont des solutions statiques des équations de Bogomolny en dimension 3

$$F_A = *d_A\varphi.$$

Ce sont les équations dérivées de \mathcal{L} dans la limite $\lambda \rightarrow 0$, introduite par Prasad et Sommerfield (1975). Pour que ces monopôles correspondent à des solutions de Yang-Mills-Higgs dans \mathbf{R}^4 d'énergie finie, on impose qu'à l'infini dans \mathbf{R}^3 (à temps fixé), on ait dans chaque direction $|\varphi| = a$; si bien que la configuration à l'infini du champ de Higgs φ fournit un invariant topologique, une classe caractéristique. Par exemple, pour le groupe de jauge $G = SU_2$, φ définit une application de la sphère à l'infini de \mathbf{R}^3 dans la sphère de rayon a de l'algèbre de Lie \mathfrak{so}_3 de SO_3 ; son degré est un nombre entier k qui s'interprète comme une charge magnétique, $b = 2\pi k g^{-1}$. Autrement dit, les plans perpendiculaires aux directions de φ définissent, en dehors

d'un compact de \mathbf{R}^3 , un fibré en plans, de groupe structural SO_2 ; le nombre k est sa classe d'Euler.

On dit alors que φ a *spontanément brisé* le groupe de jauge SU_2 sur le sous-groupe $H = SO_2$. (Attention : à proprement parler aucune symétrie n'est brisée, la symétrie de jauge est intacte.) Suivant B. Julia et A. Zee (1974, [50]) et M.K. Prasad et C.M. Sommerfield (1975, [69]), il existe aussi des *dyons*, devinés par J. Schwinger, possédant à la fois une charge magnétique et une charge électrique ; alors la condition qui remplace $bg \in 2\pi\mathbf{Z}$ s'écrit $e_n b_m - e_m b_n \in 2\pi\mathbf{Z}$, pour chaque paire de particules n et m de charges électriques e_n, e_m et magnétiques b_n, b_m (J. Schwinger, [75], D. Zwanziger, [98], 1968). C'est la condition de quantification généralisant celle que Dirac avait trouvée pour le monopôle magnétique.

Polyakov et 't Hooft (1975, 1976, cf. [42], [68]) évaluèrent l'effet des instantons et des monopôles dans la théorie des perturbations : si g est la constante de couplage de jauge, les contributions aux amplitudes calculées par intégrales de Feynman viennent avec un facteur de l'ordre de $\exp(-8\pi^2/g^2)$. 't Hooft et Polyakov remarquèrent aussi l'importance du facteur $\exp(i\theta)$ qui intervient devant les amplitudes. Cet angle θ n'apparaît que dans la combinaison complexe $(\theta/2\pi) + i(4\pi/g^2)$; il vient de la possibilité d'ajouter un terme $\frac{g^2\theta}{32\pi^2}\text{Tr}(F_A^2)$ au lagrangien de Yang-Mills-Higgs. Ce terme est de nature topologique, il ne modifie pas la théorie classique, mais s'avère fondamental pour la théorie quantique.

Dans une théorie de jauge en dimension 4, pour $G = SU_n$ avec des champs de Higgs, mais sans brisure de symétrie, après un fixage de jauge (c'est-à-dire une trivialisatation unitaire des fibrés), 't Hooft montre que la théorie renferme des champs vectoriels avec ou sans masse, des champs scalaires et des SU_2 -monopôles. En couplage fort, on espère que la théorie se réexprime, dualement, en séries de $1/g^2$ à partir de monopôles. Les physiciens attendent un comportement statistique analogue à celui des supraconducteurs, mais avec les rôles de l'électricité et du magnétisme échangé : plus précisément, ils suspectent un comportement en troupeau de moutons des paires de monopôles. Or il est établi que les champs magnétiques, qui ne sont pas piégés dans les supraconducteurs, en sont rejetés ; c'est ce qui s'appelle l'«effet Meissner». Dès lors un effet Meissner dual exclurait, ou confinerait, les courants électriques. C'est le scénario de confinement des quarks que 't Hooft proposa.

Justifier ce scénario est un des principaux problèmes de la physique des particules.

't Hooft détailla même la description d'un portrait de plusieurs phases pour la chromodynamique quantique renormalisée ; au moins trois phases : le mode du confinement, dual du mode de Higgs et une «phase de Coulomb» autoduale où

l'électricité et le magnétisme sont symétriques, qui permet des forces de longue portée. La théorie duale pour une théorie de jauge pour SU_n a pour groupe de jauge le groupe adjoint PSU_n .

Plus généralement, on peut tenter l'analyse des théories de Yang-Mills avec tout groupe de jauge compact G et des champs de Higgs ramenant spontanément la symétrie à un sous-groupe H . Goddard, Nuyts et Olive ([32]) ont démontré que la «charge magnétique» des monopôles de 't Hooft et Polyakov se généralise naturellement en une représentation d'un groupe de Lie compact H^\vee qu'on peut construire explicitement à partir de H , le «groupe dual». On a $(H^\vee)^\vee = H$. Par exemple, $U_1^\vee = U_1$, alors la charge est un entier. La conjecture est qu'il existe une vraie dualité faible/fort échangeant H et H^\vee . Les nombres quantiques de Noether (électriques) et topologiques (magnétiques) s'échangeraient. Les monopôles pour H formeraient une théorie de jauge pour H^\vee .

Il est très remarquable, mais aussi rarement signalé (A.J. Schwartz l'a pourtant fait), que le groupe dual H^\vee est exactement celui que R.P. Langlands avait défini à la fin des années 60 pour les besoins de l'arithmétique et de la théorie des groupes ([53]). Par exemple, $SU_n^\vee = PSU_n = SU_n/(\mathbf{Z}/n\mathbf{Z})$. La dualité induit la transposition des matrices de Cartan, ainsi les séries A, D, E, F, G sont auto-duales mais B et C s'échangent. Le groupe dual de Langlands intervient dans un vaste ensemble de conjectures reliant l'analyse harmonique, les formes automorphes et la théorie des nombres, pour donner des lois de réciprocité nouvelles. (Langlands parle de groupes réductifs complexes ; ce sont les complexifications des groupes compacts. Il travaille aussi avec des adèles et pas seulement sur \mathbf{R} ou \mathbf{C} .) Il serait merveilleux qu'une même théorie mathématique à venir éclaire à la fois la vraie nature des quarks et la correspondance de Langlands (dont un cas particulier traité par A. Wiles a entraîné la démonstration de la conjecture de Fermat).

2.4. En considérant des théories de jauge avec suffisamment de super-symétrie, le scénario du confinement a pu être établi presque entièrement.

Dans les théories de super-Yang-Mills-Higgs, les variables sont une super-connexion \mathcal{A} sur un G -fibré et un super-champ vectoriel Φ dont on note Φ_i , $i \in I$, les super-fonctions coordonnées (cf. §1.4). On décompose $\Phi = \varphi + \psi + F$ suivant L_d^0 , L_d^- , L_d^+ . Le champ φ est à valeurs dans un multiple de la représentation adjointe \mathfrak{g} de G . Le lagrangien est somme de deux termes $\mathcal{F}(\mathcal{A}, \Phi)$ et $\mathcal{K}(\Phi, \bar{\Phi})$. Le premier, \mathcal{F} , est une fonction holomorphe des Φ_i appelée super-potentiel, alors que le second, \mathcal{K} , est la fonction génératrice d'une métrique kählérienne $g_{i\bar{j}} = \partial^2 K / \partial \Phi_i \partial \bar{\Phi}_j$ (cf. [84],[33]).

Le principal paramètre dont dépend la théorie est issu du coefficient u du terme quadratique en Φ dans \mathcal{F} ; après renormalisation, $u = C\langle\text{Tr}\varphi^2\rangle$. Il s'appelle «vide» et il est responsable des brisures de symétrie par un mécanisme de Higgs.

Pour $G = SU_2$, spontanément brisé sur U_1 , on a $|\varphi| = a$ à l'infini dans \mathbf{R}^3 , alors $u = a^2$.

Dans \mathcal{F} , en facteur du terme quadratique en les super-courbures, on trouve

$$\tau(\Phi) = \frac{\theta}{2\pi} + \frac{4\pi i}{g^2},$$

avec $\theta, g \in \mathbf{R}$. Lorsque $N = 2$, il existe des relations entre τ et \mathcal{K} , par exemple la métrique kählérienne sur la droite complexe engendrée par a s'écrit $ds^2 = \frac{1}{2\pi}\Im(\tau da d\bar{a})$.

Soit $\tilde{\mathcal{F}}$ le super-potential renormalisé ; on introduit la «variable duale» de a , en posant $a_D = \partial\tilde{\mathcal{F}}/\partial a$. On a donc $\tau = da_D/da$.

Le groupe de renormalisation β s'exprime par la dépendance de a , a_D , o , en fonction de u et de la fréquence Λ . Seiberg et Witten ([76], [77]) démontrent l'existence d'une série

$$\tau(u, \Lambda) = \frac{i}{\pi} \ell n \frac{u}{\Lambda^2} + \sum_0^{\infty} F_n \frac{\Lambda^n}{u^{2n}}.$$

Donc, quand $\Lambda \rightarrow \infty$, on a $a_D \sim 2i(a/\pi) \ln(a/\Lambda) + ia/\pi$.

Ainsi la monodromie à l'infini en u de a_D est non triviale, il doit donc exister une singularité au moins dans le plan des u . Seiberg et Witten montrent que la positivité de ds^2 force l'existence d'un groupe de monodromie global non abélien, d'où l'existence d'au moins deux singularités $\pm u_c$. Une bonne normalisation des singularités pour rendre compte de la limite $\Lambda \rightarrow \infty$ est $u_c = \Lambda^2$.

Toujours selon Seiberg et Witten, a et a_D sont des périodes de la forme différentielle holomorphe $\frac{1}{2\pi}(x-u)\frac{dx}{y}$ sur la courbe elliptique d'équation $y^2 = (x^2 - \Lambda^4)(x-u)$.

Plus généralement, pour un groupe de jauge compact simple G de rang r , la théorie de champs renormalisée est décrite par une courbe complexe Σ de genre r variant au-dessus d'un tore de dimension r (la «branche de Coulomb»). Le groupe de renormalisation fait évoluer les couplages, interprétés comme des périodes d'une forme holomorphe, en fonction de Λ et des coordonnées sur le tore, suivant le système complètement intégrable associé à G par Hitchin (cf. [13]).

La limite de basse énergie en un point singulier u_c est la théorie duale infrarouge de la théorie de jauge $N = 2$. Elle s'appelle «théorie de Seiberg-Witten». Ses champs fondamentaux, un champ de spineurs et une connexion, sont des dégénérescences de

masse 0 d'états quantiques massifs très particuliers qui existent pour tout $u \neq \pm u_c$: les *états BPS*.

En effet, on trouve dans la théorie super-YMH, $N = 2$, des super-monopôles magnétiques et des super-dyons de charge électrique Q_e et de charge magnétique Q_m , qui sont invariants par la moitié des super-symétries ; on les appelle états BPS pour Bogomolny, Prasad, Sommerfield. Leur masse satisfait à $M^2 = C^2(Q_e^2 + Q_m^2)$. Dans cette théorie, on n'a qu'une action projective de l'algèbre de super-Poincaré $N = 2$ ([60], [97]), c'est-à-dire une action linéaire d'une extension centrale de la super-algèbre : pour un super-dyon, la charge centrale est égale à $Z = aN_e + a_D N_m$, où $N_e = Q_e - \frac{\theta}{2\pi} Q_m$ et $N_m = Q_m$ sont des nombres entiers en vertu de l'effet θ de Witten ([88]). Ces nombres Z forment un réseau dans \mathbf{C} lorsque $u \neq \pm u_c$. Le groupe de monodromie de (a, a_D) agit sur Z ; il est d'indice fini dans $SL_2(\mathbf{Z})$.

C'est seulement pour $N = 4$, avec deux super-symétries de plus, que A. Sen [79] a pu démontrer l'existence d'états BPS pour N_e, N_m premiers entre eux quelconques, établissant ainsi la conjecture de dualité $N = 4$ de Montonen et Olive [57].

Notons $N_m = r$ et $N_e = s$. L'espace des modules de monopôles de charge magnétique r pour les équations de Bogomolny sur \mathbf{R}^3 s'écrit

$$\mathbf{R}^3 \times (S^1 \times \mathcal{M}_r^0) / (\mathbf{Z}/r\mathbf{Z}),$$

où \mathcal{M}_r^0 est une variété hyperkählérienne de dimension $4r - 4$, l'espace des monopôles réduits, sur laquelle le sous-groupe $\mathbf{Z}/r\mathbf{Z}$ de S^1 agit librement ([3]). D'après Witten, Manton et Sen ([79]), les états quantiques de la théorie $N = 4$ sont des formes différentielles sur \mathcal{M}_r^0 ; la conjecture de Montonen et Olive se ramenait à prouver l'existence d'une unique forme harmonique L^2 (de degré $2r - 2$) sur laquelle l'action de $t \in \mathbf{Z}/r\mathbf{Z}$ est la multiplication par $\exp(-2\pi i t s / r)$.

A. Sen a d'abord traité le cas de $r = 2$, les autres cas ont été résolus peu après par Segal, Selby, Parrati. Pour G simple compact quelconque, cf. [26], [29], [54].

3. DUALITÉS DE CORDES

3.1. Depuis plus de 20 ans, la théorie des cordes, et surtout celle des supercordes, était considérée comme la meilleure chance pour unifier toutes les interactions avec la gravitation. Elle avait éclipsé les tentatives de champs unitaires, en particulier la supergravité $D = 11$. Mais depuis 6 ou 7 ans, à la suite des découvertes de dualités étranges dans le monde des cordes (Schwarz, Sen, Duff, Hull-Townsend, Witten,...),

le paysage a changé ; une mystérieuse théorie M en dimension 11 (ou peut-être F en dimension 12), rassemblant des cordes qui se propagent, mais aussi des membranes (surfaces dans l'espace, variétés de dimension 3 dans l'espace-temps), des trous noirs et des variétés spéciales de toutes dimensions, s'est dévoilée. Ses limites les plus intéressantes sont des supercordes, mais aussi une de ses limites est la supergravité en dimension 11. Ainsi c'est difficile de dire qui a mangé l'autre, de la corde ou de la supergravité.

L'idée des cordes est d'attribuer aux particules très énergétiques la forme d'une ficelle d'une longueur de l'ordre de 10^{-32} cm, et de «quantifier deux fois» les mouvements et les accidents de cette ficelle relativiste dans l'espace-temps de Minkowski pour retrouver, sans contradiction, toute la physique connue. D'où vient cette idée ? Quel est son rapport avec la gravitation ? Pour le comprendre, reportons-nous à la plus naïve des questions qui se pose en théorie quantique des champs : comment un champ devient-il une particule ? En effet, dans quelle approximation passe-t-on des amplitudes des connexions et des sections de fibrés spin à des particules ayant des trajectoires comme des projectiles ? La relation se fait en deux temps, en «deux quantifications» : une première quantification mène de l'action du point matériel (sans masse) dans l'espace de Minkowski à l'équation de Klein-Gordon $d * d\varphi = 0$; c'est l'équation de Schrödinger de la particule. La deuxième quantification fait passer à une assemblée d'oscillateurs, pour tenir compte des créations et annihilations.

Afin de rendre compte des interactions, il est possible d'introduire un potentiel, par exemple φ^4 , ce qui donne une théorie de champ scalaire. Pour les particules chargées, les photons ou les super-points, les interactions peuvent s'expliquer en couplant les équations de Dirac, de Maxwell et d'Einstein. De même les quarks, les bosons W et Z , les gluons sont soumis au couplage de Dirac et de Yang-Mills. L'exigence de la renormalisation sélectionne sévèrement le choix des interactions mais laisse encore un goût d'arbitraire.

Avec les cordes, la vision des interactions est changée : dans une limite où $\tilde{\alpha}'$ serait négligeable, mais pas la longueur typique des cordes, mesurée par la constante α' , inverse de la *tension de la corde*, on aurait affaire à des objets possédant une dimension d'espace et une de temps. Les trajectoires sont remplacées par des cylindres. Et, pour décrire les *interactions* de cordes, il suffit de remplacer le cylindre par des surfaces dont le bord a plus de deux composantes.

Le nombre n de trous compte le nombre de cordes à l'entrée et à la sortie de l'accident. On montre que le genre k de la surface (son nombre d'anses) mesure l'ordre des *corrections quantiques* dans les amplitudes de cordes quantiques.

La deuxième bonne nouvelle apportée par les cordes est la présence naturelle de particules véhiculant des interactions gravitationnelles : les *gravitons*. La première quantification des théories des cordes est une théorie de champ bidimensionnelle qui incorpore comme condition de renormalisation, au premier ordre en α' , la théorie de la gravitation d'Einstein classique à côté de la théorie de Yang-Mills classique dans l'espace ambiant. D'ailleurs la constante α' vaut $\sim \frac{G}{c^3}$, où G est la constante de gravitation universelle de Newton.

Mais aux ordres plus élevés en α' des corrections arrivent qui pourraient bien régulariser la gravitation aux grandes énergies et guider sa quantification. Notons que les théories de cordes conservent un sens, au moins perturbatif, si l'on remplace l'espace-temps plat de Minkowski par des variétés lorentziennes munies de métriques d'Einstein, i.e. solutions des équations d'Einstein classiques.

Une autre «prédiction» est la *super-symétrie* : sans elle, c'est-à-dire sans passer aux *supercordes*, des particules aberrantes violant les principes de la relativité, se propageant plus vite que les photons, s'imposeraient subrepticement.

Enfin, et surtout, déjà la première quantification des cordes et supercordes est une théorie quantique de champ très exigeante : si l'on veut y maintenir, au niveau quantique, les symétries de la corde, ou supercorde classique (à savoir le groupe de (super) Poincaré ambiant, les difféomorphismes de la surface source, les dilatations locales de sa métrique (transformations de Weyl), les super-symétries de super-surface), il n'y a que 5 possibilités, toutes en dimension 10 d'espace-temps. Celles-ci portent les noms suivants : Type I, Type IIA, Type IIB, Hétérotique SO_{32} , Hétérotique $E_8 \times E_8$.

3.2. Suivant Polyakov (cf. [68], [2], [5], [33]), une théorie de cordes à valeurs dans la variété W , de dimension D , équipée d'une métrique lorentzienne $(D - 1, 1)$, est composée de deux champs en dimension 2 : une métrique sur Σ (de signature $(1, 1)$) et une application X de Σ dans W ; le lagrangien est l'énergie $-\frac{1}{2\alpha'} \|\nabla X\|^2$.

Lorsque W est l'espace plat de Minkowski, un choix convenable des coordonnées ramène X à $D - 2$ solutions de l'équation des ondes (les coordonnées transverses à $\Sigma = S^1 \times \mathbf{R}$ ou $[0, 1] \times \mathbf{R}$), et deux constantes (cf. [33], [13]). La première quantification donne une algèbre de Weyl et sa représentation de Fock, et les symétries, la reparamétrisation des branches du cône de lumière fournissent une représentation ρ de l'algèbre de Virasoro $\text{Vir} \otimes \overline{\text{Vir}}$ dans l'espace de Fock.

Si N est l'opérateur de nombre, la masse d'un état pur est donnée par une formule $\alpha' M^2 = C_0 + N$, où C_0 est une constante à déterminer. L'invariance lorentzienne externe (de W) force $C_0 = -1$ et l'invariance conforme (de Σ) force $D = 26$. Alors, en décomposant la représentation ρ en facteurs irréductibles (les champs primaires) (cf. [27]), on peut identifier les vecteurs de masse 0 ; ils forment diverses représentations irréductibles de $SO(D-1, 1)$: on y trouve les espaces de sections des fibrés en $S^2, \Lambda^2, \mathbf{R}$, c'est-à-dire les 2-formes symétriques, $G_{\mu\nu}$, les 2-formes antisymétriques $B_{\mu\nu}$ et un champ scalaire Φ .

Les conditions de renormalisation du premier ordre en α' s'expriment comme des équations aux dérivées partielles sur $G_{\mu\nu}, B_{\mu\nu}, \Phi$; c'est la surprise : on tombe sur une perturbation des équations d'Einstein, celle qui dérive du lagrangien :

$$\mathcal{L}(G, B, \Phi) = \frac{1}{2\kappa^2} e^{-\Phi} (R_G + 4|\nabla\Phi|^2 - \frac{1}{12}|dB|^2).$$

On voit que $\alpha' \sim \kappa^2 e^{2\langle\Phi\rangle}$, donc Φ , nommé le *dilaton*, fixe la « constante de couplage » α' . C'est la morale des cordes : toutes les constantes y deviennent des champs.

La présence d'une particule de masse imaginaire (le « vertex » ordinaire) rend la théorie peu physique, mais cela ne l'empêche pas d'être riche de propositions mathématiques.

Avec les *supercordes*, les physiciens semblent avoir eu plus de chance, et ils ont fait un cadeau encore plus beau aux mathématiciens.

Il y a *a priori* deux définitions différentes : celle dite de Green et Schwarz, où une super-surface de Riemann s'envoie dans un espace de super-Minkowski modelé sur une puissance du fibré des spineurs S^N , et celle dite de Neveu-Schwarz-Ramond, où la super-surface s'envoie dans le super-espace modelé sur le fibré tangent $T(W)$. Mais, compte tenu des symétries, les deux approches s'avèrent équivalentes.

Pour maintenir les invariances super-Poincaré et super-conformes, il faut $D = 10$; en cette dimension, l'espace des spineurs de Dirac est de dimension 32. À côté des vecteurs de l'espace transverse à la corde, on trouve les spineurs de l'espace de dimension 8, S^+, S^- tous deux de dimension 8 aussi. (La *trialité* est l'équivalence par automorphisme extérieur des trois représentations.)

Pour décrire les solutions classiques, on doit tenir compte de plusieurs possibilités pour les ondes de spineurs $\psi(z, \bar{z}, \theta, \bar{\theta}) \in S^N$

- 1) elles se propagent soit à droite soit à gauche,
- 2) elles appartiennent à l'un ou l'autre des deux fibrés spin sur S^1 .

On met un indice + pour les ondes se déplaçant vers la droite, c'est-à-dire les fonctions de z et un indice - pour celles qui se déplacent vers la gauche, c'est-à-dire

les fonctions de \bar{z} . La seconde distinction sépare l'espace des multiples de $(zdz)^{\frac{1}{2}}$, noté R (comme Ramond) de l'espace des multiples de $(dz)^{\frac{1}{2}}$, noté NS (comme Neveu-Schwarz).

Toutes les possibilités sans tachyons ont été classifiées (c'est le contenu du Théorème de projection GSO, Goddard, Sent, Olive).

Le type I, avec une super-symétrie $N = 1$ sur W (mais toujours 32 pour Σ). Ses états de masse 0, déterminés par la représentation des algèbres super-Virasoro, appelés *champs effectifs*, sont une métrique $G_{\mu\nu}$ ($\text{NS} \otimes \text{NS}$), un dilaton Φ ($\text{NS} \otimes \text{NS}$), une 2-forme anti-symétrique $B_{\mu\nu}$ ($R \otimes R$), un champ de covecteur-spineur ξ_μ^α , le *gravitino*, et un champ de spineurs λ_α , le *dilatino*. C'est tout pour les cordes fermées. Mais, en type I, on permet des cordes ouvertes, il s'ajoute alors un champ de jauge A_μ pour le groupe SO_{32} et un champ de gaugino ψ^σ .

Le type IIA, $N = 2$, champs effectifs : G, B, Φ , tous trois $\text{NS}_+ \otimes \text{NS}_+$, ξ_1, λ_1 ($R_+ \otimes \text{NS}_+$), ξ_2, λ_2 ($\text{NS}_+ \otimes R_-$), plus une 1-forme A_μ et une 3-forme $A_{\mu\nu\rho}$, de type $R_+ \otimes R_-$.

Le type IIB (dit *chiral*), $N = 2$, champs effectifs : G, B, Φ ($\text{NS}_+ \otimes \text{NS}_+$), ξ_1, λ_1 ($\text{NS}_+ \otimes R_-$), ξ_2, λ_2 ($\text{NS}_+ \otimes R_+$), mais trois formes paires (toutes de type $R_+ \otimes R_+$) : A^0 l'*axion*, $A_{\mu\nu}^2$ une 2-forme, et une 4-forme $A_{\mu\nu\rho\sigma}$, avec la restriction $dA = *dA$.

En plus de ces trois possibilités, fut inventée la théorie des *cordes hétérotiques* ; une mise en forme géométrique fait appel à une notion de super-feuilletage dans une super-variété de dimension $4|n$ (cf. [51]) ; elle correspond à une corde (bosonique) pour les champs allant à droite, et à une supercorde (fermionique) pour les champs allant à gauche. Elle vient sous deux formes, avec des champs de Yang-Mills supplémentaires, soit SO_{32} , soit $E_8 \times E_8$. Ses champs effectifs sont G, B, Φ, A (de jauge), uniquement NS, et χ, λ, ψ , de type R, comme pour le type I.

Lorsqu'on pratique une réduction à la Kaluza-Klein pour descendre en dimension < 10 , les cinq types n'en donnent plus que trois : I, II, Het.

Les lagrangiens effectifs, qui sont les limites effectives de basses énergies des supercordes, ont été construits : ils donnent les théories de supergravité à 10 dimensions.

Par exemple, pour le type I, la partie bosonique de l'action est

$$S_I = \frac{1}{16\pi\alpha'^4} \int \text{Vol}_{10} \left(e^{-\Phi} (R_G + |\nabla\Phi|^2) - \frac{1}{12} |H_3|^2 - \frac{1}{4} e^{-\frac{\Phi}{2}} |F_2|^2 \right),$$

où F_2 est la courbure du champ de Yang-Mills SO_{32} et $H_3 = dB_2$.

Pour les théories hétérotiques, avec mes mêmes conventions,

$$S_{\text{Het}} = \frac{1}{16\pi\alpha'^4} \int \text{Vol}_{10} e^{-\Phi} (R_G + |\nabla\Phi|^2 - \frac{1}{12} |H_3|^2 - \frac{1}{4} |F_2|^2).$$

Pour la théorie de type IIA,

$$S_{\text{IIA}} = \frac{1}{16\pi\alpha'^4} \int \text{Vol}_{10} \left(e^{-\Phi} (R_G + |\nabla\Phi|^2 - \frac{1}{12}|H_3|^2) - \frac{1}{4}|F_2|^2 - \frac{1}{48}|F_4|^2 \right) + \frac{1}{2} B_2 \wedge F_4 \wedge F_4,$$

où $H_3 = dB_2$, $F_2 = dA_1$, $F_4 = dA_3$ et $F'_4 = F_4 + A_1 \wedge H_3$.

Pour la théorie de type IIB,

$$S_{\text{IIB}} = \frac{1}{16\pi\alpha'^4} \int \text{Vol}_{10} \left(e^{-\Phi} (R_G + |\nabla\Phi|^2 - \frac{1}{12}|H_3|^2) - \frac{1}{2} |\nabla A^0|^2 - \frac{1}{12} |H'_3 + A^0 H_3|^2 - \frac{1}{240} |F_5|^2 \right) + A_4 \wedge H_3 \wedge H'_3,$$

où $H_3 = dB_2$, $H'_3 = dB'_2$ et $F_5 = dA_4 + B'_2 \wedge H_3$.

3.3. Si l'on ajoute aux cinq théories toutes les théories de cordes en dimension plus petite obtenues par compactification de Kaluza-Klein, on a une grande famille qui a beaucoup embarrassé les chevaliers de la théorie unique de tout. L'espoir est revenu lorsque Witten, Hull, Townsend, Duff, Sen et al., entre 1995 et 1996, on observé que, par dualité, toutes ces théories avaient l'air d'être équivalentes ([46], [90], [66]).

Afin de préserver une partie des super-symétries, les compactifications ont surtout été faites le long de tores \mathbf{T}^k , de produits $M \times \mathbf{T}^{k-4}$ avec M une surface K3 ou de produits $Y \times \mathbf{T}^{k-6}$ avec une variété de Calabi-Yau Y de dimension complexe 3. Ces variétés possèdent des espaces de modules (ne serait-ce que le rayon d'un cercle, ou le volume) qui forment pour chaque théorie \mathcal{T} un espace $U_{\mathcal{T}}$. Cet espace se projette sur des axes de constantes de couplages, en particulier celui de α' ou $\langle\Phi\rangle$.

Une *dualité* entre deux théories de cordes \mathcal{T}_1 et \mathcal{T}_2 est une bijection f d'une partie \mathcal{O}_1 des observables de \mathcal{T}_1 sur une partie \mathcal{O}_2 des observables de \mathcal{T}_2 , et une bijection f^* d'une partie V_1 de $U_{\mathcal{T}_1}$, sur V_2 dans $U_{\mathcal{T}_2}$ qui échangent les amplitudes, c'est-à-dire $\langle f(F_1) \rangle_{f^*(u_1)} = \langle F_1 \rangle_{u_1}$.

À proprement parler, il n'y a que des *conjectures de dualité* mais elles ont franchi des tests qui portent, soit sur les théories de supergravités effectives, soit sur les solutions classiques invariantes par une partie des super-symétries, encore appelées solutions BPS (cf. [80]).

Certaines dualités échangent les couplages forts et faibles des cordes ; ce sont les *S-dualités* ([74], [19], [20], [46], [58], [59]).

Exemples : pour $D = 10$, type I \longleftrightarrow Het₃₂, pour $D = 6$, IIA/K3 \longleftrightarrow Het/ \mathbf{T}^4 (i.e. la théorie de type IIA compactifiée sur une surface K3 est duale de la théorie hétérotique compactifiée sur un tore \mathbf{T}^4), pour $D = 10$, IIB \longleftrightarrow IIB.

D'autres dualités respectent les couplages faibles et peuvent être détectées sur la théorie perturbative ; ce sont les *T-dualités*, T comme target ([25], [30], [39]).

Exemples : IIA \longleftrightarrow IIB si on compactifie sur des cercles de rayons inverses (i.e. IIA/ $S^1_R \longleftrightarrow$ IIB/ $S^1_{R^{-1}}$). Dans les mêmes conditions $\text{Het}_{E_8 \times E_8} \longleftrightarrow \text{Het}_{SO_{32}}$.

Les T -dualités constituent des sous-groupes discrets des symétries signalées à la fin du § 1. Par exemple Het/\mathbf{T}^6 a un $O(6, 22; \mathbf{Z})$ de T -dualités dans le $O(6, 22; \mathbf{R})$. Mais dans ce cas, il y a aussi un $SL_2(\mathbf{R})$ mieux caché de S -dualités agissant sur $\tau = \langle A + ie^{-\Phi} \rangle$, A l'axion, Φ le dilaton. Dedans, un $SL_2(\mathbf{Z})$ passe aux cordes quantiques. On retrouve la dualité de Sen (§ 2.4) comme une des conséquences de basse énergie.

Pour la théorie de type II compactifiée sur \mathbf{T}^6 , le groupe E_7 des symétries de la super-gravité $N = 8$ contient un sous-groupe discret $E_7(\mathbf{Z})$ dont les éléments sont appelés U -dualités, formé à partir d'un $O(6, 6; \mathbf{Z})$ de T -dualités et du $SL_2(\mathbf{Z})$ de S -dualités de la théorie IIB en dimension 10.

La difficulté à chaque fois est de construire les états BPS, analogues des dyons, avec des charges quantifiées.

C'est là que les *branes* apparaissent :

Dans toutes les théories de supercordes, l'examen non perturbatif (mais spéculatif) a révélé l'existence de solutions (semi-)classiques particulières électriquement et/ou magnétiquement chargées, jouant le rôle des instantons, des solitons et des monopôles en théorie des champs : les *p-branes*. Ainsi nommées car elles définissent soit comme lieu singulier, soit comme centre, des sous-variétés de dimension $p + 1$ de l'espace-temps, donc des sous-variétés dimension p dans l'espace (au moins pour les solitons et les monopôles). Les 2-branes s'appellent simplement des membranes.

Sur ces variétés, on peut intégrer les $(p + 1)$ -formes différentielles du lagrangien de supergravité effective ou leurs formes duales.

La dualité de Poincaré accompagne les dualités de cordes et associe des q -branes duales aux p -branes. La duale d'une p -brane est donc une $(D - p - 4)$ -brane.

Par exemple, avec $D = 10$, en théorie IIA, la corde est une 1-brane et donne une 5-brane. En théorie IIB, on trouve une 3-brane autoduale.

Les théories de supercordes induisent sur les p -branes des théories de jauge dont les constantes de couplage proviennent de la géométrie en dimension D .

Les branes associées comme courants aux formes différentielles de type $\text{NS} \otimes \text{NS}$ furent assez faciles à construire ; on trouve par exemple les 0-branes associées aux dilatons Φ de toutes les théories, une 1-brane pour Het, courant de la jauge A , duale

d'une 5-brane, des membranes associées aux formes B_2 des théories de type II et Het.

La construction de courants pour les formes $R \otimes R$ fut plus longue à venir. J. Polchinski ([64], [65]) les trouva sous formes de \mathcal{D} -branes, \mathcal{D} comme Dirichlet ; ce sont des solutions singulières, comme des déchirures et des trous noirs dans l'espace-temps, le long desquelles les cordes, *a priori* toutes fermées de IIA et IIB peuvent s'ouvrir. Elles ont permis de compléter la liste BPS conjecturée pour la S -dualité de IIB/ S^1 ou la T -dualité du type I. En théorie de type IIA (resp. IIB), il y a une p -brane qui est \mathcal{D} -brane pour tout entier pair entre 0 et 8 (resp. impair entre -1 et 9). La T -dualité échange les \mathcal{D} -branes de IIA avec celles de IIB. En théorie de type I, il y a des \mathcal{D} -branes pour $p = 1, 5, 9$ (cf. [4], [7], [18], [30]).

La dernière révolution fut la découverte de la M-théorie (Townsend, Witten, 1996, [46], [90]) : en cherchant à comprendre le couplage fort de IIA, on trouve quelque chose de dual en dimension 11 dont la limite de basse énergie est la supergravité $D = 11$ de Cremmer, Julia, Scherk !

Ce « quelque chose » fut baptisé M-théorie, comme membrane, merveille ou mystère. Cette théorie est encore largement conjecturale, sa fonction principale est de compléter le diagramme suivant :

$$\begin{array}{ccc}
 \text{M-théorie} & \longrightarrow & \text{Supercorde IIA} \\
 \downarrow & & \downarrow \\
 \text{Supergravité, } D = 11 & \longrightarrow & \text{Supergravité IIA, } D = 10
 \end{array}$$

où les flèches horizontales sont des compactifications sur un cercle S^1 et les flèches verticales sont les limites de basse énergie ([16], [17], [44], [58], [80]).

Par compactification sur un intervalle compact, on trouve la duale de Het $_{E_8 \times E_8}$ en dimension 10. La M-théorie possède une membrane et une 5-brane duale.

En compactifiant une théorie de supercorde ou la M-théorie sur une variété X de dimension k , on peut enrouler une p -brane sur un d -cycle de X et obtenir dans \mathbf{R}^{D-k} une $(p-d)$ -brane ([1], [16], [19], [20], [81], [92]).

Ainsi la corde IIA est la membrane de M enroulée sur S^1 , la 2-brane et IIA est la projection de la membrane, sa 4-brane est la 5-brane de M enroulée et sa 5-brane non singulière est la projection de la 5-brane de M. Il y a aussi en M-théorie des branes noires analogues de trous noirs de dimensions 2, 4 et 6 ([16], [34]).

Les ressources mathématiques de la M-théorie semblent inépuisables :

1) Selon Strominger, Yau, Zaslow, toute variété de Calabi-Yau X de dimension 3 sur \mathbf{C} contient une famille à trois paramètres réels de 2-branes pour IIA compactifiée sur X . Si on applique la dualité T , on trouve une famille de 0-branes pour IIB formant la variété miroir Y de X . Cf. [36], [39], [41], [45], [52], [82].

2) Selon Witten, les états BPS de la théorie super-Yang-Mills $N = 2$, $D = 4$ qui permettent d'établir la dualité électrique-magnétique quantique et de construire la théorie de Seiberg-Witten viennent de configurations de branes de IIA qui correspondent par dualité à des 2-branes, de la forme $D^2 \times \mathbf{R}$ en M-théorie accrochées sur des 5-branes de la forme $\Sigma \times \mathbf{R}^4$ dans $S^1 \times \mathbf{R}^{10}$ ([93], [94]).

Par un glissement de sens, l'ultime théorie dont les limites variées sont IIA, IIB, I, supergravité $D = 11$, $\text{Het}_{E_8 \times E_8}$, $\text{Het}_{SO_{32}}$, s'est aussi appelée M. Schwarz et Sen ont bien proposé la lettre U comme unité, mais notre époque est plus mystique.

3.4. Paracelse avait deviné la théorie M en 1525 au cours de ses recherches sur l'Astrologie. Il n'a pas dit M comme quoi. Sans doute comme *Mysterium*. Il a même parlé du M.m. en précisant plus loin M.magnum.

Écoutons d'abord Lucien Braun dans son commentaire récent ([6]) à la traduction du *Volumen Paramirum* ([61]) :

«...La recherche sur l'*ens astrale* le conduit à s'interroger, à partir de l'Astre, sur les conditions d'existence de la vie humaine. L'homme est incapable de penser sa propre origine, estime Paracelse. Dès 1525, il se heurte, dans sa réflexion têtue, au «toujours déjà-là», au primordial, au primexistant : avant tout surgissement d'un existant quelconque, quelque chose est toujours déjà donné d'où il procède. Pour le faire entendre, Paracelse se sert de l'air comme comparaison (ailleurs aussi appelé *chaos*) et comme métaphore (qui doit nous *conduire plus loin*). Nous respirons l'air, écrit-il, sans lui nous ne pouvons exister. En filant la métaphore suffisamment loin, l'air lui-même finit par procéder de quelque chose que décidément on ne peut plus nommer – même par métaphore ! Paracelse désigne cet «abyssal-toujours-premier» par la lettre M (qui est mystère). Toujours déjà il y a quelque chose, avant tout donné, qui est la condition d'apparition de tout donné possible ; et qui est réfractaire à toute analyse subséquente.»

Écoutons Paracelse lui-même avant que, quelques lignes plus loin, il ne parle du poison contenu dans le M.m. qui risque d'infecter le corps de l'homme :

«... Vous dites que si l'air n'existait pas, toutes choses dépériraient, disparaîtraient ; tout ce qui vit étoufferait et mourrait. On peut dire de même : il existe quelque chose qui maintient et conserve le corps (qui contient la vie) ; la perte de ce «quelque

chose» serait aussi dommageable pour le corps que serait pour le vivant le défaut de l'air. Il convient donc de le préserver. L'air lui-même est conservé dans et par ce «quelque chose», et si cela venait à manquer, l'air aussi disparaîtrait. Le firmament en vit également, et si cela n'était, il disparaîtrait. Ce «quelque chose», je l'appelle M. Il n'existe rien de plus éminent dans tout le domaine de la création, et pour un médecin il n'y a rien de plus utile à méditer.

Je m'efforce de vous l'expliquer : le M. ne provient pas du firmament ; il n'y est pas né ; ce n'est pas le firmament qui nous l'envoie, etc. Rien de tout cela ne vaut. Mais le M. maintient toutes les créatures en leur être, que ce soit dans le ciel ou sur la terre ; tous les éléments vivent en lui et par lui... ».

Remerciements

L'auteur est seul responsable des fautes dans le texte, mais il tient à remercier ceux qui l'ont aidé à entrer dans le sujet : C. Bachas, A. Bahraini, J. Dunaud, B. Julia, O. Maspfuhl, S. Paycha, M. Slupinski et J.-B. Zuber.

Références

- [1] O. AHARONY – *String theory dualities from M-theory*, Nuclear Physics B476 (1996), 470-483.
- [2] I. ANTONIADIS, E. CREMMER, K.S. STELLE – *Les supercordes*, 1ère partie, Gazette des Mathématiciens, S.M.F., n° 87, janvier 2001 ; 2ème partie, Gazette des Mathématiciens, S.M.F., n° 88, avril 2001.
- [3] M.F. ATIYAH, N.J. HITCHIN – *The geometry and dynamics of magnetic monopoles*, Princeton U.P., 1988.
- [4] C. BACHAS – *Lectures on D-branes*, dans [59].
- [5] J.-B. BOST – *Fibrés déterminants, déterminants régularisés et mesures sur les espaces de modules des courbes complexes*, Sémin. Bourbaki exp. n° 676, Astérisque **152-153** (1987), 113-149.
- [6] L. BRAUN – *Paracelse, de l'Astrologie*, Presses Universitaires de Strasbourg (2002).
- [7] P. BREITENLOHNER, D. MAISON, G. GIBBONS – *4-dimensional black holes from Kaluza-Klein theories*, Comm. Math. Physics (1988), 295-333.
- [8] G. BENFATTO, G. GALAVOTTI – *Renormalization group*, Physics Notes N° 1, Princeton University Press, 1995.
- [9] R. COQUEREAUX, A. JADCZYK – *Symmetries of Einstein-Yang-Mills fields and dimensional reductions*, Comm. Math. Physics, 1985.
- [10] E. CREMMER, B. JULIA – *The SO(8) supergravity*, Nuclear Physics B159 (1979), 141-212.
- [11] E. CREMMER, B. JULIA, H. LÜ, C. POPE – *Dualisation of dualities*, Nuclear Physics B523 (1998), 73-144.

- [12] E. CREMMER, B. JULIA, J. SCHERK – *Supergravity theory in 11 dimensions*, Physics Letters B76 (1978), 409-412.
- [13] P. DELIGNE, P. ETINGOF et al. – *Quantum fields and strings: A course for mathematicians*, AMS, IAS, 1999.
- [14] P.A.M. DIRAC – *Quantised Singularities in the Electromagnetic Field*, Proc. Roy. Soc. A133 (1931), 60-72.
- [15] P.A.M. DIRAC – *The Theory of Magnetic Poles*, Phys. Review 74 (1948), 817-830.
- [16] M.J. DUFF – *M-theory (the theory formerly known as strings)*, International Journal of Modern Physics A. Vol. 11, No 32 (1996), 5623-5641.
- [17] M.J. DUFF – *The world in eleven dimensions: a tribute to Oskar Klein*, arXiv:hep-th/0111237, 38 pp.
- [18] M.J. DUFF, Ramzi R. KHURI, J.X. LU – *String solitons*, Physics Reports 259 (1995), 213-326.
- [19] M.J. DUFF, J.T. LIU, R. MINASIAN – *Eleven-dimensional origin of string/string duality: a one-loop test*, Nuclear Physics B452 (1995), 261-282.
- [20] M.J. DUFF, J.T. LIU, J. RAHMFELD – *Four-dimensional string/string/string triality*, Nuclear Physics B459 (1996), 125-159.
- [21] M.J. DUFF, B.E.W. NILSSON, C.N. POPE – *Kaluza-Klein supergravity*, Physics Reports 130, Nos 1 & 2 (1986), 1-142.
- [22] M.J. DUFF, K.S. STELLE – *Multi-membrane solutions of $D = 11$ supergravity*, Physics Letters B 253 (1991), 113-118.
- [23] J. EHLERS – *Extension solutions...*, J. Diss. Univ. Hamburg, Hambourg, 1958.
- [24] S. FERRARA, J. SCHERK, B. ZUMINO – *Algebraic properties of extended supergravity theories*, Nuclear Physics B121 (1977), 393-402.
- [25] A. FONT, L.E. IBÁÑEZ, D. LÜST, F. QUEVEDO – *Strong-weak coupling duality and non-perturbative effects in string theory*, Physics Letters B 249 (1990), 35-43.
- [26] J. GAUNLETT – *Supersymmetric monopoles and duality*, dans [59].
- [27] K. GAWĘDSKI – *Conformal field theory*, Sémin. Bourbaki, exp. n° 704, Astérisque **177-178** (1989), 95-126.
- [28] R. GEROCH – *A method for generating solutions of Einstein's equations*, Journal of Math. Physics **12**, n° 6 (1971), 918-924.
- [29] G.W. GIBBONS – *The Sen conjecture for fundamental monopoles of distinct types*, Physics Letters B382 (1996), 53-59.
- [30] A. GIVEON, M. PORRATI, E. RABINOVICI – *Target space duality in string theory*, Physics Reports 244 (1994), 77-202.
- [31] J. GLIMM, A. JAFFE – *Quantum physics, a functional integral point of view*, Springer-Verlag (1981).
- [32] P. GODDARD, J. NUYTS, D. OLIVE – *Gauge theory and magnetic charge*, Nuclear Physics B125 (1977), 1-28.
- [33] M.B. GREEN, J.H. SCHWARZ, E. WITTEN – *Superstring theory*, Vol. 1 & 2, Cambridge University Press, 1987.
- [34] R. GÜVEN – *Black p-brane solutions of $D = 11$ supergravity theory*, Physics Letters B276 (1992), 49-55.
- [35] A. HANANY, E. WITTEN – *Type IIB superstrings, PBS monopoles, and three-dimensional gauge dynamics*, Nuclear Physics B492 (1997), 152-190.

- [36] J.A. HARVEY, G. MOORE – *Algebras, BPS states, and strings*, Nuclear Physics B463 (1996), 315-368.
- [37] S. HAWKING, G. ELLIS – *The large scale structure of space-time*, Cambridge University Press, 1973.
- [38] W. HEISENBERG – *Les principes de la théorie de quanta*, Gauthier-Villars, 1972.
- [39] M. HENNINGSON, G. MOORE – *Counting curves with modular forms*, Nuclear Physics B472 (1999), 518-528.
- [40] N. HITCHIN – *The Yang-Mills equation and the topology of 4-manifolds (after Simon K. Donaldson)*, Sémin. Bourbaki (1982/83), exp. n° 606, Astérisque 105-106 (1982), 167-178.
- [41] N. HITCHIN – *Lectures on special lagrangian submanifolds*, arXiv:math.DG/9907034 v1 (6/7/99).
- [42] G. 't HOOFT – *Under the spell of the gauge principle*, Advanced Series in Mathematical Physics, vol. 19, World Scientific, 1994.
- [43] G. 't HOOFT, M. VELTMAN – *Diagrammar*, CERN (1973) et G. 't HOOFT – *The renormalization group in QFT* (1998), in *Under the spell of the gauge principle*, World Scientific, 1994.
- [44] P. HORAVA, E. WITTEN – *Heterotic and Type I string dynamics from eleven dimensions*, Nuclear Physics B460 (1996), 506-524.
- [45] K. HORI, C. VAFA – *Mirror Symmetry*, ArXiv:Hep-th/00022V3 (mars 2000).
- [46] C.M. HULL, P.K. TOWNSEND – *Unity of superstring dualities*, Nuclear Physics B438 (1995), 109-137.
- [47] C. ITZYKSON, J.B. ZUBER – *Quantum field theory*, McGraw-Hill, 1980.
- [48] B. JULIA – *Dualities in the classical supergravity limits*, arXiv:hep-th/9805083.
- [49] B. JULIA – *Magics of M-gravity*, arXiv:hep-th/0105031 v1.
- [50] B. JULIA, A. ZEE – *Poles with both magnetic and electric charges in non-Abelian gauge theory*, Phys. Rev. D11 (1975), 22-27 à 22-32.
- [51] A. KONECHNY, A.S. SCHWARTZ – *On $(k \oplus \ell|q)$ -Dimensional Supermanifolds*, in *Supersymmetry and Quantum Field Theory*, Proc. Kharkov, Ukraine 1997, J. Weiss, V.P. Akulov (ed.), L. N. in Physics 509, Springer (1998), 201-206.
- [52] M. KONTSEVICH – *Mirror symmetry in dimension 3*, Sémin. Bourbaki, exp. n° 801, Astérisque **237** (1996), 275-293.
- [53] R.P. LANGLANDS – *Problems in the theory of automorphic forms*, in *Lectures in Modern Analysis and Applications III*, Lect. Notes in Math. 170 (1970).
- [54] K. LEE, E. WEINBERG, P. YI – *Moduli space of many BPS monopoles for arbitrary gauge groups*, Phys. Rev D54 (1996), 1633-1643.
- [55] Y.I. MANIN – *Gauge field theory and complex geometry*, Springer-Verlag, 1984.
- [56] C.W. MISNER, K.S. THORNE, J.A. WHEELER – *Gravitation*, N.H. Freeman.
- [57] C. MONTONEN, D. OLIVE – *Magnetic monopoles as gauge particles*, Physic Letters B72 (1977), 117-120.
- [58] N.A. OBERS, B. PIOLINE – *U-duality and M-theory*, Physics Reports 318 (1999), 113-225.
- [59] D. OLIVE, P. WEST – *Duality and supersymmetric theories*, Publications of the Newton Institute, Cambridge University Press, 1999.

- [60] H. OSBORN – *Topological charges for $N = 4$ supersymmetric gauge theories and monopoles of spin 1*, Physics Letters B83 (1979), 321-326.
- [61] T. PARACELSE – *Volumen (Medicinae) Paramirum*, Sudhoff (1525), 176-188.
- [62] M.E. PESKIN, D.V. SCHROEDER – *An introduction to Quantum Field Theory*, Addison-Wesley Publishing Company, 1995.
- [63] J. POLCHINSKI – *Renormalization and effective lagrangians*, Nuclear Physics B231 (1984), 261-295.
- [64] J. POLCHINSKI – *Combinatorics of boundaries in string theory*, Physical Review D 50 (1994), 6041-6045.
- [65] J. POLCHINSKI – *Dirichlet-branes and Ramond-Ramond charges*, arXiv:Hep-ph/9510017V3 (novembre 1995).
- [66] J. POLCHINSKI – *String duality*, Reviews of Modern Physics 68 (1996), 1245-1258.
- [67] J. POLCHINSKI, E. WITTEN – *Evidence for heterotic-Type I string duality*, Nuclear Physics B 460 (1996), 525-540.
- [68] A.M. POLYAKOV – *Gauge fields and strings*, Contemporary concepts in Physics, vol. 3, Harwood Academic Publishers (1987).
- [69] M.K. PRASAD, C.M. SOMMERFIELD – *Exact Solution for the 't Hooft Monopole and the Julia-Zee Dyon*, Phys. Rev. Letters 35 (1975), 760-762.
- [70] S.-J. REY – *Confining phase of superstrings and axionic strings*, Physical Review D 43 (1991), 526-538.
- [71] V. RIVASSEAU – *From perturbative to constructive renormalization*, Princeton Series in Physics, Princeton University Press, 1991.
- [72] A. ROGERS – *A global theory of supermanifolds*, J. Math. Phys. 21 (6), Juin 1980.
- [73] A.S. SCHWARTZ – *On the definition of superspace*, Theoret. and Math. Phys. 60 (1984), n° 1, 657-660.
- [74] J.H. SCHWARZ, A. SEN – *Duality symmetries of 4D-heterotic strings*, Physics Letters B 312 (1993), 105-114.
- [75] J. SCHWINGER – *Magnetic Charge in Quantum Field Theory*, Phys. Rev 144 (1966), 1087-1093.
- [76] N. SEIBERG, E. WITTEN – *Electric magnetic duality, monopole condensation and confinement in $N = 2$ supersymmetric Yang-Mills theory*, Nuclear Physics B 426 (1994), 19-52.
- [77] N. SEIBERG, E. WITTEN – *Monopoles, duality and chiral symmetric breaking in $N = 2$ supersymmetric QCD*, Nuclear Physics B 431 (1994), 484-550.
- [78] G. SEGAL – Graeme Segal [http://onlive.itp.ucsb.edu/onlive/geom 99/segal](http://onlive.itp.ucsb.edu/onlive/geom%2099/segal)
- [79] A. SEN – *Dyon monopole bound states, self-dual harmonic forms on the multi-monopole moduli space, and $SL_2(\mathbf{Z})$ invariance in string theory*, Physics Letters B 329 (1994), 217-221.
- [80] A. SEN – *An introduction to non-perturbative string theory*, 297-413, dans [59].
- [81] D. SOROKIN – *Superbranes and superembeddings*, Physics Reports 329 (2000), 1-101.
- [82] A. STROMINGER, S.-T. YAU, E. ZASLOW – *Mirror symmetry is T-duality*, Nuclear Physics B479 (1996), 243-259.
- [83] P.K. TOWNSEND – *The eleven-dimensional supermembrane revisited*, Physics Letters B350 (1995), 184-188.

- [84] J. WESS, J. BAGGER – *Supersymmetry and supergravity*, 2nd edition, Princeton University Press, 1992.
- [85] K.G. WILSON – *Renormalization of a scalar field theory in strong coupling*, Physical Review D, vol. 6, n° 2 (1972), 419-426.
- [86] K.G. WILSON – *The renormalization group: Critical phenomena and the Kondo problem*, Review of Modern Physics 47 (1975), 773-840.
- [87] K.G. WILSON – *The renormalization group and critical phenomena*, Reviews of Modern Physics, vol. 55, n° 3 (1983).
- [88] E. WITTEN – *Dyons of charge $e\theta/2\pi$* , Physics Letters B86 (1979), 283-287.
- [89] E. WITTEN – *Strong coupling and the cosmological constant*, Modern Physics Letters A10 (1995), 2153-2155.
- [90] E. WITTEN – *String theory dynamics in various dimensions*, Nuclear Physics B443 (1995), 85-126.
- [91] E. WITTEN – *Bound states of strings and p-branes*, Nuclear Physics B460 (1996), 335-350.
- [92] E. WITTEN – *Five-branes and M-theory on an orbifold*, Nuclear Physics B463 (1996), 383-397.
- [93] E. WITTEN – *Duality, spacetime and quantum mechanics*, Physics today, May 1997, 28-33.
- [94] E. WITTEN – *Solutions of four-dimensional field theories via M-theory*, Nuclear Physics B500 (1997), 3-42.
- [95] E. WITTEN – *Overview of K-theory applied to strings*, arXiv:hep-th/0007175 v1
- [96] E. WITTEN – *Deconstruction, G_2 holonomy, and Doublet-Triplet splitting*, arXiv:Hep-ph/0201018V2, 15.01.2002.
- [97] E. WITTEN, D. OLIVE – *Supersymmetry algebras that include topological charges*, Physics Letters B78 (1978), 97-101.
- [98] D. ZWANZIGER – *Exactly Soluble Non-relativistic Model of Particles with Both Electric and Magnetic Charges*, Phys. Rev. 176, n° 5 (1968), 1480-1495.

Daniel BENNEQUIN

Université Denis-Diderot (Paris VII)
 Équipe de Géométrie et Dynamique
 Institut de Mathématique de Jussieu
 UMR 9994 du CNRS
 2 place Jussieu
 F-75251 PARIS Cedex 05

**ALGÈBRE DE HOPF DES DIAGRAMMES DE FEYNMAN,
RENORMALISATION ET FACTORISATION DE WIENER–HOPF**
[d’après A. Connes et D. Kreimer]

par Louis BOUTET de MONVEL

1. INTRODUCTION

1.1. Renormalisation

Cet exposé a pour objet de décrire le point de vue de A. Connes et D. Kreimer sur la renormalisation en théorie quantique des champs (théorie perturbative), expliquées dans les articles [CK1, CK2]. La plupart des démonstrations sont omises, et on en trouvera les détails dans *loc. cit.*. Je remercie chaleureusement A. Connes, J. Zinn-Justin, A. Arabia pour leur aide et leurs conseils. Je remercie particulièrement P. Cartier qui a relu et corrigé ces notes dans la meilleure tradition de N. Bourbaki.

Un des effets frappants de la théorie de la renormalisation est de faire prévoir que les constantes de structure qui gouvernent un grand système physique dépendent de l’échelle d’observation. Ce fait se constate aussi dans d’autres phénomènes étendus où tous les ordres de grandeur contribuent, en particulier dans la description des transitions de phase pour lesquelles aussi l’idée de la renormalisation est pertinente (cf. [ZJ]). Il s’observe déjà dans des phénomènes plus classiques et connus depuis le 19ème siècle : ainsi une balle de ping-pong de masse $m = 2gr$, immergée dans l’eau (masse d’eau déplacée : $M = 32gr$ environ), subit une poussée vers le haut correspondant au poids $M - m = 30gr$ (force d’Archimède moins son poids). Mais, si on la lâche, elle ne remonte pas avec une accélération de $\frac{M-m}{m}g = 15g$: il y a une interaction avec l’eau environnante et le résultat final est que, tant que la vitesse reste petite, la balle remonte avec une accélération de $\frac{M-m}{m+\frac{M}{2}}g < 2g$. Vu de l’extérieur, tout se passe comme si la balle était affectée de la masse apparente $m' = m + \frac{M}{2}$, et si l’on n’avait pas le moyen de disséquer en détail l’interaction qui a lieu près de la balle, m' serait la seule « masse effective » qu’on puisse lui attribuer à cette échelle.

La théorie quantique des champs veut décrire et expliquer les phénomènes fondamentaux qui gouvernent la physique à l’échelle nucléaire ou particulaire. Dès le départ, elle se heurte à des difficultés considérables car les quantités « naturelles » qu’on veut calculer

sont décrites par des analogues d'intégrales qui n'ont pour l'instant pas beaucoup de sens mathématique, ou dans le cas plus simple de la théorie perturbative, par des séries asymptotiques d'intégrales divergentes.¹

Un des objets de la renormalisation est d'attribuer à ces intégrales une valeur («partie finie»). Les parties finies d'intégrales interviennent dans beaucoup de questions d'E.D.P. et Hadamard en a donné une description systématique par troncature ou par prolongement analytique. Ici, elles arrivent en série et il faut les organiser de façon cohérente. Les théoriciens du champ ont énoncé pour cela un certain nombre de règles qui, pour être précises et efficaces, n'en sont pas moins très compliquées.

A. Connes et D. Kreimer montrent que ces règles peuvent être résumées très élégamment par une factorisation à la Wiener-Hopf (ou Riemann-Hilbert) d'un lacet à valeurs dans un groupe associé aux diagrammes de Feynman, et que cela donne lieu à des formules «universelles» qui vivent dans le groupe «universel» des difféomorphismes formels tangents à l'identité.

1.2. Champs

Dans la théorie classique, un champ² est une fonction A sur \mathbf{R}^d qui satisfait une équation

$$\partial^2 A - F'(A) = 0$$

où ∂^2 désigne le d'Alembertien $-\partial_t^2 + \sum \partial_{x_i^2}$, et F' est la dérivée d'une fonction donnée F . Il revient au même de dire que A extrémise l'intégrale d'action $\int d^d x \mathcal{L}(A)$ où le Lagrangien est $\mathcal{L}(A) = \frac{1}{2}(\partial A)^2 + F(A)$; d est la dimension de l'espace-temps ; dans les formules finales c'est un entier D , $D = 4$ dans notre monde, $D = 6$ ci-dessous, mais de toute façon on devra «le faire varier continûment» pour la régularisation dimensionnelle.

Un champ quantique ϕ est une fonction (généralisée) à valeurs opérateurs, satisfaisant à un certain nombre d'axiomes pour lesquels je renvoie à la bibliographie. Il est déterminé par un état vide (d'énergie minimale) noté $\langle 0 \rangle$, et les «fonctions de Green»

$$G(x_1, \dots, x_N) = \langle 0 | T \phi(x_1) \dots \phi(x_N) | 0 \rangle,$$

(le signe T signifie qu'on réordonne les facteurs x_i par ordre de temps $t = x^0$ décroissant).

Pour quantifier une théorie classique, les physiciens s'inspirent de la formulation variationnelle et de la mécanique statistique, et affirment que la fonction de Green doit être

¹Il ne s'agit en aucune façon d'une théorie floue ou imprécise puisque les résultats théoriques qu'elle fournit concordent aussi bien que possible avec les résultats expérimentaux - jusqu'à 14 chiffres significatifs en électrodynamique quantique - ce qui est un exploit inégalé par la physique antérieure. C'est plutôt un exemple de plus du fait que les résultats les plus spectaculaires dans la description du monde qui nous entoure précèdent souvent les mathématiques qui les expliquent le plus élégamment.

²Pour simplifier, nous ne considérerons ici qu'un champ scalaire, i.e. A est une fonction numérique ; la physique utilise aussi des champs vectoriels, tensoriels, spinoriels, etc..

décrite par une intégrale fonctionnelle sur l'espace de tous les champs classiques:

$$G(x_1, \dots, x_N) = \mathcal{N}^{-1} \int [dA] e^{iS(A)} A(x_1) \dots A(x_N)$$

où $S(A) = \int d^d x \mathcal{L}(A)$ est l'intégrale d'action, et $\mathcal{N} = \int [dA] e^{iS(A)}$ est un facteur de normalisation assurant $\langle 0|0 \rangle = 1$. Dans cette intégrale aucun des termes \mathcal{N} , $[dA]$, \int n'a de sens, du moins pris séparément; le tout évoque une intégrale gaussienne comme celles du mouvement Brownien, mais la notation des physiciens est plus suggestive.

Dans la théorie perturbative, l'action est $\mathcal{L}(A) = \mathcal{L}_0(A) + \mathcal{L}_{int}(A)$ où

$$(1) \quad \mathcal{L}_0 = \frac{1}{2}(\partial A)^2 - \frac{m^2}{2}A^2$$

est le Lagrangien du champ libre de masse m , et $\mathcal{L}_{int}(A)$ est le terme perturbatif. Le développement en série (formelle) de l'exponentielle suggère

$$G(x_1, \dots, x_N) \sim \mathcal{N}^{-1} \sum \frac{i^n}{n!} \int [dA] e^{iS_0(A)} (S_{int}(A))^n A(x_1) \dots A(x_N)$$

avec

$$\mathcal{N} \sim \sum \frac{i^n}{n!} \int [dA] e^{iS_0(A)} (S_{int}(A))^n.$$

Dans ces expressions interviennent une foison d'intégrales «gaussiennes» de la forme

$$(2) \quad \sum \int [dA] e^{iS_0(A)} S_{int}(A)^n A(x_1) \dots A(x_N)$$

qui se ramènent à des intégrales en dimension finie.

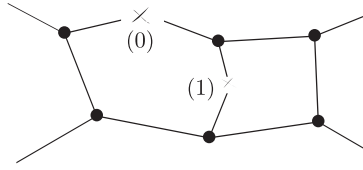
Dans [CK1, CK2] et dans beaucoup d'autres travaux, les auteurs ont choisi comme exemple le Lagrangien correctif $\mathcal{L}_{int}(A) = \frac{g}{3!}A^3$ en dimension $d = 6$ («théorie ϕ^3 »), parce que cela illustre déjà de façon simple, mais significative, le cas général, même si cela ne correspond pas à une théorie physique vraisemblable. Nous ferons de même ici.

2. INTÉGRALES ET DIAGRAMMES DE FEYNMAN

2.1. Diagrammes de Feynman

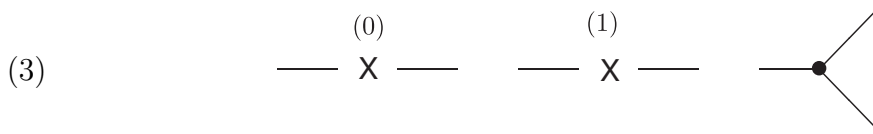
Les intégrales de Feynman (2) et les problèmes que pose leur divergence sont connus depuis le début de la théorie (Dirac). Le génie de Feynman l'a conduit à les indexer par les diagrammes qui portent son nom, qui en plus d'indiquer clairement l'intégrand et l'espace où l'on intègre, évoquent de façon imagée des interactions ou collisions virtuelles.

Un diagramme de Feynman est un graphe coloré (non orienté): ³ il a un ensemble de sommets colorés, reliés par des arêtes.



Les arêtes extérieures, ou pattes, n'ont qu'un seul sommet; les arêtes intérieures ont deux sommets distincts. Les types de sommets se distinguent par le nombre d'arêtes, et une couleur (pour les sommets à deux arêtes).

Dans la théorie ϕ^3 , il y a 3 types de sommets :



correspondant respectivement aux termes quadratiques A^2 , $(\partial A)^2$ et au terme cubique A^3 de $\mathcal{L}(A)$.

2.2. Intégrales de Feynman

Un tel diagramme Γ code une distribution U_Γ de N variables vectorielles (x_1, \dots, x_N) , N le nombre de pattes, qu'on repère plutôt par sa transformée de Fourier $U_\Gamma(p_1, \dots, p_N)$ des moments extérieurs p_e , qui vérifient des relations de conservation $\sum p_e = 0$ (une pour chaque composante connexe). Celle-ci est définie comme suit : à chaque arête i est associée une variable de moment p_i à d dimensions, libre si i est une patte extérieure, muette (variable d'intégration, notée k_i dans [CK1]) sinon. U_Γ est alors définie par une intégrale sur l'espace des moments internes (ou plutôt sur un sous-espace, car l'intégrand comporte des facteurs δ) :

$$(4) \quad U_\Gamma(p_1, \dots, p_N) = \int \prod d^d p_i \ I_\Gamma(p_i).$$

L'intégrand I_Γ est un produit de facteurs auxquels contribuent séparément les arêtes intérieures et les sommets, par des règles explicites suivantes (pour un autre \mathcal{L}_{int} il faudrait en général plus de types de sommets (ou d'arrêtes), et compléter les règles qui décrivent les facteurs associés aux sommets ou aux arêtes).

I_1 Chaque arête i produit un facteur $\frac{1}{m^2 + p_i^2}$.

I_2 Chaque sommet v produit un facteur $\delta(\sum_{i \ni v} p_i^v)$ (cela fait autant de relations linéaires que de sommets, exprimant la conservation des moments).⁴

³ on est aussi amené à colorer les arêtes, et à orienter le graphe, s'il y a plusieurs espèces de particules et qu'on doit distinguer entre particules et anti-particules

⁴ p_i^v désigne le flux de moment entrant dans v indexé par l'arête i : si v, v' sont les deux sommets d'une arête i , on a $p_i^{v'} = -p_i^v$.

I_3 Chaque 2-sommet de type 0 produit un facteur m^2 .

I_4 Chaque 2-sommet de type 1 produit un facteur p^2 (les deux moments associés sont opposés et ont même p^2).

I_5 Chaque 3-sommet produit un facteur $\mu^{3-d/2}g$.

μ est l'unité de masse – pour l'instant une variable formelle – introduite pour que les constantes qui figurent dans les intégrales à renormaliser soient des nombres purs, « sans dimension ». (Il s'agit cette fois de la dimension des physiciens : les mesures sont éléments des puissances tensorielles d'un espace vectoriel (ou sections d'un fibré vectoriel) de base. Dans une théorie scalaire relativiste comme ici, celui-ci est de rang 1, et la dimension d'une quantité physique est l'exposant de la puissance tensorielle où on la mesure ; de façon équivalente, toutes les quantités (masse, moment, ...) de la théorie sont commensurables à une puissance de l'unité de masse μ ; la « dimension » d'une quantité P est l'exposant s tel que $\mu^{-s}P$ soit un nombre pur).

On peut résumer l'effet des facteurs δ en disant que les moments extérieurs vérifient la relation de conservation $\sum p_e = 0$ et que l'intégrale est effectuée sur l'espace E_{Γ}^i de dimension $(I - V + C)d$ des moments intérieurs indépendants (I est le nombre d'arêtes internes, V le nombre de sommets, C le nombre de composantes connexes).

Ces intégrales sont le plus souvent divergentes – doublement : à distance finie elles ont des singularités car la norme utilisée est celle de la relativité : $p^2 = -p_0^2 + \sum p_i^2$; les physiciens s'empressent de la remplacer par la norme euclidienne en faisant une « rotation de Wick » (i.e. en remplaçant le temps $t = x^0$ par un temps imaginaire pur it), ce qui élimine ces singularités (sauf pour le cas limite $m = 0$); le cas relativiste s'en déduit en principe par prolongement analytique, et cela ne change de toute façon pas grand'chose au principal problème de la renormalisation, qui vient des divergences à l'infini de ces intégrales (« catastrophe ultra-violette »).

2.3. Parties finies

Hadamard nous a montré comment on peut attribuer une valeur à une intégrale divergente. Par exemple, pour une singularité à l'origine, on tronque et on enlève une partie infinie « évidente » dans un développement asymptotique :

$$\text{pf} \int_0^1 dt t^{s-1} \varphi(t) = \lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 dt t^{s-1} \varphi(t) + \sum \frac{\varphi^{(j)}(0)}{j!} \frac{\epsilon^{s+j}}{s+j}.$$

Il y a un problème supplémentaire dans le cas - le plus utile - où l'exposant s est entier.

Un autre moyen « canonique » est de remarquer que le membre de gauche est une fonction méromorphe de s , avec ici des pôles simples aux entiers négatifs. Ce prolongement redonne la partie finie pour s régulier. Aux pôles, on retient le terme constant du développement de Laurent ; c'est un peu moins canonique, car il dépend du choix de la borne supérieure d'intégration, ou d'une unité de longueur.

Une des méthodes utilisées par les physiciens, qui est celle qu'utilisent Connes et Kreimer, est la méthode de régularisation dimensionnelle, qui est une méthode de prolongement analytique : pour une fonction $f(p)$ invariante par rotation sur \mathbf{R}^d , cela revient à intégrer en coordonnées polaires :

$$\int d^d p f(p) = v_d \int_0^\infty dr r^{d-1} f(r),$$

où $v_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}$ est le $(d-1)$ -volume de la sphère unité, de sorte qu'on ait $\int d^d x e^{-x^2} = \pi^{d/2}$. Le résultat est une fonction méromorphe de d .

La régularisation dimensionnelle peut être effectuée pour les intégrales de Feynman. Elle conduit pour chacune à une distribution $U_\Gamma = U_\Gamma(p, \mu, d)$ des moments extérieurs, qui dépend du paramètre μ et, de façon méromorphe, de la dimension d , avec un pôle pour $d = D$ (en général multiple, mais d'ordre fini ; ici $D = 6$; il peut y avoir d'autres pôles en dehors de D).

REMARQUE 2.1. — Les valeurs (renormalisées) des intégrales de Feynman sont intéressantes pour elles-mêmes : en dehors de puissances de π , on trouve (entre autres) des combinaisons rationnelles de $\zeta(3), \zeta(5), \dots$ et plus généralement de nombres multi-zêtas dont le degré de complexité (poids, profondeur) est lié à la topologie du diagramme; mais ceci est une autre histoire.

3. ALGÈBRE DE HOPF DES DIAGRAMMES DE FEYNMAN

3.1. Diagrammes habillés et coproduit

Dans la théorie ϕ^3 , les diagrammes utiles (nous dirons « admissibles ») sont réunion disjointe de diagrammes « simples », i.e. connexes, et qui restent connexes si on leur enlève une arête (le sigle des physiciens est 1PI = irréductibles à une particule); comme pour les groupes, on compte que les objets de la fig. (3), les plus « simples », ne le sont pas.

Pour fabriquer des nombres à partir des diagrammes et des intégrales de Feynman, et surtout pour décrire la règle de réinsertion de diagrammes pour le calcul de ces intégrales par itération d'intégrations partielles, il convient de les « habiller ». Un diagramme « habillé » est une combinaison de diagrammes admissibles:

$$\sum \sigma_\Gamma \otimes \Gamma,$$

où chaque coefficient σ_Γ est une distribution sur l'espace E_Γ des moments extérieurs p_e (indexés par les N arêtes extérieures, et liés par la relation de conservation $\sum p_e = 0$, de sorte que E_Γ est de dimension $(N-1)d$).⁵

⁵L'espace des coefficients distribution n'est pas précisé et cela n'a guère d'importance ; on peut se limiter à des distributions invariantes par le groupe de Lorentz, et il faut en outre qu'on puisse appliquer le procédé de régularisation dimensionnelle, pour pouvoir faire varier d continûment autour de $D = 6$ en contournant le pôle.

Les diagrammes admissibles, habillés ou non, forment une algèbre commutative graduée \mathcal{H} : la loi de produit est la somme disjointe, et \otimes pour les coefficients, l'unité est le diagramme vide (affecté du coefficient 1). La graduation est définie par le nombre de boucles $L = I + V - C = \dim H^1(\Gamma)$, voir ci-dessous §3.2.

Connes et Kreimer définissent sur \mathcal{H} une structure d'algèbre de Hopf au moyen du coproduit

$$(5) \quad \Delta(\Gamma) = \Gamma \otimes 1 + 1 \otimes \Gamma + \sum_{\gamma \subset \Gamma, \iota} \gamma_{(\iota)} \otimes \Gamma/\gamma_{(\iota)}.$$

Dans cette somme, γ désigne un diagramme admissible non vide, inséré dans Γ , dont les composantes ont deux ou trois pattes : les ensembles de sommets des composantes de γ dans Γ sont deux à deux disjoints, seules des arêtes extérieures peuvent être communes.

ι est une collection de signes 0 ou 1 indexée par les composantes à deux pattes.

Le deuxième facteur, $\Gamma/\gamma_{(\iota)}$ s'obtient en contractant en un point chaque composante α , et en l'affectant de la couleur $\iota(\alpha)$ (0 ou 1) s'il s'agit d'une composante à deux pattes.

Dans le premier facteur les composantes de $\gamma_{(\iota)}$ ont deux ou trois pattes, et un habit fixé par ι : $\sigma_0 \otimes \gamma_\alpha$ si γ_α est une composante à trois pattes, ou deux pattes et la couleur est $\iota(\alpha) = 0$, $\sigma_1 \otimes \gamma_\alpha$ si γ_α a deux pattes et la couleur correspondante est 1. Les distributions $\sigma_{0,1}$ sont choisies de sorte que $\sigma_0(am^2 + bp^2) = a$, $\sigma_1(am^2 + bp^2) = b$.⁶

THÉORÈME 3.1. — *Muni du coproduit Δ ci-dessus, \mathcal{H} est une algèbre de Hopf.*

Ce théorème est démontré en détail dans [CK1] et nous ne reproduisons pas ici la démonstration. En fait, il correspond à l'idée naturelle qu'on peut dans les intégrales (4) faire des intégrations partielles par paquets sur des sous-diagrammes (pourvu que ceux-ci restent admissibles), et itérer cette opération autant de fois qu'on veut. Il y a une complication mineure venant du fait que, pour les diagrammes à deux pattes, la règle de réinsertion dépend de la couleur du diagramme qu'on réinsère.

3.2. Graduation

Un poids multiplicatif sur \mathcal{H} (valuation) est une fonction w à valeurs entières telle que

$$w(fg) = w(f) + w(g).$$

Il est complètement défini par ses valeurs sur les diagrammes simples puisque \mathcal{H} est l'algèbre de polynômes engendrée par ceux-ci. Nous nous intéressons plus particulièrement à ceux qui respectent le coproduit : $w(\Delta f) = w(f)$: les plus évidents sont

$$\begin{aligned} I(\Gamma) &= I &&= \text{le nombre d'arêtes intérieures de } \Gamma \\ v(\Gamma) &= V - C &&= \text{le nombre de sommets} - \text{le nombre de composantes.} \end{aligned}$$

⁶L'article [CK1] dit plutôt que $\gamma_{(0)}$ est le diagramme γ affecté de moments extérieurs nuls, et $\gamma_{(1)} = \frac{\partial}{\partial p^2} \gamma|_{p=0}$: j'espère avoir interprété correctement. Il faut modifier cette règle dans le cas limite $m = 0$.

DÉFINITION 3.2. — La graduation canonique de \mathcal{H} est définie par $L(\Gamma) = I - V + C = \dim H^1(\Gamma)$. Pour cette graduation \mathcal{H} est connexe, i.e. \mathcal{H}^0 est le corps des constantes (\mathbf{C} ou n'importe quel corps de caractéristique 0).

L'assertion résulte de ce que, par définition, on a $\dim H^1(\Gamma) > 0$ si Γ est non vide.

On note \mathcal{H}_+ l'idéal des éléments de degré > 0 . Il est immédiat qu'on a

$$\Delta(X) - X \otimes 1 - 1 \otimes X \in \mathcal{H}_+ \otimes \mathcal{H}_+ \quad \text{si } X \in \mathcal{H}_+.$$

Un autre poids de bigèbre utile est

$$(6) \quad \omega(\Gamma) = V_3 - N + 2,$$

où V_3 est le nombre de sommets à trois pattes, N le nombre d'arêtes extérieures.

3.3. Groupe dual

Le dual gradué \mathcal{A} de \mathcal{H} (ou plutôt un préduel) est l'espace des $\sum \varphi_\Gamma \Gamma$ où pour chaque Γ le coefficient φ_Γ parcourt un espace de fonctions test $\mathcal{S}(E_\Gamma)$. C'est une algèbre de Hopf graduée, cocommutative, mais non commutative. On notera $\widehat{\mathcal{A}}$ le complété.

Un élément $x \in \widehat{\mathcal{A}}$ est de type groupe, resp. de type Lie (primitif) si

$$\Delta_{\mathcal{A}}x = x \otimes x \quad \text{resp.} \quad \Delta_{\mathcal{A}}x = x \otimes 1 + 1 \otimes x$$

(rappelons que le coproduit $\Delta_{\mathcal{A}}$ est dual du produit de \mathcal{H}). Les éléments de type groupe sont les caractères de \mathcal{H} ; ils forment un groupe G_F gradué pronilpotent, dont l'algèbre de Lie est constituée des éléments de type Lie.

4. RENORMALISATION

4.1. Lacet fondamental

À chaque diagramme de Feynman Γ est donc associée une distribution $U_\Gamma(\mu, d)$. Dans cette théorie, il est essentiel de ne renormaliser que des nombres purs (quantités « sans dimension »).

La dimension $\dim \Gamma$ de la distribution U_Γ résulte des règles I_1, \dots, I_5 ci-dessus. Pour un diagramme simple (1PI), compte tenu des relations de conservation de moment (on intègre sur un espace de dimension Ld), de la définition $L = I - V + 1$ et de la relation de comptage élémentaire $3V_3 + 2V_2 = 2I + N$ obtenue en énumérant par sommets ou par arêtes les couples $(\bullet \text{ ---})$ constitués d'une arête et un de ses sommets (demi-arête), on obtient :

$$(7) \quad \begin{aligned} \dim \Gamma &= \dim \prod \frac{d^d p_i}{p_i^2 + m^2} \prod_{V_3} g \mu^{3-d/2} \prod_{V_2} (m^2 \text{ ou } p^2) = \\ &= Ld - 2I + V_3 \left(3 - \frac{d}{2}\right) + 2V_2 = \left(1 - \frac{N}{2}\right)d + N. \end{aligned}$$

Noter que l'exposant de g est $V_3 = \omega(\Gamma) + N - 2$.

À chaque diagramme habillé $\varphi \otimes \Gamma$ on associe alors un nombre sans dimension, dépendant de μ et de d (d voisin de 6, $d \neq 6$) :

$$(8) \quad \langle \gamma_\mu(d), \varphi \otimes \Gamma \rangle = g^{2-N} \mu^{-B} U(\Gamma) = \langle \varphi, U_\Gamma \rangle$$

avec $B = (1 - \frac{N}{2})d + N$.⁷

On a ainsi défini un lacet $\gamma = \gamma_\mu(d)$ ou $\gamma(\mu, d)$ (méromorphe en d , dépendant de μ) autour de $d = D = 6$, à valeurs dans $G_F \subset \mathcal{A}$ (il s'agit évidemment d'un caractère de \mathcal{H}).

On pose $d = D - \epsilon$ et on note \mathcal{B} le complété, pour la graduation L , de $\mathcal{A}[\frac{1}{\epsilon}, \epsilon]$, dont γ est un élément.

Il est clair que tout élément $\varphi = \sum \varphi_j \epsilon^j \in \mathcal{B}$ se décompose additivement $\varphi = \varphi_+ - \varphi_-$ avec $\varphi_+ = \sum_{j \geq 0} \varphi_j \epsilon^j$, $\varphi_- = -\sum_{j < 0} \varphi_j \epsilon^j$ (φ_\pm est holomorphe en ϵ resp. $\frac{1}{\epsilon}$; ce sont les parties positive et négative d'une série de Laurent).

De même tout élément inversible $f \in 1 + \mathcal{B}_+$ (i.e. le terme de plus bas degré pour la graduation est $f^0 = 1$) admet une unique factorisation

$$f = f_-^{-1} f_+,$$

où de nouveau f_\pm est holomorphe en ϵ , resp. $\frac{1}{\epsilon}$, le terme de plus bas degré est 1, et $f_- = 1$ pour $\frac{1}{\epsilon} = 0$ (factorisation de Wiener-Hopf, qui est liée à la classification des fibrés sur la sphère de Riemann, et au problème de Riemann-Hilbert, mais immédiate ici dans une algèbre filtrée complète). L'unicité de la factorisation assure que f_+ et f_- sont de type groupe si f l'est, de même que, dans la version infinitésimale additive, φ_+ et φ_- sont de type Lie si φ l'est.

Le résultat principal de [CK1] est le suivant :

THÉORÈME 4.1. — *Le lacet renormalisé γ_{renorm} est la valeur pour $d = D$ ($\epsilon = 0$) du facteur γ_+ dans la factorisation de Wiener-Hopf $\gamma = \gamma_-^{-1} \gamma_+$.*

Une quantité observable se réinterprète comme série à coefficients dans \mathcal{H} ou, ce qui revient au même, comme fonction $F(\gamma)$ sur le groupe G_F :

SCHOLIE 1. — *La quantité effective renormalisée (resp. effective non renormalisée, nue) s'obtient en substituant $\gamma_{\mu+}(\epsilon)$ resp. $\gamma_\mu(\epsilon), \gamma_{\mu-}$ dans son expression en fonction de γ . La valeur renormalisée est la valeur en $\gamma_{\mu+}$ pour $\epsilon = 0$.*

(Cette dernière dépend encore de l'unité de masse μ , i.e. de l'échelle d'observation).

⁷Il est parfois utile de prolonger cette définition au cas où l'habit φ est dimensionné : on remplace alors par $B = (1 - \frac{N}{2})d + N + \dim \varphi$.

Ce théorème regroupe de façon concise et élégante toutes les règles de la renormalisation par régularisation dimensionnelle. Il est démontré dans [CK1]. Il n'est pas question de le démontrer ici, ne serait-ce que parce qu'il faudrait d'abord décrire les règles en question.

Rappelons que la régularisation dimensionnelle est essentiellement équivalente aux autres procédés de régularisation, par exemple par troncature, du moins pour les intégrales de dimension finie; elle n'est pas définie pour les intégrales de dimension infinie (qui ne le sont guère non plus), et ne s'applique donc qu'à la théorie perturbative.

4.2. Groupe de renormalisation

La L -gradation et l'unité de masse sont liées de façon étroite en vertu des règles ci-dessus. De façon plus précise, notons (comme dans [CK2]) Y l'opérateur de la gradation $L : (Y \sum f_n = \sum n f_n \text{ si } f_n \text{ est de degré } n)$. C'est une dérivation (automorphisme infinitésimal) de \mathcal{H} ou de \mathcal{A} , i.e. à la fois pour le produit et le coproduit.

PROPOSITION 4.2. — *On a $e^{tY} \gamma(\epsilon, \mu) = \gamma(\epsilon, e^t \mu)$.*

La gradation et les exposants de μ ont été choisis de sorte que cela soit évident. Le groupe de renormalisation est le groupe à 1 paramètre e^{tY} .

4.3. Contreterme, fonction β

Le résultat suivant résulte simplement des règles $I_1 - I_5$ pour les intégrales de Feynman et est démontré dans [CK2]. En fait, il s'agit plutôt d'un axiome pour la renormalisation, et il faut s'assurer qu'on a fait les bons choix pour μ :

THÉORÈME 4.3. — *Dans la factorisation $\gamma = \gamma_-^{-1} \gamma_+$ le facteur γ_- est indépendant de μ .*

Il est alors important d'analyser quels sont les contretermes γ_- possibles. L'inverse $\varphi(\frac{1}{\epsilon}) = \gamma_-^{-1}$ a la propriété suivante : son terme de degré 0 est 1, il est de type groupe, et on a

$$(9) \quad e^{tY}(\varphi) = \varphi F_t(\epsilon),$$

où $F_t \in \mathcal{A}[[\epsilon]]$ est holomorphe en ϵ . F_t vérifie évidemment la condition de cocycle

$$F_{t+s} = F_t e^{tY}(F_s).$$

En particulier, pour $\epsilon = 0$, il définit un groupe à un paramètre

$$(10) \quad F_t|_{\epsilon=0} = e^{\beta t} \quad \text{avec} \quad \beta \in \mathcal{A}_+.$$

L'élément primitif $\beta \in \mathcal{A}_+$ est le «résidu» de la théorie.

Dérivant (9) pour $t = 0$ on obtient $\epsilon \varphi^{-1} \cdot Y \varphi = \frac{d}{dt} F_t(\epsilon)|_{t=0}$. Le premier membre de cette égalité est holomorphe en $\frac{1}{\epsilon}$, le second en ϵ , donc les deux sont constants et égaux à β :

$$(11) \quad Y \varphi = \varphi \frac{\beta}{\epsilon}.$$

PROPOSITION 4.4. — Soient \mathcal{C} le complété pour la graduation L de $\mathcal{A}[\frac{1}{\epsilon}]$, et $\alpha \in \mathcal{C}_+$; alors l'équation $Y\varphi = \varphi\alpha$ admet pour unique solution $\varphi \in 1 + \mathcal{C}_+$ la série

$$(12) \quad \varphi = (1 - Y^{-1}R_\alpha)^{-1}(1)$$

où R_α désigne la multiplication à droite par α . Alors φ est de type groupe ssi α est de type Lie.

Preuve : Y est inversible sur \mathcal{C}_+ ; l'équation équivaut donc à

$$Y(1 - Y^{-1}R_\alpha)\varphi = 0, \quad \text{i.e.} \quad (1 - Y^{-1}R_\alpha)\varphi = \text{constante} = 1$$

puisque le terme de degré 0 de φ est 1. On a alors aussi

$$Y(\varphi \otimes \varphi) = (\varphi \otimes \varphi)(\alpha \otimes 1 + 1 \otimes \alpha), \quad Y(\Delta\varphi) = \Delta\varphi \cdot \Delta\alpha,$$

donc $\Delta\varphi = \varphi \otimes \varphi$ ssi $\Delta\alpha = \alpha \otimes 1 + 1 \otimes \alpha$.

Ainsi, les contretermes possibles dans une théorie de renormalisation comme ci-dessus sont de la forme $(1 - Y^{-1}R_\alpha)^{-1}(1)$, avec $\alpha = \frac{\beta}{\epsilon}$, $\beta \in \mathcal{A}_+$ de type Lie.

Dans [CK2] on trouve la description suivante du contreterme : adjoignons à \mathcal{A} un nouvel élément Z_0 avec la relation $\text{ad } Z_0 = Y$. On a alors $Z_0\varphi = \varphi(Z_0 + \frac{\beta}{\epsilon})$ d'où $e^{tZ_0}\varphi = \varphi e^{t(Z_0 + \frac{\beta}{\epsilon})}$. Compte tenu que $e^{-tZ_0}\varphi e^{tZ_0} = e^{-tY}(\varphi) = \sum e^{-nt}\varphi_n \rightarrow 1$ pour $t \rightarrow \infty$, on obtient :

$$(13) \quad \varphi = e^{-tY}(\varphi) e^{-tZ_0} e^{t(Z_0 + \frac{\beta}{\epsilon})} = \lim_{t \rightarrow \infty} e^{-tZ_0} e^{t(Z_0 + \frac{\beta}{\epsilon})}.$$

L'élément $\beta \in \mathcal{A}$ qui apparaît dans la théorie est étroitement lié à la « fonction β » des physiciens.

5. CONSTANTE DE STRUCTURE ET FORMULES UNIVERSELLES.

5.1. Diagrammes réduits (cas $m = 0$)

Dans ce dernier numéro, on se limite à des diagrammes réduits correspondant au cas $m = 0$. Dans ce cas limite, on peut effacer tous les sommets de type 2 : ceux de couleur 0 produisent un facteur $m^2 = 0$, donc ne contribuent plus du tout, et ceux de couleur 1 produisent un facteur p^2 qui neutralise le facteur $\frac{1}{p^2+m^2}$ d'une des arêtes adjacentes, donc on peut aussi les effacer. La règle de réinsertion définissant le coproduit doit être modifiée : elle reste la même pour les diagrammes à trois pattes, mais les diagrammes à deux pattes sont réinsérés au milieu de n'importe quelle arête intérieure. On peut en outre oublier les habits puisqu'il ne reste qu'une couleur et un habit pour la réinsertion des diagrammes à deux pattes.

Notons \mathcal{H}_r et G_r l'algèbre de Hopf et le groupe correspondants (ce sont respectivement une sous-algèbre de \mathcal{H} et un quotient de G_F).

On repère les éléments primitifs de \mathcal{H}_r^* comme suit : à tout diagramme simple Γ on associe l'élément

$$(14) \quad \underline{\Gamma} = S(\Gamma) p\Gamma \in \mathcal{H}^*,$$

où $S(\Gamma)$ est le nombre d'automorphismes de Γ et p désigne le projecteur canonique sur les éléments primitifs (si $x \in \mathcal{H}_+$, px est l'unique élément primitif égal à x mod. \mathcal{H}_+^{*2}). Suivant F. Patras [Pat1, Pat2] p se calcule ainsi : si $a, b \in L(\mathcal{H}^*)$, on pose $a*b = M \circ (a \otimes b) \circ \Delta$ (où $M : \mathcal{H}^* \otimes \mathcal{H}^* \rightarrow \mathcal{H}^*$ est la multiplication) ; il s'agit d'un produit associatif, qui prolonge celui de l'algèbre de Hopf $\mathcal{H}^* \otimes \mathcal{H}$, et pour lequel l'unité est le projecteur p_0 sur le corps de base (la counité) ; on a $Id = p_0 + p_+$ où p_+ est le projecteur sur l'idéal des éléments de degré > 0 de \mathcal{H}^* . On a alors $p = \text{Log}^* Id = \sum \frac{(-1)^{k-1}}{k} p_+^{*k}$.

Les $\underline{\Gamma}$ forment une base de l'algèbre de Lie $\mathcal{G}_r \subset \mathcal{H}_r^*$, et la règle de réinsertion définissant le coproduit donne

$$(15) \quad [\underline{\Gamma}, \underline{\Gamma}'] = \sum_{v'} \underline{\Gamma'} \circ^{v'} \underline{\Gamma} - \sum_v \underline{\Gamma} \circ^v \underline{\Gamma}' ,$$

où v' resp. v parcourt l'ensemble des sommets (resp. lignes pour les diagrammes à deux pattes) et \circ^v signifie qu'on greffe le diagramme de gauche à la v -ième place du diagramme de droite.

Pour un diagramme simple et réduit à N pattes, on a $V = 2(L - 1) + N$, $I = 3(L - 1) + N$. Observons qu'il y a (à un isomorphisme près) un seul diagramme simple $\Gamma_{2,L}$ à deux pattes de degré L , resp. $\Gamma_{3,L}$ à trois pattes, et pour ceux-ci on a $I = 3L - 1, V = 2L$, resp. $I = 3L, V = 2L + 1$; en outre, le degré du diagramme greffé $\Gamma \circ^v \Gamma'$ est toujours $L + L'$.

On obtient donc pour les crochets de diagrammes simples à 2 ou 3 pattes :

$$(16) \quad \begin{aligned} [\underline{\Gamma}_{2,L}, \underline{\Gamma}_{2,L'}] &= 3(L' - L) \underline{\Gamma}_{2,L+L'} \\ [\underline{\Gamma}_{3,L}, \underline{\Gamma}_{3,L'}] &= 2(L' - L) \underline{\Gamma}_{3,L+L'} \\ [\underline{\Gamma}_{3,L}, \underline{\Gamma}_{2,L'}] &= 3L' \underline{\Gamma}_{3,L+L'} - 2L \underline{\Gamma}_{2,L+L'}. \end{aligned}$$

5.2. Groupe des difféomorphismes formels de la droite

Notons G_2 le groupe des difféomorphismes formels tangents à l'identité sur la droite :

$$\varphi = X + \sum_2^\infty a_n X^n.$$

Son algèbre de Hopf \mathcal{H}_2 est l'algèbre des polynômes en les $a_n = a_n(\varphi)$; si T désigne la série de Taylor $X + \sum a_n X^n$ ($T_\varphi = \langle T, \varphi \rangle = \varphi(X)$), le coproduit est défini par

$$\langle \Delta T, \varphi \otimes \psi \rangle = T_\varphi \circ T_\psi.$$

C'est un groupe gradué, la graduation correspondant à l'action du groupe des homothéties (a_n est de degré $n-1$). Son algèbre de Lie admet pour base graduée les champs de vecteurs

$$Z'_k = X^{k+1} \frac{\partial}{\partial X}.$$

Z_k est de degré k et on a

$$(17) \quad [Z'_p, Z'_q] = (p - q)Z'_{p+q}.$$

Les formules de crochet (16) impliquent :

PROPOSITION 5.1. — *Posons $\rho_\Gamma = 1$ si Γ a trois pattes, $\rho_\Gamma = \frac{3}{2}$ s'il en a deux ; il existe un unique homomorphisme ρ d'algèbres de Lie ou de groupes formels $G_r \rightarrow G_2$ tel que*

$$(18) \quad \rho(\underline{\Gamma}) = \rho_\Gamma Z'_{2L} \quad (L = L(\Gamma)).$$

5.3. Constante de couplage et difféomorphismes

La renormalisation a pour effet de remplacer les quantités observables par des séries de diagrammes (dépendant de ϵ et μ); en particulier, à la constante de couplage g est associée la série génératrice

$$(19) \quad g_0 = X + \sum_2^\infty \alpha_n X^n = (gZ_1)(Z_3)^{-\frac{3}{2}} \quad (\text{série impaire})$$

avec

$$(20) \quad gZ_1 = \left(X + \sum_{\Gamma \in F_3} X^{2L+1} \frac{\Gamma}{S(\Gamma)} \right) \quad Z_3 = \left(\left(1 - \sum_{\Gamma \in F_2} X^{2L} \frac{\Gamma}{S(\Gamma)} \right) \right)$$

où F_2 resp. F_3 désigne l'ensemble des diagrammes simples et réduits à deux resp. trois pattes ($2L$, resp. $2L+1$ est l'exposant de g dans ces diagrammes); $S(\Gamma)$ est le nombre d'automorphismes de Γ .

La série génératrice g_0 définit une fonction $\Phi^* : G_r \rightarrow G_2$: si $m \in G_r$, $\Phi^*(m) = \varphi(X)$ est l'évaluation en m de la série $g_0(X)$ (de façon duale : un homomorphisme d'algèbres $\Phi : \mathcal{H}_2 \rightarrow \mathcal{H}_r$ qui envoie a_n sur le coefficient α_n).

THÉORÈME 5.2. — *La fonction $\Phi^* : G_F \rightarrow G_2$ définie par la série génératrice g_0 est un homomorphisme de groupes (de façon duale : Φ est un homomorphisme d'algèbres de Hopf).*

En fait, Connes et Kreimer montrent qu'on a $\Phi^* = \rho$, en prouvant que les deux transportent de la même façon le champ de vecteurs invariant à droite ∂_Z associé à un élément primitif $Z \in \mathcal{H}_r$ ($\partial_Z a = \langle Z \otimes Id, \Delta a \rangle$); il suffit de vérifier ceci quand Z parcourt l'ensemble des éléments basiques $\underline{\Gamma}$.

5.4. Remarques finales

L'homomorphisme ρ défini précédemment est compatible avec les graduations (plus exactement, il est homogène de degré 2) puisque a_{n+1} et $X^{n+1} \frac{\partial}{\partial X}$ sont de degré n). Le paramètre g est essentiellement reflété dans la graduation, car on a, élémentairement, en remettant dans la notation le paramètre g qui n'y figurait pas : $\gamma_{\mu,g}(\epsilon) = e^{2tY} \gamma_{\mu,1}(\epsilon)$.

On montre que l'image $\rho(\beta)$ de l'élément générateur β du n° 4.3 est exactement la « fonction » β des physiciens, d'où la notation.

Le théorème 5.2 implique que la renormalisation pour g_{eff} se lit directement dans la factorisation $g_{eff} = g_{eff-}^{-1} g_{eff+}$ dans le groupe G_2 , et G_r n'intervient plus. En ce sens G_2 apparaît comme groupe universel pour ces calculs, et on peut imaginer qu'un groupe des difféomorphismes à plusieurs variables intervient de même pour une théorie de champ vectorielle.

Le lacet universel $\gamma_\mu(d)$ définit, par recollement via g_{eff} un fibré P sur la droite projective P_1 , de fibre la droite formelle X . Ici, comme probablement dans toute théorie perturbative, le lacet se factorise et ce fibré est trivialisable; on peut imaginer que dans une théorie non perturbative, comme dans le problème de Riemann-Hilbert non nilpotent, interviendront de tels fibrés non triviaux.

BIBLIOGRAPHIE

- [CK1] Connes, A.; Kreimer, D. – *Renormalization in quantum field theory and the Riemann-Hilbert problem I : The Hopf algebra structure of graphs and the main theorem*. Comm. Math. Phys. 210 (2000), no. 1, 249–273.
- [CK2] Connes, A.; Kreimer, D. – *Renormalization in quantum field theory and the Riemann-Hilbert problem. II: the β -function, diffeomorphisms and the renormalization group*. Comm. Math. Phys. 216 (2001), no. 1, 215–241.
- [BK1] Broadhurst, D.J.; Kreimer, D. – *Knots and numbers in ϕ^4 theory to 7 loops and beyond*. Internat. J. Modern Phys. C 6 (1995), no. 4, 519–524.
- [BK2] Broadhurst, D.J.; Kreimer, D. – *Association of multiple zeta values with positive knots via Feynman diagrams up to 9 loops*. Phys. Lett. B 393 (1997), no. 3-4, 403–412.
- [BK3] Broadhurst, D.J.; Gracey, J.A.; Kreimer, D. – *Beyond the triangle and uniqueness relations: non-zeta counterterms at large N from positive knots*. Z. Phys. C 75 (1997), no. 3, 559–574.
- [BK4] Broadhurst, D.J.; Kreimer, D. – *Feynman diagrams as a weight system: four-loop test of a four-term relation*. Phys. Lett. B 426 (1998), no. 3-4, 339–346.

- [BK5] Broadhurst, D.J.; Kreimer, D. – *Renormalization automated by Hopf algebra*. J. Symbolic Comput. 27 (1999), no. 6, 581–600; hep-th/9810087.
- [BK6] Broadhurst, D.J.; Kreimer, D. – *Combinatoric explosion of renormalization tamed by Hopf algebra: 30-loop Padé-Borel resummation*. Phys. Lett. B 475 (2000), no. 1-2, 63–70; hep-th/9912093.
- [BK7] Broadhurst, D.J.; Kreimer, D. – *Towards cohomology of renormalization: bi-grading the combinatorial Hopf algebra of rooted trees*. Comm. Math. Phys. 215 (2000), no. 1, 217–236.
- [CK3] Connes, A.; Kreimer, D. – *Hopf algebras, renormalization and noncommutative geometry*. Comm. Math. Phys. 199 (1998), no. 1, 203–242.
- [CK4] Connes, A.; Kreimer, D. – *Hopf algebras, renormalization and noncommutative geometry*. Quantum field theory: perspective and prospective (Les Houches, 1998), 59–108, NATO Sci. Ser. C Math. Phys. Sci., 530, Kluwer Acad. Publ., Dordrecht (1999).
- [CK5] Connes, A.; Kreimer, D. – *Renormalization in quantum field theory and the Riemann-Hilbert problem*. J. High Energy Phys. 1999, no. 9, Paper 24, 8 pp. (electronic); hep-th/9909126.
- [CK6] Connes, A.; Kreimer, D. – *Lessons from quantum field theory: Hopf algebras and spacetime geometries*. Moshé Flato (1937–1998). Lett. Math. Phys. 48 (1999), no. 1, 85–96; hep-th/9904044.
- [CK7] Connes, A.; Kreimer, D. – *From local perturbation theory to Hopf- and Lie-algebras of Feynman graphs*. Mathematical physics in mathematics and physics (Siena, 2000), 105–114, Fields Inst. Commun. 30, Amer. Math. Soc. (2001).
- [DK] Delbourgo, R.; Kreimer, D. – *Using the Hopf algebra structure of QFT in calculations*. Phys. Rev. D (3) 60 (1999), no. 10; hep-th/9903249.
- [K1] Kreimer, D. – *Renormalization and knot theory*. J. Knot Theory and Ramifications 6 (1997), no. 4, 479–581.
- [K2] Kreimer, D. – *On the Hopf algebra structure of perturbative quantum field theories*. Adv. Theor. Math. Phys. 2 (1998), no. 2, 303–334.
- [K3] Kreimer, D. – *On overlapping divergences*. Comm. Math. Phys. 204 (1999), no. 3, 669–689; hep-th/9810022.
- [K4] Kreimer, D. – *Chen’s iterated integral represents the operator product expansion*. Adv. Theor. Math. Phys. 3 (1999), no. 3; hep-th/9901099.
- [K5] Kreimer, D. – *Knots and Feynman diagrams*. Cambridge Lecture Notes in Physics, 13. Cambridge University Press, Cambridge (2000).
- [K6] Kreimer, D. – *Shuffling quantum field theory*. Lett. Math. Phys. 51 (2000), no. 3, 179–191.

Renormalisation

- [Coll] Collins, J. – *Renormalization*,. Cambridge monographs in math. phys., Cambridge University Press, Cambridge (1984).
- [Dres] Dresden, M. – *Renormalization in historical perspective - The first stage*. in *Renormalization*, ed. L. Brown, Springer-Verlag, New York, Berlin, Heidelberg (1994).
- [Drou] Drouffe, J.-M.; Itzykson, C. – *Théorie statistique des champs (2 volumes)*. Savoirs actuels, InterEditions/Editions du C.N.R.S. (1989).
- [EG] Epstein, H.; Glaser, V. – *The role of locality in perturbation theory*. Ann. Inst. H. Poincaré A19 (1973), 211-295.
- [FMRS] Feldman, J.; Magnen, J.; Rivasseau, V.; Seneor, R. – *Massive Gross-Neveu model: a rigorous perturbative construction*. Phys. Rev. Lett. 54 (1985).
- [GK] Gawedski, K.; Kupianen, A. – *Exact renormalization of the Gross-Neveu model of quantum fields*. Phys. Rev. Lett 54 (1985).
- [GJ] Glimm, J.; Jaffe, A. – *Quantum Physics*. Springer-Verlag, New York, Berlin, Heidelberg (1987).
- [LeBe] Le Bellac, M. – *Des phénomènes critiques aux champs de jauge*. Savoirs actuels, InterEditions/Editions du C.N.R.S. (1988).
- [ZJ] Zinn-Justin, J. – *Quantum Field Theory and Critical Phenomena*. International series of monographies on physics 92, Oxford science publications (1996).

Wiener-Hopf, Riemann-Hilbert, algèbres de Hopf

- [Beau] Beauville, A. – *Monodromie des systèmes différentiels linéaires à pôles simples sur la sphère de Riemann*. Sémin. Bourbaki, exp. n° 765 (mars 1993), Astérisque 216 (1993), 103-119.
- [Boli] Bolibruch, A. – *Fuchsian systems with reducible monodromy and the Riemann-Hilbert problem*. Lecture Notes in Math. 1250 (1992), 139-155.
- [BKI] Bourbaki, N. – *Éléments de mathématique. Algèbre. Chapitres 1 à 3*. Masson, Paris 1982.
- [ENS] *Mathématique et Physique, Séminaire de l'E.N.S. 1979-82*, L. Boutet de Monvel, A. Douady, J.-L. Verdier ed. – Progress in Math. vol. 35, Birkhäuser (1983).
- [Drin] Drinfel'd, V. G. – *Almost cocommutative Hopf algebras*. Algebra i Analiz 1 (1989), no. 2, 30–46; translation in Leningrad Math. J. 1 (1990), no. 2, 321–342.
- [WH] *Constructive methods of Wiener-Hopf factorization*. Edited by I. Gohberg and M. A. Kaashoek, Operator Theory: Advances and Applications, 21. Birkhäuser Verlag, Basel (1986).
- [LP] Lappo-Danilevskii, I. – *Mémoire sur la théorie des systèmes d'équations différentielles linéaires*. Chelsea, New York (1953).

- [Pat1] Patras, F., Reutenauer, C. – *Higher Lie idempotents*. J. Algebra 222 (1999), no. 1, 51–64.
- [Pat2] Patras, F. – *La décomposition en poids des algèbres de Hopf*. Ann. Inst. Fourier 43 (1993), no. 4, 1067–1087.
- [Spe] Speck, F.-O. – *General Wiener-Hopf factorization methods. With a foreword by E. Meister*. Research Notes in Mathematics, 119. Pitman (Advanced Publishing Program), Boston, MA (1985).

Général

- [BCE] Brodzki, J.; Connes, A., Ellwood, D. – *Polarized modules and Fredholm modules*. Mat. Fiz. Anal. Geom. 2 (1995), no. 1, 15–24.
- [C1] Connes, A. – *Noncommutative geometry*. – Academic Press, Inc., San Diego, CA, 1994 (661 pp.).
- [C2] Connes, A. – *Non-commutative geometry and physics*. Gravitation et quantifications (Les Houches, 1992), 805–950, North-Holland, Amsterdam, 1995.
- [C3] Connes, A. – *Géométrie non commutative et physique quantique*. *Mathématiques quantiques*. 20 pp., SMF Journ. Annu., 1992, Soc. Math. France, Paris, 1992.
- [C4] Connes, A. – *The action functional in noncommutative geometry*. Comm. Math. Phys. 117 (1988), no. 4, 673–683.
- [CM] Connes, A.; Moscovici, H. – *Hopf algebras, cyclic cohomology and the transverse index theorem*. Comm. Math. Phys. 198 (1998), no. 1, 199–246.
- [CS] Connes, A.; Størmer, E. – *A connection between the classical and the quantum mechanical entropies*. *Operator algebras and group representations*. Vol. I (Neptun, 1980), 113–123, Monographs Stud. Math., 17, Pitman, Boston, Mass.-London (1984).

Louis BOUTET de MONVEL

Université de Paris VI, analyse algébrique

Institut de Math. de Jussieu

UMR 7586 du CNRS

Case 82, 4 place Jussieu

F-75252 Paris Cedex 05

E-mail : boutet@math.jussieu.fr

COBORDISME DES VARIÉTÉS ALGÈBRIQUES
[d'après M. Levine et F. Morel]

par **François LOESER**

INTRODUCTION

Le but de cet exposé est de décrire les travaux de Levine et Morel concernant une nouvelle théorie (co)homologique des variétés algébriques, le cobordisme algébrique, ainsi que ses applications à la formule du degré de Rost. Traditionnellement, deux variétés \mathcal{C}^∞ compactes sans bord de dimension n sont dites cobordantes si leur somme disjointe est le bord d'une variété \mathcal{C}^∞ compacte de dimension $n+1$. On définit ainsi une relation d'équivalence et l'ensemble des classes d'équivalence peut être naturellement muni d'une structure d'anneau gradué (la graduation provenant de la dimension), l'anneau de cobordisme réel N_* . Cet anneau a été introduit par Thom dans [21], qui a démontré que les N_n sont isomorphes aux groupes d'homotopie stable des «espaces de Thom » $MO(n)$ et que N_* est isomorphe à une algèbre de polynômes sur $\mathbf{Z}/2\mathbf{Z}$. En considérant de façon similaire des variétés \mathcal{C}^∞ stablement presque complexes, on définit l'anneau de cobordisme complexe U_* dont la construction est rappelée dans la section 1. Milnor a démontré dans [17] que U_* est isomorphe à une algèbre de polynômes sur \mathbf{Z} et qu'elle est engendrée comme algèbre par les classes de certaines variétés projectives complexes, en fait les espaces projectifs complexes et les hypersurfaces de bidegré $(1,1)$ dans le produit de deux espaces projectifs complexes (cf. 2.3.9).

Dans un travail fondamental [20], Quillen démontre plus précisément que U_* est canoniquement isomorphe à l'anneau de Lazard \mathbf{L}_* classifiant les lois de groupe formel commutatif de rang 1. Plus généralement, il considère les théories homologiques et cohomologiques $X \mapsto U_*(X)$ et $X \mapsto U^*(X)$ représentées par le spectre de Thom complexe. Que U_* soit canoniquement isomorphe à l'anneau de Lazard \mathbf{L}_* peut être compris comme le fait que, dans la théorie U_* , la classe de Chern

du produit tensoriel de deux fibrés en droite est «la plus compliquée possible » contrairement à ce qui se passe en homologie singulière où $c_1(L \otimes M) = c_1(L) + c_1(M)$. Ceci reflète une propriété plus générale de U_* qui est d'être universelle parmi «les théories homologiques avec orientation complexe ».

En géométrie algébrique, pour les variétés sur un corps k , il existe un analogue naturel du cobordisme qui consiste à considérer un morphisme projectif $W \rightarrow \mathbf{A}^1$, avec, disons, W lisse et irréductible, qui est transverse à $\{0\}$ et à $\{1\}$. Les fibres W_0 et W_1 correspondantes sont alors lisses sur k et on aurait envie de dire qu'elles sont algébriquement cobordantes. La théorie que l'on obtient (cf. 2.3.3) de cette façon est malheureusement bien trop naïve : par exemple deux courbes cobordantes en ce sens ont toujours même genre, ce qui ne se produit pas avec la théorie Ω_* de Levine et Morel dont une des propriétés fondamentales est que sa valeur $\Omega_*(\text{Spec } k)$ sur le point est canoniquement isomorphe à l'anneau de Lazard \mathbf{L}_* . De plus, comme on le verra au cours de l'exposé, la théorie homologique Ω_* est également universelle parmi les théories homologiques «orientées » pour les variétés algébriques sur un corps k .

Pour conclure remarquons que, comme il l'exprime lui-même dans l'introduction de [20], le travail de Quillen est déjà fortement influencé par les travaux de Grothendieck en géométrie algébrique : «I have been strongly influenced by Grothendieck's theory of motives in algebraic geometry ... and like to think of a cobordism theory as a universal contravariant functor on the category of \mathcal{C}^∞ manifolds endowed with Gysin homomorphisms for a class of proper «oriented maps »... ».

J'ai bénéficié de l'aide et des conseils de J.-B. Bost, A. Chambert-Loir, M. Levine et F. Morel pendant la préparation de cet exposé. Qu'ils en soient amicalement remerciés ici.

Table des matières

Introduction	1
1. Préliminaires sur le cobordisme complexe	3
2. Construction du cobordisme algébrique	6
3. Propriétés fondamentales	13
4. Théorèmes de comparaison	16
5. La formule du degré	21
6. Images inverses en cobordisme algébrique	24
7. Relations avec la théorie homotopique des schémas	26
Références	28

1. PRÉLIMINAIRES SUR LE COBORDISME COMPLEXE

1.1. Cobordisme complexe

La notion d'orientation complexe pour un morphisme $f : Z \rightarrow X$ de variétés \mathcal{C}^∞ généralise celle de structure faiblement complexe sur Z lorsque X est un point. Soit f un tel morphisme. On suppose que les variétés Z et X sont équidimensionnelles, et on définit la dimension relative de f comme $\dim f := \dim Z - \dim X$. Commençons par supposer que f est de dimension relative paire. Dans ce cas, une orientation complexe de f consiste en la donnée d'une classe d'équivalence de factorisations de f en le composé d'un morphisme structural de fibré complexe $p : E \rightarrow X$ et d'un plongement $i : Z \rightarrow E$ dont le fibré normal est muni d'une structure complexe. Deux factorisations $f = p \circ i$ et $f = p' \circ i'$ sont considérées comme équivalentes si les fibrés complexes E et E' peuvent être plongés dans le même fibré complexe E'' de façon que, dans E'' , i et i' soient isotopes de façon compatible avec la structure complexe dont sont munis leurs fibrés normaux respectifs. Quand f est de dimension relative impaire, on remplace E par $E \times \mathbf{R}$ dans la définition précédente. On dira morphisme orienté pour morphisme muni d'une orientation complexe.

Deux morphismes propres orientés $f_1 : Z_1 \rightarrow X$ et $f_2 : Z_2 \rightarrow X$ sont dits cobordants s'il existe un morphisme propre orienté $b : W \rightarrow X \times \mathbf{R}$ tel que les morphismes $\epsilon_i : X \rightarrow X \times \mathbf{R}$ qui à x associent (x, i) soient transverses à b et que le pull-back de b par les ϵ_i soient isomorphes comme morphismes orientés aux f_i .

Le cobordisme est une relation d'équivalence et on note $U^q(X)$ l'ensemble des classes de cobordisme de morphismes propres orientés de dimension relative $-q$. Il sera aussi commode de passer en notation homologique en posant $U_q(X) := U^{\dim X - q}(X)$.

1.1.1. Le cobordisme U^* est un foncteur contravariant : si $g : Y \rightarrow X$ est un morphisme de variétés \mathcal{C}^∞ et $f : Z \rightarrow X$ un morphisme propre orienté, quitte à bouger g par une homotopie, on peut supposer, par le théorème d'isotopie de Thom que le morphisme g est transverse à f . La classe de cobordisme du morphisme $Y \times_X Z \rightarrow Y$ ne dépend que de celle de f , et on obtient ainsi un morphisme

$$g^* : U^q(X) \rightarrow U^q(Y).$$

1.1.2. Tout morphisme propre orienté $g : X \rightarrow Y$ de dimension relative d définit un morphisme de Gysin

$$g_* : U^q(X) \rightarrow U^{q-d}(Y),$$

par composition.

1.1.3. Le produit cartésien induit un produit externe naturel

$$\otimes : U^*(X_1) \times U^*(X_2) \rightarrow U^*(X_1 \times X_2).$$

On en tire une structure d'anneau gradué sur $U^*(X)$ en posant $x_1 \cdot x_2 := \Delta^*(x_1 \otimes x_2)$, Δ désignant le morphisme diagonal. L'unité est la classe de cobordisme $1 = 1_X$ du morphisme identité $X \rightarrow X$.

1.2. L'anneau de Lazard

Rappelons qu'une loi de groupe formel commutatif de rang 1 à coefficients dans un anneau commutatif A est une série formelle

$$(1.2.1) \quad F(u, v) = \sum_{(i,j) \in \mathbf{N}^2} a_{i,j} u^i v^j$$

appartenant à $A[[u, v]]$ et satisfaisant les relations

$$(1.2.2) \quad F(u, 0) = F(0, u) = u$$

$$(1.2.3) \quad F(u, v) = F(v, u)$$

et

$$(1.2.4) \quad F(u, F(v, w)) = F(F(u, v), w).$$

D'après Lazard [9], il existe une loi de groupe formel commutatif de rang 1 universelle $(\mathbf{L}, F_{\mathbf{L}})$. La construction de $(\mathbf{L}, F_{\mathbf{L}})$ est très simple. On considère l'anneau $\tilde{\mathbf{L}}$ des polynômes à coefficients entiers en les variables $A_{i,j}$, pour i, j dans \mathbf{N} et la série formelle universelle $\tilde{F} = \sum_{i,j} A_{i,j} u^i v^j$ dans $\tilde{\mathbf{L}}[[u, v]]$. Il suffit alors de définir \mathbf{L} comme le quotient de l'anneau $\tilde{\mathbf{L}}$ par les relations obtenues en imposant les relations (1.2.2), (1.2.3) et (1.2.4) à \tilde{F} et de définir la série $F_{\mathbf{L}} := \sum_{i,j} a_{i,j} u^i v^j$ comme l'image de \tilde{F} dans $\mathbf{L}[[u, v]]$. On munit l'anneau \mathbf{L} d'une structure d'anneau gradué \mathbf{L}_* en donnant le degré $i + j - 1$ à $a_{i,j}$. On aura aussi à considérer l'anneau gradué \mathbf{L}^* défini par $\mathbf{L}^n := \mathbf{L}_{-n}$. Un fait remarquable, démontré par Lazard [9], est que l'anneau de Lazard \mathbf{L} est un anneau de polynômes à coefficients entiers en un nombre dénombrable de variables.

1.3. Le théorème de Quillen

Soit E un fibré vectoriel complexe de rang n sur une variété X . On note $i : X \rightarrow E$ le morphisme donné par la section nulle. La classe d'Euler de E est l'élément

$$(1.3.1) \quad e(E) (= c_n(E)) := i^* i_* 1$$

de $U^{2n}(X)$.

Notons $q : \mathbf{P}(E) \rightarrow X$ le fibré projectif représentant les quotients de rang 1 de E et γ le fibré en droites sur $\mathbf{P}(E)$ quotient canonique de $q^*(E)$. Dans les sections suivantes ce fibré en droites sera parfois également noté $O(1)$ (c'est le fibré qui a des sections globales et non son dual). On note ξ la classe d'Euler du fibré γ . On vérifie alors par un calcul classique, cf. [8], que $U^*(\mathbf{P}(E))$ est, via q^* , un $U^*(X)$ -module libre de base $1, \xi, \dots, \xi^{n-1}$. De plus si E est somme directe de fibrés en droites L_1, \dots, L_n , on a

$$(1.3.2) \quad \prod_{i=1}^n (\xi - e(L_i)) = 0.$$

On en tire en particulier que

$$(1.3.3) \quad U^*(\mathbf{P}^n) \simeq U^*(\text{point})[u]/u^{n+1}$$

avec $u = e(\gamma)$ et que

$$(1.3.4) \quad U^*(\mathbf{P}^n \times \mathbf{P}^n) \simeq U^*(\text{point})[u, v]/(u^{n+1}, v^{n+1}),$$

avec $u = e(\text{pr}_1^*(\gamma))$ et $v = e(\text{pr}_2^*(\gamma))$. On peut alors construire une loi de groupe formel commutative sur $U^*(\text{point})$ de la façon suivante. D'après (1.3.4), on peut écrire

$$(1.3.5) \quad e(\text{pr}_1^*(\gamma) \otimes \text{pr}_2^*(\gamma)) = \sum_{i,j \leq n} c_{i,j}^n u^i v^j.$$

De plus $c_{i,j}^n$ ne change pas quand on augmente n . En faisant tendre n vers l'infini on en tire une série formelle $F(u, v)$ dans $U^*(\text{point})[[u, v]]$ de la forme

$$(1.3.6) \quad F(u, v) = \sum_{i,j} c_{i,j} u^i v^j$$

avec $c_{i,j}$ dans $U^{2-2i-2j}(\text{point})$. On vérifie facilement que c'est une loi de groupe formel sur $U^*(\text{point})$. Par exemple, on démontre l'associativité (1.2.4) en évaluant de deux façons différentes la classe d'Euler de $\text{pr}_1^*(\gamma) \otimes \text{pr}_2^*(\gamma) \otimes \text{pr}_3^*(\gamma)$ sur $\mathbf{P}^n \otimes \mathbf{P}^n \otimes \mathbf{P}^n$.

Comme tout fibré en droites complexes sur une variété X est image inverse de γ pour un morphisme $X \rightarrow \mathbf{P}^n$ convenable, on en déduit que pour toute paire de fibrés en droites complexes E_1 et E_2 sur la variété X on a

$$(1.3.7) \quad e(L_1 \otimes L_2) = F(e(L_1), e(L_2)).$$

Par la propriété universelle de l'anneau de Lazard, on dispose d'un morphisme canonique $\delta : \mathbf{L}^* \rightarrow U^*(\text{point})$.

1.3.1. THÉORÈME (Quillen [20]). — *Le morphisme canonique $\delta : \mathbf{L}^* \rightarrow U^*$ (point) est un isomorphisme d'anneaux gradués.*

Dans cet isomorphisme, les coefficients $a_{i,j}$ de la série de Lazard admettent une interprétation géométrique en terme des classes des espaces projectifs et de celles des hypersurfaces de Milnor $H_{i,j}$ (cf. 2.3.9, la proposition 2.3.10 restant bien sûr vraie en remplaçant Ω_* par U_* ; voir aussi [2] Proposition 10.6). La preuve de Quillen a été exposée dans ce séminaire par M. Karoubi [8] (voir aussi [2]). Elle utilise de façon essentielle que $U^q(X)$ est de type fini lorsque X a le type d'homotopie d'un CW-complexe fini, ce qui résulte de l'interprétation homotopique à la Thom (Théorème 7.1.1) et des énoncés de finitude usuels en homotopie, le lien entre opérations de Steenrod en cobordisme complexe et opérations de Landweber-Novikov ainsi qu'un calcul explicite faisant intervenir les espaces lenticulaires.

2. CONSTRUCTION DU COBORDISME ALGÈBRIQUE

2.1. Notations

On fixe un corps k . On note \mathcal{L}_k la catégorie des variétés lisses quasi-projectives sur k et \mathcal{V}_k celle des schémas de type fini sur k . On note \mathcal{L}'_k et \mathcal{V}'_k les catégories formées des mêmes objets mais où les morphismes sont les morphismes projectifs. On dira parfois variété lisse pour variété quasi-projective lisse.

2.2. Cycles de cobordisme

2.2.1. Soit X dans \mathcal{V}_k . On note $\mathcal{M}_*^+(X)$ le groupe abélien libre sur les classes d'isomorphisme de cycles $f : Y \rightarrow X$ avec f projectif et Y lisse et intègre sur k , gradué par la dimension de Y .

On désire définir le groupe de cobordisme algébrique $\Omega_*(X)$ comme quotient de $\mathcal{M}_*^+(X)$ par une relation d'équivalence convenable, le «cobordisme algébrique».

Pour des raisons de commodité technique, en particulier pour disposer dès le départ de classes de Chern de fibrés en droites, il est préférable de considérer la donnée supplémentaire de familles de fibrés en droites sur Y . Ainsi on définit un cycle de cobordisme sur X comme la donnée d'un morphisme $f : Y \rightarrow X$ avec f projectif et Y lisse et intègre sur k ainsi que de r fibrés en droites L_1, \dots, L_r sur Y , avec r éventuellement nul. On a une notion évidente d'isomorphisme pour les cycles de cobordisme sur X (on tolère le renumérotage des L_i). La dimension d'un tel cycle est l'entier $\dim_k(Y) - r$. On notera 1_X le cycle $\text{Id} : X \rightarrow X$ et $1 = 1_{\text{Spec } k}$.

2.2.2. On note $\mathcal{Z}(X)$ le groupe abélien libre sur les classes d'isomorphisme de cycles de cobordisme sur X . En considérant la dimension des cycles de cobordisme, on le munit d'une structure de groupe abélien gradué $\mathcal{Z}_*(X)$.

Si $g : X \rightarrow X'$ est un morphisme projectif, on dispose d'un morphisme d'image directe

$$g_* : \mathcal{Z}_*(X) \longrightarrow \mathcal{Z}_*(X')$$

qui à $[f : Y \rightarrow X, (L_1, \dots, L_r)]$ associe $[g \circ f : Y \rightarrow X', (L_1, \dots, L_r)]$. On obtient ainsi un foncteur

$$\mathcal{Z}_* : \mathcal{V}'_k \longrightarrow \mathbf{Ab}_*$$

qui est additif, autrement dit, le morphisme canonique

$$\bigoplus_{1 \leq i \leq r} \mathcal{Z}_*(X_i) \longrightarrow \mathcal{Z}_*\left(\prod_{1 \leq i \leq r} X_i\right)$$

est un isomorphisme.

Si $g : X' \rightarrow X$ est un morphisme lisse équidimensionnel de dimension relative d on dispose d'un morphisme d'image inverse

$$g^* : \mathcal{Z}_*(X) \longrightarrow \mathcal{Z}_{*+d}(X')$$

qui à $[f : Y \rightarrow X, (L_1, \dots, L_r)]$ associe $[p_2 : Y \times_X X' \rightarrow X', (p_1^*L_1, \dots, p_1^*L_r)]$. Enfin, pour tout fibré en droites L sur la variété X , on dispose d'un morphisme tautologique, la première classe de Chern,

$$\tilde{c}_1(L) : \mathcal{Z}_*(X) \longrightarrow \mathcal{Z}_{*-1}(X)$$

qui à $[f : Y \rightarrow X, (L_1, \dots, L_r)]$ associe $[f : Y \rightarrow X, (L_1, \dots, L_r, f^*L)]$.

Ces données vérifient les conditions suivantes :

- (A1) On a $\text{Id}^* = \text{Id}$ et, chaque fois que cela a un sens, $(f \circ g)^* = g^* \circ f^*$.
- (A2) Si $f : X \rightarrow Z$ est projectif et $g : Y \rightarrow Z$ est lisse équidimensionnel, on a $g^* f_* = f'_* g'^*$, pour tout diagramme cartésien

$$\begin{array}{ccc} W & \xrightarrow{g'} & X \\ f' \downarrow & & \downarrow f \\ Y & \xrightarrow{g} & Z. \end{array}$$

- (A3) La première classe de Chern est compatible à l'image directe :

$$f_* \circ \tilde{c}_1(f^*L) = \tilde{c}_1(L) \circ f_*$$

et à l'image inverse :

$$\tilde{c}_1(f^*L) \circ f^* = f^* \circ \tilde{c}_1(L).$$

De plus on a toujours

$$\tilde{c}_1(L) \circ \tilde{c}_1(M) = \tilde{c}_1(M) \circ \tilde{c}_1(L)$$

et $\tilde{c}_1(L) = \tilde{c}_1(M)$ si L et M sont isomorphes.

On pose $\mathcal{Z}_*(k) := \mathcal{Z}_*(\text{Spec } k)$. On a un produit externe évident

$$\mathcal{Z}_*(X) \times \mathcal{Z}_*(Y) \longrightarrow \mathcal{Z}_*(X \times_k Y)$$

qui est associatif, commutatif, et pour lequel l'élément 1 de $\mathcal{Z}_0(k)$ est une unité. Ainsi $\mathcal{Z}_*(k)$ est muni d'une structure d'anneau gradué et $\mathcal{Z}_*(X)$ est naturellement muni d'une structure de $\mathcal{Z}_*(k)$ -module gradué.

Ce produit externe est compatible en un sens évident aux images directes et inverses, ainsi qu'aux classes de Chern \tilde{c}_1 .

2.2.3. Un foncteur additif

$$H_* : \mathcal{V}'_k \longrightarrow \mathbf{Ab}_*$$

muni de morphismes d'image inverse $g^* : H_*(X) \rightarrow H_{*+d}(X')$ pour les morphismes $g : X' \rightarrow X$ lisses équidimensionnels de dimension d et de classes de Chern $\tilde{c}_1(L) : H_*(X) \rightarrow H_{*-1}(X)$ pour les fibrés en droites, vérifiant l'analogie des conditions (A1)-(A3) est appelé foncteur de Borel-Moore orienté sur \mathcal{V}_k . On écrira $H_*(k)$ pour $H_*(\text{Spec } k)$. Si de plus, $H_0(k)$ contient un élément 1 et pour toute paire (X, Y) d'objets de \mathcal{V}_k , on dispose d'un produit externe, donné par une forme bilinéaire graduée

$$H_*(X) \times H_*(Y) \longrightarrow H_*(X \times Y),$$

qui est strictement commutatif, associatif et pour lequel 1 est une unité et que ce produit externe est compatible aux images directes et inverses, ainsi qu'aux classes de Chern \tilde{c}_1 , on dira que H_* est un foncteur de Borel-Moore orienté avec produit.

Remarquons que le foncteur \mathcal{Z}_* est universel parmi les foncteurs de Borel-Moore orientés avec produit : en effet, si H_* est un tel foncteur, il existe un unique morphisme de foncteurs de Borel-Moore orientés $\theta : \mathcal{Z}_* \rightarrow \mathcal{H}_*$ tel que $\theta(1) = 1$, car nécessairement $\theta([f : Y \rightarrow X, (L_1, \dots, L_r)]) = f_* \circ \tilde{c}_1(L_1) \circ \dots \circ \tilde{c}_1(L_r)(1)$.

2.3. De \mathcal{Z}_* à Ω_*

2.3.1. *Condition dimensionnelle.* — Un premier axiome naturel que doit satisfaire la théorie Ω_* que nous cherchons à construire est la condition de dimensionalité

(Dim) Pour toute variété lisse Y sur k et toute famille (L_1, \dots, L_n) de fibrés en droites sur Y , on a

$$\tilde{c}_1(L_1) \circ \dots \circ \tilde{c}_1(L_n)(1_Y) = 0$$

si $n > \dim(Y)$.

En quotientant les $\mathcal{Z}_*(X)$ par les sous-groupes engendrés par les éléments de la forme $[Y \rightarrow X, (\pi^*L_1, \dots, \pi^*L_n, M_1, \dots, M_m)]$ pour $\pi : Y \rightarrow Z$ un morphisme lisse équidimensionnel entre variétés lisses et $n > \dim(Z)$, on obtient un foncteur de Borel-Moore orienté $\underline{\mathcal{Z}}_*(X)$ universel parmi les foncteurs de Borel-Moore orientés avec produit satisfaisant (Dim). Remarquons que, par construction, $\underline{\mathcal{Z}}_n(X) = 0$ si $n < 0$.

2.3.2. Zéros d'une section et \tilde{c}_1 . — Une autre condition naturelle est que la première classe de Chern corresponde aux zéros d'une section générale. Plus précisément

(Sec) Pour toute variété lisse Y sur k et tout fibré en droites L sur Y , pour toute section s de L qui est transverse à la section nulle,

$$\tilde{c}_1(L)(1_Y) = [i : Z \rightarrow Y]$$

avec $i : Z \rightarrow Y$ l'inclusion du lieu des zéros de s .

En quotientant les $\underline{\mathcal{Z}}_*(X)$ par les sous-groupes engendrés par les éléments de la forme

$$[Y \rightarrow X, (L, L_1, \dots, L_n)] - [Z \rightarrow X, (i^*L_1, \dots, i^*L_n)]$$

on obtient un foncteur $\underline{\Omega}_*(X)$ universel parmi les foncteurs de Borel-Moore orientés avec produit satisfaisant (Dim) et (Sec).

2.3.3. Remarque. — Soit $W \rightarrow X \times \mathbf{A}^1$ un morphisme projectif avec W lisse et irréductible. On suppose que le morphisme composé $W \rightarrow \mathbf{A}^1$ est transverse à $\{0\}$ et à $\{1\}$ de telle sorte que les fibres W_0 et W_1 correspondantes soient lisses sur k . Par analogie avec le cas complexe, on définit le groupe de cobordisme naïf $\Omega_*^{\text{naïf}}(X)$ comme le quotient de $\mathcal{M}_*^+(X)$ par le sous-groupe engendré par les différences $[W_0 \rightarrow X] - [W_1 \rightarrow X]$ (cobordisme naïf). En appliquant (Sec) au fibré trivial sur W , on vérifie qu'on a un morphisme $\Omega_*^{\text{naïf}}(X) \rightarrow \underline{\Omega}_*(X)$. En général, ce morphisme n'est pas surjectif.

2.3.4. *R_{*}-foncteurs de Borel-Moore orientés.* — Si R_* est un anneau gradué (commutatif avec unité), un R_* -foncteur de Borel-Moore orienté est un foncteur de Borel-Moore orienté avec produit A_* muni d'un morphisme d'anneaux gradués $R_* \rightarrow A_*(k)$. Dans ce cas, tous les groupes $A_*(X)$ sont naturellement munis d'une structure de R_* -module pour laquelle toutes les opérations précédemment considérées sont R_* linéaires.

2.3.5. *Construction de Ω_* , enfin !* — On désire que Ω_* soit un \mathbf{L}_* -foncteur de Borel-Moore orienté et que la classe de Chern dans Ω_* du produit de deux fibrés en droites soit donnée par la loi de Lazard universelle $F_{\mathbf{L}}$.

Soit A_* un \mathbf{L}_* -foncteur de Borel-Moore orienté vérifiant (Dim). On note F_A l'image de la série de Lazard $F_{\mathbf{L}}$ dans $A_*(k)[[u, v]]$. On considère la condition

(LGF) Pour toute variété lisse Y sur k et toute paire de fibrés en droites (L, M) sur Y , on a

$$F_A(\tilde{c}_1(L), \tilde{c}_1(M))(1_Y) = \tilde{c}_1(L \otimes M)(1_Y).$$

Remarquons que l'évaluation de F_A a un sens grâce à l'axiome (Dim).

On peut maintenant définir Ω_* . Pour cela on considère comme précédemment le quotient des $\mathbf{L}_* \otimes_{\mathbf{Z}} \underline{\Omega}_*(X)$ par des relations convenables (cf [12]) pour obtenir le foncteur Ω_* universel parmi les \mathbf{L}_* -foncteurs de Borel-Moore orientés avec produit satisfaisant (Dim), (Sec) et (LGF).

2.3.6. Considérons le foncteur $X \mapsto \text{CH}_*(X)$ qui à une variété X sur k associe son groupe de Chow. C'est un foncteur de Borel-Moore orienté avec produit. De plus on a la relation

$$\tilde{c}_1(L \otimes M) = \tilde{c}_1(L) + \tilde{c}_1(M)$$

pour deux fibrés en droites sur une même base. On en tire une structure de \mathbf{L}_* -foncteur sur CH_* correspondant à la loi additive $F_a(u, v) = u+v$ qui munit CH_* d'une structure de \mathbf{L}_* -foncteur de Borel-Moore orienté avec produit satisfaisant (Dim), (Sec) et (LGF). On dispose donc d'un morphisme canonique de foncteurs $\Omega_* \rightarrow \text{CH}_*$.

2.3.7. Un autre exemple est fourni par le foncteur de G_0 -théorie qui à une variété X sur k associe le groupe de Grothendieck $G_0(X)$ de la catégorie des \mathcal{O}_X -modules cohérents. On pose $G_0(X)[\beta, \beta^{-1}] := G_0(X) \otimes_{\mathbf{Z}} \mathbf{Z}[\beta, \beta^{-1}]$. On munit ce groupe d'une graduation en décrétant que β est de degré 1. On définit $f_* : G_0(Y)[\beta, \beta^{-1}] \rightarrow G_0(X)[\beta, \beta^{-1}]$ pour $f : Y \rightarrow X$ un morphisme projectif par extension des scalaires à partir de $f_* : G_0(Y) \rightarrow G_0(X)$. Par contre, pour l'image inverse par un morphisme $f : Y \rightarrow X$ lisse équidimensionnel de dimension d , on multiplie de plus par β^d . Enfin,

si L est un fibré en droites sur X , on définit $\tilde{c}_1(L)$ comme la multiplication par $(1 - [\mathcal{L}^\vee])\beta$, \mathcal{L} désignant le faisceau des sections de L . On a la relation

$$\tilde{c}_1(L \otimes M) = \tilde{c}_1(L) + \tilde{c}_1(M) - \beta \tilde{c}_1(L) \circ \tilde{c}_1(M)$$

pour deux fibrés en droites sur une même base. On en tire une structure de \mathbf{L}_* -foncteur sur $G_0[\beta, \beta^{-1}]$ correspondant à la loi multiplicative $F_m(u, v) = u + v - \beta uv$. On vérifie qu'on a ainsi une structure de \mathbf{L}_* -foncteur de Borel-Moore orienté avec produit satisfaisant (Dim), (Sec) et (LGF) et donc un morphisme canonique de foncteurs $\Omega_* \rightarrow G_0[\beta, \beta^{-1}]$.

2.3.8. Classe d'un diviseur. — À tout diviseur de Cartier D dans X on peut associer la classe $[D \rightarrow X] := \tilde{c}_1(O_X(D))(1_X)$, $O_X(D)$ désignant le fibré des sections de $\mathcal{O}_X(D)$. Si D et D' sont linéairement équivalents on a $[D \rightarrow X] = [D' \rightarrow X]$.

2.3.9. Hypersurfaces de Milnor. — On considère des entiers $n > 0$ et $m > 0$. On note γ_n le fibré sur \mathbf{P}^n dont le faisceau des sections est $\mathcal{O}(1)$ et on considère le fibré $\gamma_{n,m} := p_1^* \gamma_n \otimes p_2^* \gamma_m$ sur $\mathbf{P}^n \times \mathbf{P}^m$. On considère l'hypersurface $H_{n,m}$ de $\mathbf{P}^n \times \mathbf{P}^m$ lieu des zéros d'une section de $\gamma_{n,m}$ transverse à la section nulle. Si $m \leq n$, on peut choisir $H_{n,m}$ définie par l'équation bihomogène $\sum_{0 \leq i \leq m} X_i Y_i = 0$.

L'énoncé suivant résulte aisément des constructions précédentes.

2.3.10. PROPOSITION. — *On a les relations*

$$[H_{n,m} \rightarrow \mathbf{P}^n \times \mathbf{P}^m] = \sum_{i \geq 0}^n \sum_{j \geq 0}^m a_{i,j} [\mathbf{P}^{n-i} \times \mathbf{P}^{m-j} \rightarrow \mathbf{P}^n \times \mathbf{P}^m]$$

dans $\Omega_*(\mathbf{P}^n \times \mathbf{P}^m)$, et

$$[H_{n,m}] = [\mathbf{P}^n][\mathbf{P}^{m-1}] + [\mathbf{P}^{n-1}][\mathbf{P}^m] + \sum_{i \geq 1}^n \sum_{j \geq 1}^m a_{i,j} [\mathbf{P}^{n-i}][\mathbf{P}^{m-j}]$$

dans $\Omega_*(k)$, les $a_{i,j}$ étant les coefficients de la série de Lazard $F_{\mathbf{L}}$.

Ceci donne une interprétation géométrique aux coefficients $a_{i,j}$. En particulier, pour $i, j > 1$, $a_{i,j}$ est égal à la classe de $H_{i,j}$ dans $\Omega_*(k)$ modulo des éléments décomposables. Aussi l'image de \mathbf{L}_* dans $\Omega_*(k)$ est engendrée par les images des espaces projectifs et des hypersurfaces de Milnor $H_{i,j}$. On en déduit que pour toute variété X sur k le morphisme canonique $\mathcal{Z}_*(X) \rightarrow \Omega_*(X)$ est surjectif : les $a_{i,j}$ étaient en fait déjà là avant qu'on ne les y mette de force !

En fait le morphisme canonique $\mathcal{M}_*(X) \rightarrow \Omega_*(X)$ est aussi surjectif, autrement dit $\Omega_*(X)$ est engendré comme groupe abélien par des cycles naïfs de la forme $[Y \rightarrow X]$. Pour se débarrasser des fibrés L_i dans les générateurs, on commence par

remarquer que sur une variété quasi-projective, pour tout fibré en droites L , il existe un fibré en droites M tel que M et $M \otimes L$ soient tous les deux très amples. Comme $\tilde{c}_1(L)$ s'exprime en fonction de $\tilde{c}_1(M)$ et $\tilde{c}_1(L \otimes M)$ on peut supposer que les L_i sont très amples. Quand k est infini, on a alors des sections transverses à la section nulle par Bertini et on conclut par (SEC). Quand k est fini, un petit argument supplémentaire est nécessaire ([12] Lemma 4.17).

2.4. Premières propriétés

On regroupe ici quelques propriétés fondamentales dont la démonstration n'est pas très difficile à partir de ce qui précède.

2.4.1. Formule pour un éclatement. — Soit $i : Z \rightarrow X$ une immersion fermée entre variétés lisses. On note X_Z l'éclatement de Z dans X et η le fibré conormal à i . La déformation au cône normal de i est l'éclatement $\pi : Y \rightarrow X \times \mathbf{P}^1$ de $X \times \mathbf{P}^1$ le long de $Z \times \{0\}$. Le diviseur exceptionnel P de π s'identifie au projectifié du fibré conormal à $Z \times \{0\}$ et X_Z au transformé strict de $X \times \{0\}$. On note $O_Y(P)$ le fibré en droites associé au diviseur P sur Y et $O_P(-1)$ le dual du fibré canonique quotient sur P . La série $\chi(u)$ donnant l'inverse pour la loi de groupe formel $F_{\mathbf{L}}$ (i.e. vérifiant $F_{\mathbf{L}}(u, \chi(u)) = u$) étant divisible par u , on peut l'écrire $\chi(u) = ug(u)$.

2.4.2. PROPOSITION. — *Avec les notations précédentes, on a les égalités*

$$[X_Z \rightarrow Y] = [X \rightarrow Y] + \chi([O_Y(P)])$$

dans $\Omega_*(Y)$ et

$$[X_Z \rightarrow X] = [X \rightarrow X] + i_*q_*([g(O_P(-1))])$$

dans $\Omega_*(X)$, q désignant la projection $P \rightarrow X$.

2.4.3. Trace et dimension zéro. — Le lemme suivant se démontre en considérant un cobordisme naïf (cf. remarque 2.3.3) convenable.

2.4.4. LEMME. — *Soit L une k -algèbre séparable finie. Alors $[\mathrm{Spec} L] = ([L : k])1$ dans $\Omega_0(k)$.*

On en déduit aisément :

2.4.5. PROPOSITION. — *Soit $L|k$ une extension séparable finie de corps. Alors la composition des morphismes canoniques*

$$\Omega_*(X) \longrightarrow \Omega_*(X \otimes L) \longrightarrow \Omega_*(X)$$

est la multiplication par $[L : k]$.

D'où finalement :

2.4.6. THÉORÈME. — *Pour tout corps k le morphisme canonique $\mathbf{L}_0 \rightarrow \Omega_0(k)$ est un isomorphisme. Pour toute variété lisse de dimension 0, $X = \text{Spec } A$ sur k , $[X] = (\dim_k A) 1$ dans $\Omega_0(k)$.*

3. PROPRIÉTÉS FONDAMENTALES

On aura besoin dans cette section de supposer que k admet la résolution des singularités, ce qui est vérifié pour k de caractéristique 0. Par souci de simplicité, on supposera dans la suite de cet exposé que k est toujours de caractéristique 0, même si ce n'est pas explicite et/ou utile.

3.1. Le théorème de localisation

L'énoncé suivant est vraiment le résultat technique essentiel qui permet la démonstration des propriétés fondamentales de Ω_* .

3.1.1. THÉORÈME. — *On suppose que le corps k admet la résolution des singularités. Soient X une variété sur k , $i : Z \rightarrow X$ une sous-variété fermée et $j : U \rightarrow X$ l'ouvert complémentaire. Alors on a une suite exacte*

$$\Omega_*(Z) \longrightarrow \Omega_*(X) \longrightarrow \Omega_*(U) \longrightarrow 0.$$

Expliquons pourquoi l'hypothèse que le corps k admette la résolution des singularités permet de démontrer la surjectivité de $j^* : \Omega_*(X) \rightarrow \Omega_*(U)$. On part de $Y \rightarrow U$ un morphisme projectif avec Y lisse quasi-projective. En appliquant la résolution des singularités et le lemme de Chow à une adhérence convenable de Y on parvient à construire un morphisme projectif $\tilde{Y} \rightarrow X$ prolongeant $Y \rightarrow U$ avec \tilde{Y} lisse quasi-projective, ce qui permet de démontrer la surjectivité au niveau de \mathcal{M}_*^+ . La suite de la démonstration est bien plus délicate. Elle procède en plusieurs étapes successives, pour lesquelles nous renvoyons à [12], et consiste à prolonger, de façon plus ou moins analogue, certaines relations de U à X . Elle utilise en particulier de façon fondamentale l'énoncé suivant, qui est une conséquence (d'une variante de) la proposition 2.4.2 et de la résolution des singularités.

3.1.2. PROPOSITION. — *Soit $W' \rightarrow W$ un morphisme birationnel projectif entre variétés quasi-projectives lisses. On note F et E les lieux exceptionnels de W et W' ,*

respectivement. On suppose que E est un diviseur à croisements normaux. Alors il existe η dans $\Omega_*(F)$ tel que

$$[W' \rightarrow W] = [W \rightarrow W] + i_*(\eta),$$

i désignant le morphisme d'inclusion de F dans W .

3.2. Invariance par homotopie

3.2.1. THÉORÈME. — Soit X une variété sur k . Pour tout entier n le morphisme d'image inverse $p^* : \Omega_*(X) \rightarrow \Omega_{*+n}(X \times \mathbf{A}^n)$ associé à la projection $p : X \times \mathbf{A}^n \rightarrow X$ est un isomorphisme.

On se ramène au cas $n = 1$ et on commence par démontrer :

3.2.2. LEMME. — Avec les notations précédentes, pour tout a dans k , le morphisme $T_a : X \times \mathbf{A}^1 \rightarrow X \times \mathbf{A}^1$ donné par la translation par a sur le deuxième facteur induit l'identité sur $\Omega_*(X \times \mathbf{A}^1)$

Démonstration. — Soit $f : Y \rightarrow X \times \mathbf{A}^1$. On considère $F : X \times \mathbf{A}^1 \times \mathbf{A}^1 \rightarrow X \times \mathbf{A}^1$ donné par $(x, t_1, t_2) \mapsto (x, t_1 + at_2)$. Le produit fibré F^*Y donne un cobordisme naïf (au sens de la remarque 2.3.3) entre f qui est isomorphe à la fibre en 0 et son translaté par a qui est isomorphe à la fibre en 1. \square

Suite de la preuve. — Démontrons la surjectivité de p^* . Soit $f : Y \rightarrow X \times \mathbf{A}^1$. D'après le lemme précédent on peut supposer que le composé $Y \rightarrow \mathbf{A}^1$ est lisse en 0. On considère le produit fibré $Y_m := Y \times_{\mathbf{A}^1} (\mathbf{A}^1 \times \mathbf{A}^1)$, le morphisme $\mathbf{A}^1 \times \mathbf{A}^1 \rightarrow \mathbf{A}^1$ étant celui de multiplication, et le cycle $g : Y_m \rightarrow X \times \mathbf{A}^1 \times \mathbf{A}^1$ qui s'en déduit en composant avec la projection. On vérifie que ce cycle donne un cobordisme naïf entre f qui est isomorphe à la préimage de $X \times \mathbf{A}^1 \times \{1\}$ et la préimage de $X \times \mathbf{A}^1 \times \{0\}$ qui s'identifie à $p^*(f^{-1}(X \times \{0\}) \times X)$.

On va démontrer l'injectivité de p^* en exhibant un inverse à gauche. Soit $f : Y \rightarrow X \times \mathbf{P}^1$ un cycle sur $X \times \mathbf{P}^1$ de classe η dans $\Omega_*(X \times \mathbf{P}^1)$. La classe $\psi(\eta)$ du cycle $f^{-1}(X \times \{a\}) \rightarrow X$ dans $\Omega_{*-1}(X)$ est bien définie et indépendante de a , pour a assez général. En effet $\psi = q_* \circ \tilde{c}_1(\gamma_1)$ avec $q : X \times \mathbf{P}^1 \rightarrow X$ la projection. En particulier ψ s'annule sur l'image de $\Omega_*(X \times \{\infty\})$ dans $\Omega_*(X \times \mathbf{P}^1)$. Par le théorème de localisation 3.1.1, il s'ensuit que ψ se factorise par un morphisme $\Omega_*(X \times \mathbf{A}^1) \rightarrow \Omega_{*-1}(X)$ qui est l'inverse à gauche de p^* désiré. \square

3.3. Fibrés projectifs

Soit X une k -variété et soit E un fibré vectoriel de rang n sur X . On considère le fibré projectif $q : \mathbf{P}(E) \rightarrow X$ représentant les quotients de rang 1 de E . On note $O(1) \rightarrow \mathbf{P}(E)$ le fibré en droites quotient canonique de $q^*(E)$. Pour tout entier i dans $\{0, \dots, n\}$ on considère le morphisme

$$\xi^{(i)} : \Omega_{*-n+i}(X) \longrightarrow \Omega_*(\mathbf{P}(E))$$

défini en posant $\xi^{(i)} := \tilde{c}_1(O(1))^i \circ q^*$.

Par des méthodes essentiellement classiques, Levine et Morel obtiennent à partir des théorèmes 3.1.1 et 3.2.1, la formule suivante pour les fibrés projectifs.

3.3.1. THÉORÈME. — *Soit X une k -variété et soit E un fibré vectoriel de rang n sur X . Le morphisme canonique*

$$\bigoplus_{i=0}^{n-1} \xi^{(i)} : \bigoplus_{i=0}^{n-1} \Omega_{*-n+i}(X) \longrightarrow \Omega_*(\mathbf{P}(E))$$

est un isomorphisme.

En utilisant ce résultat joint au théorème 3.1.1, ils obtiennent la forme plus générale suivante de l'invariance par homotopie.

3.3.2. THÉORÈME. — *Soit X une k -variété, soit E un fibré vectoriel de rang n sur X et soit $p : V \rightarrow X$ un E -torseur. Alors le morphisme canonique*

$$p^* : \Omega_*(X) \longrightarrow \Omega_{*+n}(V)$$

est un isomorphisme.

3.4. Classes de Chern

Soit X une k -variété et soit E un fibré vectoriel de rang n sur X . On déduit du théorème 3.3.1 que $\Omega_*(\mathbf{P}(E))$ est muni d'une structure canonique de $\text{End } \Omega_*(X)$ -module gradué à gauche. On en tire l'existence de morphismes

$$\tilde{c}_i(E) : \Omega_*(X) \longrightarrow \Omega_*(X)$$

pour $0 \leq i \leq n$, caractérisés par les relations

$$\sum_{i=0}^n \tilde{c}_i(E) \tilde{c}_1(O(1))^{n-i} = 0$$

et $\tilde{c}_0(E) = \text{id}$, les classes de Chern de E , qui vérifient les propriétés «usuelles» des classes de Chern.

Il est aussi possible de définir ces classes de Chern à partir des classes de Chern des fibrés en droites en utilisant le principe de scindage. Plus généralement considérons l'anneau gradué $\mathbf{Z}[t] = \mathbf{Z}[t_1, \dots, t_n, \dots]$ des polynômes à coefficients entiers en une infinité de variables t_i de degré 1 et posons

$$\Omega_*(X)[t] := \Omega_*(X) \otimes_{\mathbf{Z}} \mathbf{Z}[t].$$

Pour tout fibré en droites L sur X on considère l'automorphisme $\tilde{c}_t(L) := \sum_{i=0}^{\infty} \tilde{c}_1(L)^i t_i$ de $\Omega_*(X)[t]$. Par le principe de scindage, cette construction s'étend à tout fibré vectoriel E de rang n sur X . On développe alors $\tilde{c}_t(E)$ en

$$\tilde{c}_t(E) = \sum_I \tilde{c}_I(E) t^I,$$

I parcourant $\mathbf{N}^{\mathbf{N}}$. Les coefficients $\tilde{c}_I(E)$ sont les classes de Chern de Conner et Floyd attachées à E . On retrouve les classes de Chern usuelles comme coefficients des t_i .

4. THÉORÈMES DE COMPARAISON

4.1. Loi de groupe formel multiplicative et invariance birationnelle

On s'intéresse dans cette section à des foncteurs de Borel-Moore orientés avec produit satisfaisant (Dim), (Sec) et (LGF) avec loi de groupe formel multiplicative, c'est-à-dire de la forme

$$u + v - \beta uv,$$

avec β non nécessairement inversible. Le foncteur $\Omega_*^\beta := \Omega_* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta]$ est universel parmi de tels foncteurs. On notera avec un indice α_β l'image d'un élément α de $\Omega_*(X)$ dans $\Omega_*^\beta(X)$. Il résulte de la proposition 2.3.10 que $[\mathbf{P}^1]_\beta = \beta$. De plus on peut montrer, à l'aide de la proposition 2.4.2, du théorème 3.3.2 et de l'astuce de Jouanolou, que, pour tout fibré vectoriel E sur X lisse quasi-projective de rang n , on a

$$(4.1.1) \quad [\mathbf{P}(E) \rightarrow X]_\beta = \beta^{n-1} [X \rightarrow X]_\beta.$$

En particulier on a $[\mathbf{P}^n]_\beta = \beta^n$. La formule pour un éclatement 2.4.2 se simplifie dramatiquement alors pour donner :

4.1.1. PROPOSITION. — *Soit X une variété quasi-projective lisse quasi-projective sur k et soit Z une sous-variété lisse fermée. Alors on a*

$$[X_Z \rightarrow X]_\beta = [X \rightarrow X]_\beta$$

dans $\Omega_*^\beta(X)$.

En utilisant le théorème de factorisation faible de Abramovich, Karu, Matsuki, et Włodarczyk ([1], [3]), on en déduit :

4.1.2. COROLLAIRE. — Soient Z et Z' deux variétés projectives lisses birationnellement équivalentes. Alors $[Z]_\beta = [Z']_\beta$ dans $\Omega_*^\beta(k)$.

Plus précisément, on a le résultat suivant :

4.1.3. PROPOSITION. — Le noyau du morphisme canonique

$$\Omega_*(k) \longrightarrow \Omega_*^\beta(k)$$

est l'idéal I engendré par les différences $[Z] - [Z']$ avec Z et Z' des variétés projectives lisses birationnellement équivalentes.

Démonstration. — Au vu du corollaire 4.1.2, il suffit de démontrer que $\Omega_*(k) = I + \mathbf{Z}[\mathbf{P}^1]$, ce qui résulte du fait que les hypersurfaces de Milnor $H_{n,m}$ sont birationnellement équivalentes à \mathbf{P}^{n+m-1} et de ce que $\Omega_*(k)$ est engendré par les classes des hypersurfaces de Milnor et des espaces projectifs (ce dernier point est conséquence de l'isomorphisme $\mathbf{L}_* \rightarrow \Omega_*(k)$ démontré en 4.3.2). \square

On en déduit le théorème :

4.1.4. THÉORÈME. — Le foncteur $X \mapsto \Omega_*(X) \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta]$ est universel parmi les \mathbf{L}_* -foncteurs de Borel-Moore orientés avec produit satisfaisant (Dim), (Sec) et (LGF) satisfaisant la condition d'invariance birationnelle suivante : pour tout morphisme projectif birationnel $f : Y \rightarrow X$ entre variétés lisses irréductibles, on a $f_* 1_Y = 1_X$.

Démonstration. — Il résulte de la proposition 4.1.3 que si A_* est un \mathbf{L}_* -foncteur de Borel-Moore orienté avec produit satisfaisant (Dim), (Sec) et (LGF) satisfaisant la condition d'invariance birationnelle, alors on a un morphisme de foncteurs $\Omega_* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta] \rightarrow A_*$. Il reste à vérifier que le foncteur $\Omega_* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta]$ satisfait la condition d'invariance birationnelle, ce qui résulte de la proposition 4.1.1 et du théorème de factorisation faible. \square

4.2. Comparaison avec le K^0

Dans ce numéro on restreindra des foncteurs de Borel-Moore orientés avec produit A_* à la catégorie \mathcal{L}'_k . On passera en numérotation cohomologique en posant $A^*(X) = A_{\dim X - *}(X)$.

Pour X dans \mathcal{L}'_k le morphisme canonique entre groupes de Grothendieck $K^0(X) \rightarrow G_0(X)$ est un isomorphisme (on rappelle que $K^0(X)$ désigne le groupe de Grothendieck des faisceaux localement libres de type fini sur X). On identifiera donc la restriction du foncteur $G_0[\beta, \beta^{-1}]$ à \mathcal{L}'_k au foncteur $K^0[\beta, \beta^{-1}] := K^0 \otimes_{\mathbf{Z}} \mathbf{Z}[\beta, \beta^{-1}]$, en convenant maintenant que β est de degré -1 .

4.2.1. THÉORÈME. — *Soit k un corps de caractéristique zéro. Le morphisme canonique de foncteurs de Borel-Moore orientés avec produit sur \mathcal{L}'_k*

$$\Omega^* \longrightarrow K^0[\beta, \beta^{-1}]$$

induit un isomorphisme

$$\Omega^* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta, \beta^{-1}] \simeq K^0[\beta, \beta^{-1}].$$

Ébauche de preuve. — Compte tenu du caractère universel de Ω_* , il suffit de montrer qu'il existe un unique morphisme de foncteurs de Borel-Moore orientés avec produit sur \mathcal{L}'_k :

$$\lambda : K^0[\beta, \beta^{-1}] \longrightarrow \Omega^* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta, \beta^{-1}].$$

Remarquons que nécessairement λ doit envoyer β sur la classe de $[\mathbf{P}^1]$, autrement dit β . De plus, pour X dans \mathcal{L}_k , on a, pour tout fibré en droites L de faisceau des sections \mathcal{L} , la relation $[\mathcal{L}] = 1 - c_1(L^\vee)\beta$ dans $K^0[\beta, \beta^{-1}]$, et donc nécessairement λ doit envoyer $[\mathcal{L}]$ sur $1 - c_1(L^\vee)\beta$ dans $\Omega^* \otimes_{\mathbf{L}_*} \mathbf{Z}[\beta, \beta^{-1}]$, ce qui garantit l'unicité de λ par le principe de scindage. Pour définir λ , on pose $\lambda([\mathcal{E}]\beta^n) = (r - c_1(E^\vee)\beta)\beta^n$, pour E un fibré vectoriel de rang r sur X de faisceau des sections \mathcal{E} . Par le principe de scindage, λ est un morphisme d'anneaux gradués. Le fait qu'il commute à l'image inverse par les morphismes lisses étant clair, il reste à vérifier qu'il commute à l'image directe par les morphismes projectifs. Ceci se démontre de façon tout à fait analogue⁽¹⁾ à la preuve du théorème de Riemann-Roch-Grothendieck donnée dans [5]. Il suffit de vérifier l'énoncé pour une projection $\mathbf{P}_k^n \times X \rightarrow X$ sur un schéma lisse et pour une immersion fermée $Y \rightarrow X$ entre schémas lisses. Pour le premier cas on se ramène au cas où $X = \text{Spec } k$, par compatibilité avec le produit externe, qui est de vérification directe. Le cas de l'immersion fermée se démontre lui par déformation au cône normal. \square

⁽¹⁾En fait, d'après [12] remarque 8 page 93, le théorème est essentiellement équivalent au théorème de Riemann-Roch-Grothendieck.

4.3. Calcul de $\Omega_*(k)$

Commençons par déterminer $\Omega_*(k) \otimes_{\mathbf{L}_*} \mathbf{Z}$.

4.3.1. PROPOSITION. — *Soit k un corps de caractéristique zéro. On a*

$$\Omega_*(k) \otimes_{\mathbf{L}_*} \mathbf{Z} \simeq \mathbf{Z}.$$

Démonstration. — Posons $\Omega_*^{\text{ad}}(k) = \Omega_*(k) \otimes_{\mathbf{L}_*} \mathbf{Z}$. D'après le théorème 2.4.6, $\Omega_0^{\text{ad}}(k) \simeq \mathbf{Z}$. Il suffit donc de démontrer que $\Omega_n^{\text{ad}}(k) = 0$, pour $n > 0$, autrement dit que la classe de X dans $\Omega_*^{\text{ad}}(k)$ est nulle, pour toute variété projective lisse X de dimension > 0 . D'après 4.1, la classe de \mathbf{P}^n est nulle pour $n > 0$. Si $n > 0$, la classe de toute hypersurface lisse Y dans \mathbf{P}^{n+1} est nulle; en effet une telle hypersurface est linéairement équivalente au diviseur $d\mathbf{P}^n$ et $[d\mathbf{P}^n] = d[\mathbf{P}^n] = 0$ dans $\Omega_*^{\text{ad}}(k)$. On utilise ici l'additivité de la loi de groupe formel associée à $\Omega_*^{\text{ad}}(k)$ ainsi que les notations de 2.3.8. Considérons maintenant une variété projective lisse irréductible Y . Elle est birationnellement équivalente à une hypersurface réduite (peut-être singulière) \bar{Y} . Par le théorème d'Hironaka, il existe un morphisme $\mu : S \rightarrow \mathbf{P}^{n+1}$ composé d'éclatements de centres lisses tel que le diviseur $\mu^{-1}(\bar{Y})$ soit à croisements normaux. En particulier la transformée stricte \tilde{Y} de \bar{Y} est lisse et birationnellement équivalente à Y . On a donc $[Y] = [\tilde{Y}]$ dans $\Omega_*^{\text{ad}}(k)$. Il résulte de (4.1.1) appliqué aux centres des éclatements successifs que $[\tilde{Y}] = [\mu^{-1}(\bar{Y})]$ dans $\Omega_*^{\text{ad}}(k)$. Pour conclure, on remarque que la classe de $\mu^{-1}(\bar{Y})$ est égale à celle de $\mu^{-1}(D)$ pour tout diviseur D de même degré que \bar{Y} , et que $\mu^{-1}(D)$ est lisse et birationnellement équivalent à D pour D assez général. \square

4.3.2. THÉORÈME. — *Soit k un corps de caractéristique zéro. Le morphisme canonique*

$$\Psi : \mathbf{L}_* \longrightarrow \Omega_*(k)$$

est un isomorphisme.

Ébauche de preuve. — Démontrons l'injectivité. Lorsque k est un sous-corps de \mathbf{C} , on peut considérer le foncteur cohomologique $X \mapsto U^{2*}(X(\mathbf{C}))$ sur les variétés lisses sur k . Par le caractère universel du cobordisme algébrique, on en tire un morphisme canonique $\Omega^*(X) \rightarrow U^{2*}(X(\mathbf{C}))$, et en particulier un morphisme canonique d'anneaux $\theta : \Omega_*(k) \rightarrow U_{2*}(\text{point})$. Mais le composé $\theta \circ \Psi$ n'est autre que l'isomorphisme de Quillen $\delta : \mathbf{L}^* \rightarrow U^*(\text{point})$, ce qui prouve l'injectivité de Ψ . Pour démontrer l'injectivité sans hypothèse sur k , Levine et Morel considèrent le

morphisme canonique⁽²⁾

$$\Omega_*(k) \longrightarrow \mathbf{Z}[t_1, \dots, t_n, \dots]$$

qui associe à une variété projective lisse sur k « l'ensemble de ses nombres de Chern ». En composant avec le morphisme canonique $\mathbf{L}_* \rightarrow \Omega_*(k)$, on en tire un morphisme

$$\mathbf{L}_* \longrightarrow \mathbf{Z}[t_1, \dots, t_n, \dots].$$

Pour conclure, Levine et Morel démontrent que ce morphisme correspond à la loi de groupe formel $\lambda^{-1}(\lambda(u) + \lambda(v))$, avec $\lambda(u)$ la série $\sum_{i \geq 0} t_i u^i$ et $\lambda^{-1}(u)$ son inverse pour la composition, dont il est connu qu'il est injectif [20]. Cette preuve n'utilise pas le théorème de Quillen.

Enfin, il résulte de la proposition 4.3.1 que, pour $i > 0$, $\Omega_i(k) = \sum_{1 \leq j \leq i} \mathbf{L}_j \Omega_{i-j}(k)$, et la surjectivité suit. \square

4.3.3. Remarque. — La preuve du théorème précédent est toute différente (et indépendante) de celle du théorème de Quillen. En effet, comme nous l'avons déjà signalé, celle-ci repose de façon essentielle sur les propriétés de finitude du cobordisme complexe, dont l'analogue strict en cobordisme algébrique n'est vraisemblablement pas vérifié en général. Par contre on a utilisé des arguments géométriques qui n'ont pas d'analogue en topologie : la suite exacte de localisation, le théorème d'Hironaka, le fait que toute variété soit birationnellement une hypersurface, etc.

4.3.4. COROLLAIRE. — *Soit $k \subset K$ une extension de corps de caractéristique zéro. Le morphisme canonique $\Omega_*(k) \rightarrow \Omega_*(K)$ est un isomorphisme.*

4.4. Comparaison avec les groupes de Chow

4.4.1. THÉORÈME. — *Soit k un corps de caractéristique zéro. Le morphisme canonique de foncteurs de Borel-Moore orientés*

$$\Omega_* \longrightarrow \mathrm{CH}_*$$

induit un isomorphisme

$$\Omega_* \otimes_{\mathbf{L}_*} \mathbf{Z} \longrightarrow \mathrm{CH}_*.$$

Démonstration. — Commençons par remarquer que si $\tilde{Z} \rightarrow Z$ est un morphisme projectif birationnel entre variétés lisses, alors l'image de $[\tilde{Z} \rightarrow Z]$ dans $\Omega_*(Z) \otimes_{\mathbf{L}_*} \mathbf{Z}$ est égale à 1_Z . En effet cela résulte du théorème 5.2.1 car les termes du second membre de (5.2.1) sont nuls dans $\Omega_*(Z) \otimes_{\mathbf{L}_*} \mathbf{Z}$ pour des raisons de degré.

⁽²⁾Merkurjev a démontré, sans hypothèse sur k , que ce morphisme se factorise en une rétraction $\Omega_*(k) \rightarrow \mathbf{L}_*$ du morphisme canonique $\mathbf{L}_* \rightarrow \Omega_*(k)$.

Soit Z un sous-schéma fermé intègre d'un schéma X de type fini sur k . Il résulte de la remarque précédente que la classe de $[\tilde{Z} \rightarrow X]$ dans $\Omega_*(X) \otimes_{\mathbf{L}^*} \mathbf{Z}$, pour $\tilde{Z} \rightarrow Z$ un morphisme projectif birationnel avec \tilde{Z} lisse, ne dépend que de Z . On la note $[Z \hookrightarrow X]$. On définit ainsi un morphisme de groupes abéliens $\Phi : Z_*(X) \rightarrow \Omega_*(X) \otimes_{\mathbf{L}^*} \mathbf{Z}$ dont le composé avec le morphisme $\Omega_*(X) \otimes_{\mathbf{L}^*} \mathbf{Z} \rightarrow \mathrm{CH}_*(X)$ est le morphisme canonique $Z_*(X) \rightarrow \mathrm{CH}_*(X)$. Comme il résulte du théorème 5.2.1 que le morphisme Φ est surjectif, il reste à savoir que Φ se factorise par $\mathrm{CH}_*(X)$, ce qui est donné par le lemme suivant. \square

4.4.2. LEMME. — *Soit W un sous-schéma fermé intègre d'un schéma X de type fini sur k et soit f dans $k(W)^*$ une fonction rationnelle non nulle de diviseur $\mathrm{div}(f)$ dans $Z_*(X)$. Alors $\Phi(\mathrm{div}(f)) = 0$ dans $\Omega_*(X) \otimes_{\mathbf{L}^*} \mathbf{Z}$.*

Ébauche de preuve. — Grâce à la résolution des singularités on peut considérer un morphisme $\pi : \tilde{W} \rightarrow W$ avec \tilde{W} quasi-projective lisse tel que f induise un morphisme $\tilde{f} : \tilde{W} \rightarrow \mathbf{P}^1$. On peut de plus supposer que le diviseur de \tilde{f} soit de la forme $D_0 - D_\infty$ avec D_0 et D_∞ des diviseurs à croisements normaux ne s'intersectant pas. Considérons les classes $[D_0 \rightarrow \tilde{W}]$ et $[D_\infty \rightarrow \tilde{W}]$ définies dans 2.3.8. Comme on a $[D_0 \rightarrow \tilde{W}] = [D_\infty \rightarrow \tilde{W}]$, il suffit de démontrer que $\Phi(D_0)$ est égale à l'image de $[D_0 \rightarrow \tilde{W}]$ dans $\Omega_*(X) \otimes_{\mathbf{L}^*} \mathbf{Z}$, et de même pour D_∞ , ce qui résulte des formules explicites pour la classe d'un diviseur à croisements normaux de [12] et de la proposition 4.3.1. \square

En particulier, pour X une variété algébrique complexe lisse et projective, on déduit du théorème 4.4.1 l'existence d'un morphisme canonique

$$\mathrm{CH}^*(X) \longrightarrow U^{2*}(X(\mathbf{C})) \otimes_{\mathbf{L}^*} \mathbf{Z}.$$

Ce morphisme avait été introduit par B. Totaro dans son travail sur les cycles algébriques de torsion [22].

5. LA FORMULE DU DEGRÉ

On rappelle que l'on fait partout l'hypothèse que k est de caractéristique zéro. On renvoie à [12], [14], [15], [16] pour une discussion de ce qui reste connu en caractéristique positive.

5.1. Définition du degré

Soit X une variété irréductible sur k de corps des fractions $k(X)$. Pour tout ouvert U non vide de X on dispose d'un morphisme canonique $\Omega_*(U) \rightarrow \Omega_*(k(X))$ et on vérifie que le morphisme canonique

$$\varinjlim \Omega_*(U) \rightarrow \Omega_*(k(X))$$

est un isomorphisme. En particulier on dispose d'un morphisme $\Omega_*(X) \rightarrow \Omega_*(k(X))$, qui composé avec l'inverse de l'isomorphisme du corollaire 4.3.4 fournit un morphisme canonique $\deg : \Omega_*(X) \rightarrow \Omega_*(k)$. Plus généralement, si X_1, \dots, X_n sont les composantes irréductibles de X , on définit de façon similaire des morphismes degré $\deg_i : \Omega_*(X) \rightarrow \Omega_*(k)$, pour $1 \leq i \leq n$.

5.2. Formule générale du degré

Nous sommes maintenant en mesure d'énoncer la formule générale du degré de Levine et Morel.

5.2.1. THÉORÈME (Formule générale du degré). — *Soit X un schéma de type fini sur k de composantes irréductibles X_1, \dots, X_n . Pour tout sous-schéma fermé Z de X , on fixe un morphisme projectif birationnel $\tilde{Z} \rightarrow Z$ avec \tilde{Z} dans \mathcal{L}_k . Alors, pour tout élément α de $\Omega_*(X)$, on peut écrire*

$$(5.2.1) \quad \alpha - \sum_{1 \leq i \leq n} \deg_i(\alpha)[\tilde{X}_i \rightarrow X_i] = \sum_{Z, \text{codim} Z > 0} \omega_Z[\tilde{Z} \rightarrow X]$$

pour un choix convenable d'éléments ω_Z dans $\Omega_*(k)$.

Ébauche de preuve. — On raisonne par récurrence sur la dimension de X , en utilisant le théorème de localisation 3.1.1. \square

5.2.2. COROLLAIRE. — *Pour tout schéma de type fini X , $\Omega_*(X)$ est engendré comme $\Omega_*(k)$ -module par les classes de degré $\leq \dim X$. Si, de plus, X est irréductible, fixons un morphisme projectif birationnel $\tilde{X} \rightarrow X$ avec \tilde{X} dans \mathcal{L}_k . Alors $\Omega_*(X)$ est engendré comme $\Omega_*(k)$ -module par les classes de degré $\leq \dim X - 1$ et $[\tilde{X} \rightarrow X]$.*

5.2.3. Remarque. — Le corollaire précédent est l'analogue algébrique d'un résultat de Quillen [20] selon lequel l'anneau de cobordisme complexe $U_*(X)$ est engendré comme \mathbf{L}_* -module par les classes de degré $\leq 2 \dim X$, si X est un complexe fini.

Soit X une variété projective irréductible et lisse sur k . Considérons le noyau $\tilde{\Omega}_*(X)$ du morphisme $\deg : \Omega_*(X) \rightarrow \Omega_*(k)$. Remarquons que le morphisme

composé $\Omega_*(k) \rightarrow \Omega_*(X) \rightarrow \Omega_*(k)$ étant l'identité, on a un scindage canonique $\Omega_*(X) = \Omega_*(k) \oplus \tilde{\Omega}_*(X)$.

Il résulte du théorème 5.2.1 que l'image de $\tilde{\Omega}_*(X)$ par le morphisme d'image directe $\Omega_*(X) \rightarrow \Omega_*(k)$ associé au morphisme structural $X \rightarrow \text{Spec } k$ coïncide avec l'idéal $M(X)$ de $\Omega_*(k)$ engendré par les classes $[Y]$, pour Y projective lisse de dimension $< \dim X$ admettant un k -morphisme $Y \rightarrow X$.

L'énoncé suivant qui est également une conséquence du théorème 5.2.1, a été conjecturé par Rost (cf. [14]).

5.2.4. THÉORÈME. — *Soit $f : Y \rightarrow X$ un morphisme entre variétés projectives lisses irréductibles. Alors $[Y] - \deg(f)[X]$ appartient à l'idéal $M(X)$ de $\Omega_*(k)$.*

5.3. La formule du degré de Rost

Soit $f : Y \rightarrow X$ un morphisme entre variétés projectives lisses irréductibles de même dimension d . Remarquons que dans ce cas $\deg(f)$ appartient à $\Omega_0(k)$, on peut donc le voir comme un entier, égal au degré usuel d'après le lemme 2.4.4. On peut ainsi réécrire (5.2.1) en

$$(5.3.1) \quad [Y] - \deg(f)[X] = \sum_{Z, \text{codim } Z > 0} \omega_Z [\tilde{Z} \rightarrow X]$$

pour un choix convenable d'éléments ω_Z dans $\Omega_{d-\dim Z}(k)$.

Pour toute variété projective lisse X de dimension d , on considère N_d , le d -ième polynôme de Newton en les classes de Chern du fibré tangent dans $\text{CH}_0(X)$ et on note $s_d(X)$ l'entier $-\deg N_d$ (il s'agit ici du degré usuel $\text{CH}_0(X) \rightarrow \mathbf{Z}$). Les faits suivants sont bien connus des topologues (cf. [2]).

(S1) Si d est de la forme $d = p^n - 1$ avec p un nombre premier, alors l'entier $s_d(X)$ est divisible par p .

(S2) Pour X et X' de dimension d et d' tous deux > 0 , on a $s_{d+d'}(X \times X') = 0$.

On va maintenant déduire de la formule générale du degré l'énoncé suivant, dû à Rost (cf. [14], [15]), qui généralise des résultats antérieurs de Voevodsky.

5.3.1. THÉORÈME (Formule du degré de Rost). — *Soit $f : Y \rightarrow X$ un morphisme entre variétés projectives lisses irréductibles de même dimension $d > 0$. On suppose que d est de la forme $d = p^n - 1$ avec p un nombre premier. Alors il existe un zéro-cycle z dans $\text{CH}_0(X)$ de degré égal à*

$$\frac{s_d(Y)}{p} - \deg(f) \frac{s_d(X)}{p}.$$

Démonstration. — Si on applique s_d à l'égalité (5.3.1), on obtient

$$s_d(Y) - \deg(f)s_d(X) = \sum_{Z, \dim Z=0} s_d(\omega_Z[Z \rightarrow X])$$

pour un choix convenable d'éléments ω_Z dans $\Omega_d(k)$, car les Z de dimension > 0 disparaissent grâce à (S2). Le zéro-cycle $z = \sum \frac{s_d(\omega_Z)}{p} Z$ vérifie alors la propriété requise. \square

5.3.2. Remarque. — Il est important de remarquer qu'il est vraiment nécessaire de travailler avec Ω_* pour pouvoir utiliser (S2) dans la preuve précédente. En particulier celle-ci ne fonctionnerait plus si on se contentait de travailler directement dans l'anneau de Chow.

5.3.3. Remarque. — La version plus générale du théorème 5.3.1 due à Borghesi [4] peut également être déduite du théorème 5.2.1.

La formule du degré de Rost admet un certain nombre de conséquences frappantes (cf. [15]). Elle permet en particulier de retrouver les résultats suivants de la théorie des formes quadratiques (cf. [15]).

5.3.4. THÉORÈME (Hoffmann [6]). — *Soient Q_1 et Q_2 deux quadriques anisotropes sur k . Si Q_1 est de dimension $\geq 2^r - 1$ et si Q_2 est anisotrope sur le corps des fonctions de Q_1 , alors Q_2 est de dimension $\geq 2^r - 1$.*

5.3.5. THÉORÈME (Izhboldin [7]). — *Soient Q_1 et Q_2 deux quadriques anisotropes sur k . Si Q_1 est de dimension $2^r - 1$ et si Q_2 est isotrope sur le corps des fonctions de Q_1 , alors Q_1 est isotrope sur le corps des fonctions de Q_2 .*

6. IMAGES INVERSES EN COBORDISME ALGÈBRE

On présente dans cette section des résultats contenus dans [13].

6.1. Théories homologiques de Borel-Moore orientées

Pour les besoins de la construction de Ω_* , nous avons été minimalistes pour ce qui est des propriétés demandées a priori. En particulier nous n'avons considéré que des images inverses par des morphismes lisses. Il est souhaitable d'avoir des images inverses en plus grande généralité, comme celle fournie par le formalisme des théories homologiques de Borel-Moore orientées.

Une théorie de Borel-Moore orientée sur \mathcal{V}_k consiste en les données suivantes [13] :

(D1) Un foncteur additif

$$H_* : \mathcal{V}'_k \longrightarrow \mathbf{Ab}_*.$$

(D2) Pour tout morphisme $f : Y \rightarrow X$ dans \mathcal{V}_k localement d'intersection complète⁽³⁾ de dimension relative d , un morphisme

$$f^* : H_*(X) \longrightarrow H_{*+d}(Y).$$

(D3) Un produit externe bilinéaire

$$H_*(X) \times H_*(Y) \longrightarrow H_*(X \times_k Y)$$

qui est associatif, commutatif, et admettant une unité 1 de $H_0(k)$.

On demande de plus que ces données vérifient les propriétés suivantes :

(BM1) Functorialité de f^* .

(BM2) Si $f : X \rightarrow Z$ est un morphisme projectif et $g : Y \rightarrow Z$ est un morphisme localement d'intersection complète et que g est transverse à f , on a $g^*f_* = f'_*g'^*$, pour tout diagramme cartésien

$$\begin{array}{ccc} W & \xrightarrow{g'} & X \\ f' \downarrow & & \downarrow f \\ Y & \xrightarrow{g} & Z. \end{array}$$

(BM3) Compatibilité de f_* et f^* au produit externe.

(PB) Soit E un fibré vectoriel de rang n sur X dans \mathcal{V}_k , le morphisme canonique

$$\bigoplus_{i=0}^{n-1} \xi^{(i)} : \bigoplus_{i=0}^{n-1} H_{*-n+i}(X) \longrightarrow H_*(\mathbf{P}(E))$$

est un isomorphisme, le morphisme $\xi^{(i)} : H_{*-n+i}(X) \rightarrow H_*(\mathbf{P}(E))$ étant défini comme en (3.3) par $\xi^{(i)} := \tilde{c}_1(O(1))^i \circ q^*$.

(H) Soit X dans \mathcal{V}_k , soit E un fibré vectoriel de rang n sur X et soit $p : V \rightarrow X$ un E -torseur. Alors $p^* : H_*(X) \rightarrow H_{*+n}(V)$ est un isomorphisme.

Bien sûr un tel foncteur est un foncteur de Borel-Moore orienté avec produits au sens de (2.2.3), la première classe de Chern d'un fibré en droites \mathcal{L} sur une variété X étant définie par $\tilde{c}_1(\mathcal{L}) := i^* \circ i_*$, pour $i : X \rightarrow \mathcal{L}$ la section nulle (ce qui permet de définir $\xi^{(i)}$ dans (PB)).

⁽³⁾i.e. qui admet une factorisation $f = p \circ i$ avec i une immersion régulière fermée et p un morphisme lisse.

6.2. Images inverses en cobordisme algébrique

La théorie de l'intersection de Fulton [5] permet de munir le foncteur $X \mapsto \mathrm{CH}_*(X)$ d'une structure de théorie de Borel-Moore orientée. On peut également munir le foncteur $G_0[\beta, \beta^{-1}]$ d'une telle structure.

6.2.1. THÉORÈME. — *Si k est de caractéristique zéro, le cobordisme algébrique Ω_* admet une unique structure de théorie de Borel-Moore orientée compatible avec les données déjà définies. De plus le cobordisme algébrique est universel parmi les théories de Borel-Moore orientées.*

Le point-clé dans cet énoncé est la construction d'images inverses fonctorielles $f^* : \Omega_*(X) \rightarrow \Omega_{*+d}(Y)$ pour $f : Y \rightarrow X$ une immersion régulière fermée de dimension relative d vérifiant les propriétés demandées. Ce travail est mené à bien dans [13], et nécessite de reprendre en les étendant au cadre du cobordisme algébrique l'essentiel des constructions de la théorie de l'intersection de [5].

7. RELATIONS AVEC LA THÉORIE HOMOTOPIQUE DES SCHÉMAS

7.1. Le spectre du cobordisme complexe

Commençons par rappeler l'interprétation homotopique du cobordisme complexe.

On note $G_{n,m}$ la grassmannienne des n -plans dans \mathbf{C}^{n+m} et $\xi_{n,m}$ le fibré complexe tautologique de rang n sur $G_{n,m}$. Le classifiant $\mathrm{BU}(n)$ du groupe $U(n)$ (et des fibrés complexes de rang n) est la colimite des espaces $G_{n,m}$. Il est muni du fibré ξ_n , colimite des fibrés $\xi_{n,m}$. On note $\mathrm{MU}(n)$ l'espace de Thom du fibré ξ_n et on définit \mathbf{MU} , le spectre du cobordisme complexe, par $\mathbf{MU}_{2n} = \mathrm{MU}(n)$ et $\mathbf{MU}_{2n+1} = \Sigma \mathrm{MU}(n)$, Σ désignant le foncteur de suspension.

L'énoncé suivant est une reformulation du théorème de Thom sur le cobordisme (dans le cas complexe).

7.1.1. THÉORÈME. — *Pour toute variété X , $U^q(X)$ est canoniquement isomorphe à $\mathrm{colim} [\Sigma^{2k-q} X, \mathrm{MU}(k)]$. Autrement dit, le spectre \mathbf{MU} représente⁽⁴⁾ la théorie cohomologique U^* dans la catégorie homotopique stable.*

⁽⁴⁾Au grain de sel près que les objets de la catégorie homotopique stable ne sont pas tous des variétés différentiables.

7.2. Le spectre du cobordisme algébrique

Rappelons la construction de la catégorie homotopique stable en géométrie algébrique d'après [23], [18]. Morel et Voevodsky ont défini dans [19] la catégorie homotopique $\mathcal{H}(k)$ des schémas lisses sur un corps k . C'est une catégorie de modèles fermée dont les objets sont des faisceaux simpliciaux pour la topologie de Nisnevich. En particulier, tout fibré vectoriel \mathcal{E} sur une k -variété admet un espace de Thom $\mathrm{Th}(\mathcal{E})$ dans $\mathcal{H}(k)$. Ainsi on peut définir dans la catégorie $\mathcal{H}(k)$ l'analogue noté $\mathrm{MGL}(n)$ de l'espace $\mathrm{MU}(n)$ considéré en (7.1). Remarquons que dans la catégorie homotopique $\mathcal{H}(k)$ on dispose de deux « cercles » différents, le cercle simplicial S_s^1 et le cercle de Tate $S_t^1 := \mathbf{A}^1 \setminus \{0\}$. On vérifie que leur smash-produit T est faiblement équivalent à la droite projective. La catégorie homotopique stable $\mathcal{SH}(k)$ est définie⁽⁵⁾ en localisant par rapport au smash-produit $\Sigma_T : X \mapsto T \wedge X$. On définit alors les T -spectres de façon usuelle, comme une suite d'objets E_i munis de morphismes $\Sigma_T(E_i) \rightarrow E_{i+1}$. À tout T -spectre \mathbf{E} est associée une théorie cohomologique bigraduée

$$E^{p,q}(X) := \mathrm{Hom}_{\mathcal{SH}(k)}(\Sigma^\infty, S^{p,q} \wedge \mathbf{E})$$

sur $\mathcal{H}(k)$, en posant $S^{p,q} := (S_s^1)^{p-q} \wedge (S_t^1)^q$.

En particulier, on peut associer naturellement à la suite d'espaces $\mathrm{MGL}(n)$ un T -spectre noté \mathbf{MGL} et une théorie cohomologique bigraduée $\mathrm{MGL}^{*,*}$. Ceci illustre le principe général selon lequel en géométrie algébrique les théories cohomologiques sur la catégorie des variétés lisses sur un corps sont naturellement bigraduées. Ainsi la cohomologie singulière $H^*(X, \mathbf{Z})$ est-elle remplacée par la cohomologie motivique $H^{*,*}(X, \mathbf{Z})$, la K -théorie complexe par la K -théorie algébrique $K^{*,*}(X)$ (avec $K^{n,i}(X) = K_{2i-n}^{\mathrm{Quillen}}(X)$), \dots , le premier degré correspondant au degré cohomologique et le second au poids. En particulier, on obtient, pour X lisse, un morphisme

$$\Omega^*(X) \longrightarrow \mathrm{MGL}^{2*,*}(X)$$

dont Levine et Morel conjecturent qu'il est toujours un isomorphisme.

Pour conclure, signalons que Hopkins et Morel ont annoncé, sans imposer d'hypothèse sur k , que pour toute variété lisse (simpliciale) X sur k on a un isomorphisme

$$\mathrm{MGL}^{*,*}(X) \otimes_{\mathbf{L}^*} \mathbf{Z}[\beta, \beta^{-1}] \simeq K^{*,*}(X) \otimes_{\mathbf{Z}} \mathbf{Z}[\beta, \beta^{-1}],$$

⁽⁵⁾Pour être correct il faudrait considérer ici des espaces pointés, mais nous négligeons délibérément ce point.

β étant maintenant de bidegré $(-2, -1)$. Ils ont également annoncé, sous l'hypothèse que le corps k est de caractéristique zéro, l'existence d'une suite spectrale de type Atiyah-Hirzebruch reliant la cohomologie motivique à $MGL^{*,*}$.

Références

- [1] D. Abramovich, K. Karu, K. Matsuki, J. Włodarczyk – *Torification and factorization of birational morphisms*, preprint 2000, AG/9904135.
- [2] J.F. Adams – *Stable homotopy and generalised homology*, University of Chicago Press, 1974.
- [3] L. Bonavero – *Factorisation faible des applications birationnelles*, Séminaire Bourbaki, Exp. No. 880, Novembre 2000.
- [4] S. Borghesi – *Algebraic Morava K-theories and the higher degree formula*, preprint 2000, disponible à l'adresse <http://www.math.uiuc.edu/K-theory/0412/index.html>.
- [5] W. Fulton – *Intersection theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **2**, Springer-Verlag, Berlin, 1984.
- [6] D. Hoffmann – *Isotropy of quadratic forms over the function field of a quadric*, Math. Z. **220** (1995), 461–476.
- [7] O. Izhbolin – *Motivic equivalence of quadratic forms II*, Man. Math. **102** (2000), 41–52.
- [8] M. Karoubi – *Cobordisme et groupes formels (d'après D. Quillen et T. tom Dieck)*, Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 408, 141–165. Lecture Notes in Math., Vol. 317 Springer, Berlin, 1973.
- [9] M. Lazard – *Sur les groupes de Lie formels à un paramètre*, Bull. Soc. Math. France **83** (1955), 251–274.
- [10] M. Levine, F. Morel – *Cobordisme algébrique I*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), 723–728.
- [11] M. Levine, F. Morel – *Cobordisme algébrique II*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), 815–820.
- [12] M. Levine, F. Morel – *Algebraic cobordism I*, preprint 2002, disponible à l'adresse <http://www.math.uiuc.edu/K-theory/0547/index.html>.
- [13] M. Levine – *Algebraic cobordism II*, en préparation.
- [14] A. Merkurjev – *Degree Formula*, preprint 2000, disponible à l'adresse <http://www.math.ohio-state.edu/~rost/chain-lemma.html>.
- [15] A. Merkurjev – *Rost's degree formula (Notes of mini-course in Lens, June 2001)*, preprint 2001, disponible à l'adresse <http://www.math.ucla.edu/~merkurev/publicat.htm>.
- [16] A. Merkurjev – *Algebraic oriented cohomology theories*, preprint 2002, disponible à l'adresse <http://www.math.uiuc.edu/K-theory/0535/index.html>.
- [17] J. Milnor – *On the cobordism ring Ω^* and a complex analogue. I*, Amer. J. Math. **82** **1960**, 505–521.
- [18] F. Morel – *Voevodsky's proof of Milnor's conjecture*, Bull. Amer. Math. Soc. **35** (1998), 123–143.

- [19] F. Morel, V. Voevodsky – \mathbf{A}^1 -homotopy theory of schemes, Inst. Hautes Études Sci. Publ. Math. **90** (1999), 45–143.
- [20] D. Quillen – *Elementary proofs of some results of cobordism theory using Steenrod operations*, Advances in Math. **7** (1971), 29–56.
- [21] R. Thom – *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. **28** (1954), 17–86.
- [22] B. Totaro – *Torsion algebraic cycles and complex cobordism*, J. Amer. Math. Soc. **10** (1997), 467–493.
- [23] V. Voevodsky – *The Milnor Conjecture*, preprint 1996, disponible à l’adresse <http://www.math.uiuc.edu/K-theory/0170/index.html>.

François LOESER

École normale supérieure,
Département de mathématiques
et applications,
45 rue d’Ulm,
75230 Paris Cedex 05, France
(UMR 8553 du CNRS)

E-mail : Francois.Loesper@ens.fr

URL : <http://www.dma.ens.fr/~loesper/>

LA CONJECTURE DE KATO

[d'après Pascal Auscher, Steve Hofmann, Michael Lacey, John Lewis,
Alan McIntosh et Philippe Tchamitchian]

par Yves MEYER

1. INTRODUCTION

La conjecture de Kato concerne les racines carrées d'opérateurs accréatifs. Cette conjecture est magique au sens suivant : son énoncé est innocent et ne fait appel qu'à des notions très naturelles concernant les opérateurs agissant sur un espace de Hilbert. Il s'agit de démontrer que le domaine de la racine carrée d'un opérateur accréatif coïncide avec celui de la forme ayant servi à construire l'opérateur. Mais la conjecture de Kato implique d'autres conjectures importantes. Par exemple, un cas particulier, en dimension 1, de la conjecture de Kato entraîne une conjecture de Calderón portant sur la continuité du noyau de Cauchy pour les graphes lipschitziens. Quand Alan McIntosh observa ce fait en 1980, il traça du même coup le chemin conduisant à la preuve de la conjecture de Calderón. Il ne restait plus qu'à terminer la démonstration en ramenant le problème à une estimation portant sur une fonctionnelle quadratique et en utilisant alors la technique des « mesures de Carleson » (section 6). Les mesures de Carleson ont été introduites par Lennart Carleson pour résoudre un problème portant sur les idéaux de fonctions holomorphes et bornées dans le disque unité. Leur usage en théorie des opérateurs a été systématisé par R. Coifman et ses collaborateurs. Ces mêmes estimations quadratiques et ces mêmes mesures de Carleson joueront un rôle essentiel dans la résolution de la conjecture de Kato.

2. LA PREMIÈRE VERSION DE LA CONJECTURE DE KATO

Commençons par une définition introduite par Ralph Phillips en 1957. La partie réelle du nombre complexe z est notée $\operatorname{Re} z$.

DÉFINITION 2.1. — Soit H un espace de Hilbert sur \mathbb{C} et désignons par $\langle x, y \rangle$ et $\|x\|$ la forme sesquilinéaire et la norme correspondantes. Soit V un sous-espace vectoriel de

H et $T : V \rightarrow H$ un opérateur linéaire. On dira que T est dissipatif si

$$(2.1) \quad \operatorname{Re} \langle T(x), x \rangle \leq 0, \quad x \in V.$$

L'opérateur T est dissipatif maximal s'il est dissipatif et s'il ne peut être prolongé en un opérateur dissipatif $\tilde{T} : \tilde{V} \rightarrow H$ où \tilde{V} contienne strictement V . Nous dirons alors que V est le domaine de T .

Rappelons également la définition d'un semi-groupe de contractions.

DÉFINITION 2.2. — Un semi-groupe de contractions est une famille $S(t)$, $t \geq 0$, d'opérateurs linéaires continus sur l'espace de Hilbert H vérifiant, pour tout $x \in H$,

$$(2.2) \quad S(t + \tau) = S(t)S(\tau), \quad t, \tau \geq 0$$

$$(2.3) \quad \|S(t)(x) - x\| \rightarrow 0, \quad t \rightarrow 0$$

$$(2.4) \quad \|S(t)(x)\| \leq \|x\|, \quad t \geq 0.$$

Avec ces notations, le théorème de Phillips est l'énoncé suivant

THÉORÈME 2.3. — Un opérateur T est le générateur infinitésimal d'un semi-groupe de contractions $S(t) : H \rightarrow H$ si et seulement si T est dissipatif maximal et si son domaine V est dense dans H .

Selon une suggestion de Friedrichs, un opérateur T est appelé accréatif si $-T$ est dissipatif.

Voici un exemple des situations que nous venons d'étudier. L'opérateur $\Delta = \frac{\partial^2}{\partial x_1^2} + \dots + \frac{\partial^2}{\partial x_n^2}$ est dissipatif quand $H = L^2(\kappa^n)$ et le semi-groupe associé $\exp(t\Delta)$ est le semi-groupe de la chaleur. Notre propos sera de définir la racine carrée d'un opérateur accréatif maximal. Dans le cas particulier du laplacien, cette racine carrée est l'opérateur de Calderón $\Lambda = \sqrt{-\Delta}$. Son domaine est celui de la forme de Dirichlet $B(u, v) = \int \nabla u \cdot \overline{\nabla v} \, dx$. Le semi-groupe $S(t) = \exp(-t\Lambda)$, $t \geq 0$, décrit l'évolution $u(x, 0) \rightarrow u(x, t)$ des fonctions harmoniques $u(x, t)$ dans $\Omega = \{(x, t) \in \kappa^n \times (0, \infty)\}$ qui sont nulles à l'infini.

Si la construction d'opérateurs accréatifs est aisée, celle d'opérateurs accréatifs maximaux est un peu plus délicate. On commence par la définition d'une forme sesquilinéaire accréative.

DÉFINITION 2.4. — Soit H un espace de Hilbert pour lequel le produit scalaire et la norme seront notés $\langle x, y \rangle$ et $\|x\|$. Soit H_1 un second espace de Hilbert. On désigne par $\|x\|_1$ la norme correspondante et l'on suppose que H_1 est un sous-espace vectoriel dense dans H , cette injection étant continue.

Soit $B(x, y)$ une forme sesquilinéaire définie et bicontinue sur $H_1 \times H_1$:

$$(2.5) \quad |B(x, y)| \leq C\|x\|_1\|y\|_1.$$

Nous dirons que B est accréitive si, pour tout x dans H_1 , on a

$$(2.6) \quad \operatorname{Re} B(x, x) \geq 0.$$

Si β est un nombre réel positif, nous dirons que $B(x, y)$ est β -accréitive si, pour tout $x \in H_1$, on a

$$(2.7) \quad \operatorname{Re} B(x, x) + \|x\|^2 \geq \beta \|x\|_1^2.$$

Dans ce cas, nous dirons que H_1 est le domaine de la forme $B(x, y)$.

Ceci étant, voici comment on construit des opérateurs accréitifs maximaux.

LEMME 2.5. — Soit $B : H_1 \times H_1 \rightarrow \mathbb{C}$ une forme sesquilinéaire β -accréitive. On désigne par V l'ensemble des $x \in H$ pour lesquels il existe une constante $C = C(x)$ de sorte que, pour tout $y \in H$, on ait

$$(2.8) \quad |B(x, y)| \leq C \|y\|.$$

Alors l'opérateur $T : V \rightarrow H$ défini par

$$(2.9) \quad \langle T(x), y \rangle = B(x, y), \quad x \in V, \quad y \in H$$

est accréitif maximal.

En restant toujours dans un cadre hilbertien général, voici la définition, puis la construction de la racine carrée accréitive d'un opérateur accréitif maximal.

DÉFINITION 2.6. — Soit H un espace de Hilbert et $T : V \rightarrow H$ un opérateur accréitif maximal de domaine $V \subset H$. Alors il existe un et un seul opérateur accréitif maximal S tel que $S^2 = T$. On écrit $S = \sqrt{T}$. En outre, sur le domaine V de T , on a

$$(2.10) \quad S = \frac{2}{\pi} \int_0^\infty (1 + \lambda^2 T)^{-1} T \, d\lambda = \frac{8}{\pi} \int_0^\infty (1 + \lambda^2 T)^{-3} \lambda^2 T^2 \, d\lambda.$$

Le problème posé par Kato dans [7] est la question suivante

Supposons que l'opérateur maximal accréitif T soit défini à l'aide d'une forme β -accréitive B au sens du lemme 2.5. Le domaine de la racine carrée accréitive maximale \sqrt{T} est-il alors le même que le domaine H_1 de la forme qui a servi à construire T ?

Ceci est évidemment vrai si l'opérateur T est auto-adjoint, c'est-à-dire si $B(u, u) \geq 0$ pour tout $u \in V$.

J.-L. Lions a démontré dans [7] que la conjecture de Kato est équivalente à l'énoncé suivant : il existe une constante C telle que, pour tout $x \in H_1$, on ait, en désignant par T^* l'adjoint de T

$$(2.11) \quad \|\sqrt{T}x\| \leq C \|x\|_1 \quad \|\sqrt{T^*}x\| \leq C \|x\|_1, \quad x \in H_1.$$

Alan McIntosh a trouvé un contre-exemple à cette version hilbertienne de la conjecture de Kato. Il restait à savoir si la conjecture de Kato est exacte dans le cas particulier qui a motivé le travail de Tosio Kato, celui où T est un opérateur différentiel accréatif du second ordre, écrit sous forme de divergence.

3. LA CONJECTURE DE KATO PRÉCISÉE

On pose $H = L^2(\mathfrak{R}^n)$ et $H_1 \subset H$ désignera l'espace de Sobolev $H^1(\mathfrak{R}^n)$ qui se compose des fonctions de carré intégrable dont le gradient, pris au sens des distributions, est aussi de carré intégrable. Ensuite on pose $\partial_j = \frac{\partial}{\partial x_j}$ et l'on considère n^2 fonctions $a_{j,k}(x)$, $1 \leq j, k \leq n$, à valeurs réelles ou complexes, mesurables et bornées (appartenant à $L^\infty(\mathfrak{R}^n)$). On suppose qu'il existe deux constantes $C \geq \beta > 0$ telles que l'on ait, pour tout $\xi = (\xi_1, \dots, \xi_n) \in \mathfrak{C}^n$ et pour presque tout $x \in \mathfrak{R}^n$,

$$(3.1) \quad \beta |\xi|^2 \leq \operatorname{Re} \left(\sum_{1 \leq j, k \leq n} a_{j,k}(x) \xi_k \bar{\xi}_j \right) \leq C |\xi|^2.$$

On désigne par $\Omega(\beta, C)$ l'ensemble des matrices $A(x) = (a_{j,k}(x))_{1 \leq j, k \leq n}$ qui vérifient (3.1) et $\|a_{j,k}\|_\infty \leq C$, $1 \leq j, k \leq n$. Ensuite Ω est la réunion des $\Omega(\beta, C)$, $0 < \beta \leq C < \infty$. Alors Ω est un ensemble ouvert dans $(L^\infty(\mathfrak{R}^n))^{n^2}$ de matrices $n \times n$.

LEMME 3.1. — *Si (3.1) est vérifiée,*

$$(3.2) \quad B(f, g) = \sum_{1 \leq j, k \leq n} \int a_{j,k}(x) \partial_k f(x) \overline{\partial_j g(x)} dx$$

est une forme sesquilinéaire $\min(1, \beta)$ -accréative dont le domaine est l'espace de Sobolev $H^1(\mathfrak{R}^n)$.

L'opérateur accréatif maximal défini par cette forme est alors

$$(3.3) \quad T(f) = - \sum_{1 \leq j, k \leq n} \partial_j \left(a_{j,k}(x) \partial_k f(x) \right)$$

et son domaine V est le sous-espace vectoriel de $H^1(\mathfrak{R}^n)$ caractérisé par la condition suivante : $f \in V$ si et seulement si les fonctions $g_j \in L^2(\mathfrak{R}^n)$ définies par $g_j(x) = \sum_{1 \leq k \leq n} a_{j,k}(x) \partial_k f(x)$ vérifient $\sum_{1 \leq j \leq n} \partial_j g_j \in L^2(\mathfrak{R}^n)$. Ce domaine dépend non linéairement des fonctions mesurables $a_{j,k}$ et n'est pas un espace fonctionnel classique.

Pascal Auscher et ses collaborateurs (S. Hofmann, M. Lacey, J. Lewis, A. McIntosh et Ph. Tchamitchian) ont démontré le résultat suivant

THÉORÈME 3.2. — *Soit T l'opérateur maximal accréatif défini par (3.3) lorsque la matrice $A(x)$ vérifie les conditions (3.1). Alors le domaine de l'opérateur maximal accréatif \sqrt{T} est l'espace de Sobolev $H^1(\mathfrak{R}^n)$ et l'application $\Phi : A \rightarrow \sqrt{T}$ est holomorphe sur l'ouvert Ω .*

La seconde assertion du théorème 3.2 conduit à étudier \sqrt{T} en utilisant son développement en série entière. Lorsque ce théorème était encore une conjecture, plusieurs équipes concurrentes établissaient la continuité des opérateurs multilinéaires $T_k(A, A, \dots, A)(f)$, $k \in \mathcal{N}$, apparaissant dans cette série entière, mais nous (R. Coifman, G. David, J-L. Journé, C. Kenig, Y.M., etc.) n'étions capables de démontrer la convergence de cette série que si la matrice $A(x)$ est suffisamment proche de l'identité: $\|A(x) - I\|_\infty \leq \varepsilon_n$ et le résultat le plus précis dans cette direction (fournissant la valeur la moins faible de la constante ε_n) avait été obtenu par J-L. Journé [6].

Si les fonctions $a_{j,k}$ sont suffisamment régulières, \sqrt{T} est alors un opérateur pseudo-différentiel classique d'ordre 1 et, à ce titre, est continu sur l'espace de Sobolev H^1 . La même remarque s'applique à T^* et le théorème de J-L. Lions implique alors la conjecture de Kato.

Dans le cas général où les fonctions $a_{j,k}(x)$ sont mesurables et bornées, on ne peut plus utiliser les ressources du calcul pseudo-différentiel. Les méthodes hilbertiennes abstraites sont également insuffisantes. Des outils plus puissants sont nécessaires. Ces outils ont été forgés par Alberto Calderón et «l'École de Chicago», mais il faut aussi mentionner les travaux de De Giorgi, Moser, Carleson, etc. Le programme de Calderón a consisté à s'affranchir des hypothèses superflues de régularité qui limitent la portée du calcul pseudo-différentiel classique et à construire des algèbres incluant à la fois les opérateurs de multiplication ponctuelle par des fonctions mesurables et bornées, comme les $a_{j,k}(x)$, et les opérateurs de dérivation, comme les ∂_j .

Après avoir attaqué la conjecture de Kato dans la perspective donnée par le programme de Calderón (voir le théorème 5.3), Pascal Auscher et Philippe Tchamitchian ont finalement décidé d'aborder cette conjecture par une voie un peu différente, en s'inspirant des travaux de De Giorgi et de Moser que nous rappelons maintenant.

4. ENNIO DE GIORGI ET JÜRGEN MOSER

En 1956, E. De Giorgi étudia la régularité des solutions faibles u de l'équation aux dérivées partielles $T(u) = 0$ où T est défini par (3.3) et où les coefficients $a_{j,k}(x)$ sont réels, mesurables et bornés. Il démontra que toute solution faible d'une telle équation aux dérivées partielles est höldérienne d'exposant $\mu > 0$ et que μ ne dépend que des normes L^∞ des coefficients, de la constante d'accrétivité et de la dimension.

Plus précisément on a, en posant $B_R = \{x; |x - x_0| < R\}$ et $B_\rho = \{x; |x - x_0| < \rho\}$

THÉORÈME 4.1. — Soit T l'opérateur accréatif maximal défini par (3.3). Supposons que les fonctions $a_{j,k}(x)$, $1 \leq j, k \leq n$ soient à valeurs réelles. Alors il existe un exposant positif μ et une constante positive C tels que, pour tout $\rho \in]0, R[$ et pour toute solution $u \in H^1(B_R)$ de $T(u) = 0$, on ait

$$(4.1) \quad \int_{B_\rho} |\nabla u|^2 dx \leq C \left(\frac{\rho}{R}\right)^{n-2+2\mu} \int_{B_R} |\nabla u|^2 dx.$$

En outre μ et C ne dépendent que des normes L^∞ des fonctions $a_{j,k}(x)$, de la constante β d'ellipticité et de la dimension n .

La propriété (4.1) implique la régularité höldérienne. Si les fonctions $a_{j,k}(x)$ sont à valeurs complexes, le théorème 3 cesse d'être vrai, comme l'ont établi Maz'ya, Nazarov et Plamenevskii [9] en dimension $n \geq 5$. Nous dirons que l'opérateur T a la propriété de Dirichlet si (4.1) est vérifiée.

Une autre définition utile concerne le « noyau de la chaleur » $K_t(x, y)$ défini par

$$(4.2) \quad \exp(-tT)[f(x)] = \int K_t(x, y) f(y) dy.$$

On peut penser que sous l'hypothèse (3.1) le noyau de la chaleur $K_t(x, y)$ possède des propriétés analogues à celles du noyau de la chaleur usuel (correspondant à $A(x) = I$). Cette remarque conduit à la définition suivante

DÉFINITION 4.2. — L'opérateur accréatif maximal T est de type gaussien s'il existe un exposant $\mu \in (0, 1]$ et deux constantes positives $\beta > 0, c > 0$ tels que l'on ait d'une part

$$(4.3) \quad |K_t(x, y)| \leq ct^{-n/2} \exp\left(-\beta \frac{|x-y|^2}{t}\right)$$

$$(4.4) \quad |K_t(x, y) - K_t(x+h, y)| \leq c \left(\frac{|h|}{t^{1/2} + |x-y|}\right)^\mu t^{-n/2} \exp\left(-\beta \frac{|x-y|^2}{t}\right)$$

pour $t > 0, x, y, h \in \mathfrak{R}^n$ tels que $|h| \leq t^{1/2} + |x-y|$. On impose, d'autre part, la propriété (4.4) en échangeant les rôles de x et y .

J. Nash, D. Aronson, E. Fabes et D. Stroock ont établi le théorème suivant

THÉORÈME 4.3. — Si les n^2 fonctions $a_{j,k}(x)$ sont toutes à valeurs réelles, alors l'opérateur accréatif maximal T défini par (3.3) est du type gaussien.

Ce théorème a été complété par un résultat remarquable dû à P. Auscher. En effet on a

THÉORÈME 4.4. — Revenons au cas où les $a_{j,k}(x)$ sont à valeurs réelles ou complexes. Alors T défini par (3.3) est de type gaussien si et seulement si T et T^* ont la propriété de Dirichlet.

Ceci a toujours lieu si $n = 1$ et $n = 2$, mais il existe des contre-exemples en dimension $n \geq 5$ [9].

Venons-en aux contributions de Jürgen Moser. On considère un opérateur accréatif maximal T défini par (3.3) et l'on suppose encore que les fonctions $a_{j,k}(x)$ soient réelles. On a alors

$$(4.5) \quad c_0 |\xi|^2 \leq \sum_{j,k} a_{j,k}(x) \xi_j \xi_k \leq \frac{1}{c_0} |\xi|^2$$

uniformément en x . En 1961, J. Moser utilisa les ressources de l'espace BMO qui venait d'être créé pour démontrer l'inégalité de Harnack qui s'énonce ainsi

THÉORÈME 4.5. — *Soit Ω un ouvert de \mathfrak{R}^n et $K \subset \Omega$ un ensemble compact. Il existe alors une constante $C = C(c_0, K, \Omega)$ telle que, pour toute fonction positive $f \in H^1(\Omega)$ vérifiant $T(f) = 0$ au sens des distributions, ont ait*

$$(4.6) \quad \sup\{f(x); x \in K\} \leq C \inf\{f(x); x \in K\}.$$

5. ALBERTO CALDERÓN

Intrigué par le théorème de De Giorgi, A. Calderón se proposait de le redémontrer en construisant de nouvelles algèbres d'opérateurs incluant les opérateurs accréatifs donnés par (3.1). Calderón s'est d'abord posé le problème d'établir l'estimation L^2 de base en utilisant des conditions minimales de régularité sur noyau-distribution $K(x, y)$ d'un opérateur T . On veut savoir si, pour toute fonction $f \in L^2(\mathfrak{R}^n)$, on a $\|T(f)\|_2 \leq C\|f\|_2$ quand $T(f)(x) = \text{v.p.} \int K(x, y)f(y) dy$. Lars Hörmander a proposé la condition suivante

$$(5.1) \quad \int_{\{|x-y| \geq 2|x'-x|\}} |K(x', y) - K(x, y)| dy \leq C.$$

Cette propriété, combinée à l'estimation $L^2(\mathfrak{R}^n)$ qui s'écrit

$$(5.2) \quad \|T(f)\|_2 \leq C\|f\|_2$$

entraîne les estimations $L^p(\mathfrak{R}^n)$ pour $2 \leq p < \infty$, ainsi que l'estimation limite (L^∞, BMO) où, là encore, BMO est l'espace de John et Nirenberg. C'est-à-dire que $f \in BMO$ si et seulement s'il existe une constante C telle que l'on ait

$$(5.3) \quad \sup_Q \left\{ \int_Q |f(x) - m_Q(f)|^2 \frac{dx}{|Q|} \right\} \leq C.$$

On a désigné par $|Q|$ le volume du cube Q , la borne supérieure est calculée sur tous les cubes $Q \subset \mathfrak{R}^n$ et $m_Q(f)$ désigne la moyenne de f sur Q .

DÉFINITION 5.1. — *Si à la fois T et son adjoint T^* vérifient (5.1) et (5.2), nous dirons que T est un opérateur de Calderón-Zygmund.*

Les opérateurs pseudo-différentiels classiques d'ordre 0 sont des opérateurs de Calderón-Zygmund, mais la réciproque n'est pas vraie. Les opérateurs de Calderón-Zygmund jouent un rôle fondamental dans la preuve, par Guy David, de la conjecture de Vitushkin (caractérisation des ensembles compacts du plan complexe de capacité analytique nulle et de mesure de Hausdorff unidimensionnelle finie comme étant les ensembles totalement non rectifiables).

Le problème essentiel dans l'étude des opérateurs de Calderón-Zygmund est celui de la continuité de l'opérateur T sur $L^2(\mathfrak{R}^n)$. Ce programme de travail a débouché sur le «théorème $T(1)$ » de David et Journé que nous énonçons sous une forme simplifiée

THÉORÈME 5.2. — *Supposons que l'on ait*

$$(5.4) \quad K(x, y) = -K(y, x), \quad |K(x, y)| \leq |x - y|^{-n}, \quad |\nabla K(x, y)| \leq |x - y|^{-(n+1)}.$$

Alors une condition nécessaire et suffisante de continuité de T sur $L^2(\mathfrak{R}^n)$ est $T(1) \in BMO$.

Les conditions imposées à $K(x, y)$ peuvent être considérablement allégées ; une régularité höldérienne suffit. Mais l'on ne sait pas démontrer le «théorème $T(1)$ » sous la condition (5.1). L'action de T sur la fonction identiquement égale à 1 doit être définie en donnant un sens à l'intégrale $g(x) = \int K(x, y)dy$. Cette fonction $g(x)$ est définie, modulo une fonction constante. En principe le théorème $T(1)$ pourrait s'appliquer à notre propos, grâce à la relation entre la conjecture de Kato et les opérateurs de Calderón-Zygmund. On a, en effet

THÉORÈME 5.3. — *En conservant les notations du théorème 4.4, supposons que T soit de type gaussien et que la conjecture de Kato soit vraie, c'est-à-dire que (2.11) soit vérifiée lorsque H_1 est l'espace de Sobolev $H^1(\mathfrak{R}^n)$. Alors il existe n opérateurs de Calderón-Zygmund U_1, \dots, U_n , tels que l'on ait*

$$(5.5) \quad \sqrt{T} = U_1 \partial_1 + \dots + U_n \partial_n.$$

Ce théorème a été établi par Auscher et Tchamitchian avant que la preuve complète de la conjecture de Kato ne soit obtenue. Ce théorème nous incite à démontrer la conjecture de Kato en utilisant le «théorème $T(1)$ » de David et Journé. La preuve générale doit beaucoup à cette remarque. Voici cependant des exemples où le théorème $T(1)$ ne permet pas de conclure. C'est, par exemple, le cas si T est l'opérateur de Cauchy sur une courbe lipschitzienne Γ qui est le graphe d'une fonction lipschitzienne $a(x)$. Dans ce cas, l'opérateur T est défini par le noyau singulier $vp \frac{1}{\pi} (z(x) - z(y))^{-1}$ où $z(x) = x + ia(x)$. Le calcul de $T(1)$ ne peut se faire et c'est pourquoi un nouveau théorème a été recherché,

puis démontré. Il s'agit du «théorème $T(b)$ » dont on trouvera la preuve dans [10]. On part d'une fonction $b(x)$, appartenant à L^∞ , qui est accréitive au sens où sa partie réelle $\operatorname{Re} b(x)$ vérifie $\operatorname{Re} b(x) \geq \beta > 0$. Dans ces conditions, le théorème $T(1)$ se généralise et l'on a

THÉORÈME 5.4. — *En conservant les hypothèses du théorème 5.2, supposons qu'il existe une fonction accréitive $b(x) \in L^\infty(\mathfrak{R}^n)$ telle que $T(b)$ appartienne à BMO . Alors T est borné sur $L^2(\mathfrak{R}^n)$.*

En revenant à l'opérateur de Cauchy, il suffit, pour appliquer le théorème 5.4, de choisir $b(x) = 1 + ia'(x)$. En effet, cette fonction satisfait évidemment la condition d'accrétivité et l'on a alors $T(b) = 0$, comme le montre immédiatement la formule de Cauchy. La force du «théorème $T(b)$ » réside dans la possibilité de choisir la fonction $b(x)$ «après coup», en prenant en compte les propriétés de l'opérateur que l'on étudie.

Une autre version du théorème $T(b)$ peut être trouvée dans un remarquable travail de M. Christ et J.L. Journé [4].

Comme nous le verrons dans la huitième section, la preuve de la conjecture de Kato dépend d'une troisième version (adaptée) du «théorème $T(b)$ ». Cette nouvelle version, qui étend celle de M. Christ et J.L. Journé, s'applique à des situations beaucoup plus générales que celles décrites dans le théorème 5.4. Il n'est plus nécessaire de supposer que T soit de type gaussien, ce qui, comme nous l'avons vu, n'est pas toujours vérifié.

6. LES MESURES DE CARLESON

La définition des mesures de Carleson apparaît pour la première fois dans la preuve par Lennart Carleson de la «conjecture de la couronne». Il s'agit de l'énoncé suivant : si $f_1(z), \dots, f_n(z)$ sont n fonctions holomorphes et bornées dans le disque unité D et s'il existe une constante C telle que l'on ait, pour tout $z \in D$,

$$(6.1) \quad 1 \leq \sum_{1 \leq k \leq n} |f_k(z)| \leq C$$

alors on peut trouver n autres fonctions holomorphes et bornées dans D , notées $g_1(z), \dots, g_n(z)$, telles que

$$(6.2) \quad \sum_{1 \leq k \leq n} f_k(z)g_k(z) = 1.$$

Pour démontrer ce théorème, Carleson définit ce qui est aujourd'hui connu sous le nom de mesures de Carleson et démontra le théorème 6.2 qui suit.

DÉFINITION 6.1. — Soit $\Omega \subset \mathfrak{R}^n$ un ensemble ouvert. On désigne alors par $\widehat{\Omega}$ l'ensemble des couples $(x, t) \in \mathfrak{R}^n \times]0, \infty[$ tels que la boule $B(x, t)$ de centre x et de rayon t soit incluse dans Ω . Soit $d\mu(x, t)$ une mesure de Radon positive sur $\mathfrak{R}^n \times]0, \infty[$. On dit que $d\mu$ est une mesure de Carleson s'il existe une constante C telle que pour toute partie ouverte $\Omega \subset \mathfrak{R}^n$, on ait

$$(6.3) \quad \mu(\widehat{\Omega}) \leq C|\Omega|,$$

où $|E|$ est la mesure de Lebesgue de $E \subset \mathfrak{R}^n$.

La plus petite constante C pouvant figurer dans (6.3) est la norme de la mesure de Carleson μ et sera notée $\|\mu\|_C$. Pour établir (6.3), il suffit de le faire si Ω est une boule arbitraire de \mathfrak{R}^n . Le passage au cas général utilise un recouvrement de Vitali de Ω par une suite de boules B_j , $j \in \mathcal{N}$. Ce faisant, la constante figurant dans le second membre de (6.3) aura changé.

Pour vérifier qu'une mesure de Radon positive $d\mu(x, t)$ est une mesure de Carleson, on utilise le plus souvent une variante de la condition (6.3). Si Q est une cube arbitraire de \mathfrak{R}^n de côté l_Q , on désignera par \tilde{Q} le cube $Q \times [0, l_Q]$ de \mathfrak{R}^{n+1} et la condition de Carleson s'écrit simplement $\mu(\tilde{Q}) \leq C|Q|$.

Nous arrivons maintenant au théorème annoncé. Il concerne le calcul des intégrales $I = \int_0^\infty \int_{\mathbb{R}^n} f(x, t) d\mu(x, t)$ où $f(x, t)$ est une fonction borélienne positive ou nulle et où $d\mu$ est une mesure de Carleson de norme $\|\mu\|_C$.

THÉORÈME 6.2. — *Posons*

$$(6.4) \quad f^*(x) = \sup_{|y-x| \leq t} f(y, t).$$

Alors on a

$$(6.5) \quad \int_0^\infty \int_{\mathbb{R}^n} f(x, t) d\mu(x, t) \leq \|\mu\|_C \int_{\mathbb{R}^n} f^*(x) dx.$$

7. LA PREUVE DE LA CONJECTURE DE KATO (PREMIÈRE PARTIE)

La preuve de la conjecture de Kato utilise deux idées maintenant classiques en théorie des opérateurs. La première qui a été développée par Elias Stein, mais qui peut être rattachée aux travaux de Littlewood et Paley, concerne l'utilisation de fonctionnelles quadratiques. Supposons qu'un opérateur T soit naturellement décomposé en une série $\sum_{j \in J} T_j$. Pour démontrer que cet opérateur T est borné sur $L^2(\mathfrak{R}^n)$, c'est-à-dire pour établir

l'estimation fondamentale $\|T(f)\|_2 \leq C\|f\|_2$, il suffit, dans certains cas, de démontrer que

$$(7.1) \quad \left(\sum_{j \in J} \|T_j(f)\|_2^2 \right)^{1/2} \leq C\|f\|_2.$$

On dira alors que les morceaux $T_j, j \in J$, sont presque orthogonaux.

Le second ingrédient est l'utilisation des mesures de Carleson pour établir l'estimation quadratique (7.1). La preuve du théorème $T(1)$ de David et Journé suit ce programme.

Commençons par la réduction à une fonctionnelle quadratique. On part de T défini par (3.3) et l'on considère sa racine carrée accréitive, définie par (2.10). Nous devons démontrer que l'on a

$$(7.2) \quad \|\sqrt{T}(f)\|_2 \leq C\|\nabla f\|_2$$

et cela nous amène à calculer

$$(7.3) \quad \sup\{|\langle \sqrt{T}(f), g \rangle|; \|g\|_2 \leq 1\}.$$

Mais ce produit scalaire s'écrit, grâce à (2.10),

$$(7.4) \quad \langle \sqrt{T}(f), g \rangle = \frac{8}{\pi} \int_0^\infty \langle W_\lambda(f), V_\lambda(g) \rangle \frac{d\lambda}{\lambda},$$

où

$$(7.5) \quad W_\lambda = (1 + \lambda^2 T)^{-1} \lambda T, \quad V_\lambda = (1 + \lambda^2 T^*)^{-2} \lambda^2 T^*.$$

Arrivés à ce point, on applique les estimations quadratiques « abstraites » de McIntosh et Yagi ([1] est la meilleure référence) qui fournissent

$$(7.6) \quad \int_0^\infty \|V_\lambda(g)\|_2^2 \frac{d\lambda}{\lambda} \leq C\|g\|_2^2.$$

La conjecture de Kato résultera donc de l'estimation quadratique

$$(7.7) \quad \int_0^\infty \|W_\lambda(f)\|_2^2 \frac{d\lambda}{\lambda} \leq C\|\nabla f\|_2^2.$$

L'opérateur Θ_λ est défini sur les n -uples de fonctions de $L^2(\mathfrak{R}^n)$ par

$$(7.8) \quad \Theta_\lambda(f_1, \dots, f_n)(x) = \lambda(1 + \lambda^2 T)^{-1} \sum_{j,k} \partial_j [(a_{j,k}(x))(f_k(x))].$$

Alors (7.7) s'écrit

$$(7.9) \quad \int_0^\infty \|\Theta_\lambda \nabla f\|_2^2 \frac{d\lambda}{\lambda} \leq C\|\nabla f\|_2^2.$$

Nous désignons ensuite par $\varphi(x)$ une fonction de la classe de Schwartz, portée par la boule unité, d'intégrale égale à 1 et par Φ_λ l'opérateur de convolution avec $\varphi_\lambda(x) = \lambda^{-n}\varphi(x/\lambda)$. On pose ensuite

$$(7.10) \quad \gamma_\lambda(x) = [(1 + \lambda^2 T)^{-1} \lambda \partial_j a_{j,k}(x)]_{1 \leq k \leq n}.$$

En utilisant des estimations sur l'action «à longue portée» de l'opérateur $(I + \lambda^2 T)^{-1}$, on obtient

$$(7.11) \quad \Theta_\lambda(\nabla f)(x) = \gamma_\lambda(x) \cdot \Phi_\lambda \nabla f(x) + \rho(x, \lambda),$$

où

$$(7.12) \quad \int_0^\infty \int_{\mathbb{R}^n} |\rho(x, \lambda)|^2 dx \frac{d\lambda}{\lambda} \leq C \|\nabla f\|_2^2.$$

La conjecture de Kato découlera alors de

$$(7.13) \quad \int_0^\infty \int_{\mathbb{R}^n} |\gamma_\lambda(x)|^2 |\Phi_\lambda \nabla f(x)|^2 dx \frac{d\lambda}{\lambda} \leq C \|\nabla f\|_2^2.$$

Le théorème de Hardy et Littlewood sur la fonction maximale implique

$$(7.14) \quad \left\| \sup_{|y-x| \leq \lambda} |\Phi_\lambda \nabla f(y)| \right\|_2 \leq C \|\nabla f\|_2.$$

La conjecture de Kato sera donc établie si nous montrons que la mesure $d\mu(x, \lambda) = |\gamma_\lambda|^2 dx \frac{d\lambda}{\lambda}$ est une mesure de Carleson. La preuve de ce fait est classique. Partant d'une boule arbitraire B de \mathbb{R}^n , on découpe la fonction mesurable et bornée $a_{j,k}(x)$ en deux morceaux. Le premier (noté u) est porté par la boule double \tilde{B} , de même centre que B et de rayon double et le second (noté v) est nul sur \tilde{B} . Pour traiter la contribution de u dans la mesure $d\mu$, on observe que $\|u\|_2 \leq C \|a_{j,k}\|_\infty |B|^{1/2}$ et l'estimation recherchée proviendra de

$$(7.15) \quad \int_0^\infty \|(1 + \lambda^2 T)^{-1} \lambda \nabla f\|_2^2 \frac{d\lambda}{\lambda} \leq C \|f\|_2^2.$$

Mais cette dernière estimation n'est autre que l'estimation (7.7) que nous nous proposons de démontrer. Nous sommes donc dans une situation de cercle vicieux !

8. UN THÉORÈME T(b) POUR LES MESURES DE CARLESON

Le paragraphe précédent nous a appris que la preuve de l'estimation (H^1, L^2) pour l'opérateur \sqrt{T} repose sur le fait que la mesure $d\mu(x, \lambda) = |\gamma_\lambda|^2 dx \frac{d\lambda}{\lambda}$ est une mesure de Carleson, ce qui semble essentiellement dépendre de l'estimation cherchée. P. Auscher et Ph. Tchamitchian ont réussi à sortir de cette impasse en démontrant un «théorème $T(b)$ pour les mesures de Carleson». La fonction accréitive b est ici remplacée par une collection F_Q de fonctions de test adaptées à l'opérateur T . Soyons plus précis. Nous désignons par

\mathcal{Q} l'ensemble de tous les cubes $Q \subset \mathfrak{r}^n$ et, pour tout cube $Q \in \mathcal{Q}$, de côté l_Q , nous supposons avoir pu construire une fonction F_Q ayant les cinq propriétés suivantes

$$(8.1) \quad F_Q \in H^1(5Q)$$

$$(8.2) \quad \int_{5Q} |\nabla F_Q|^2 dx \leq C|Q|$$

$$(8.3) \quad T(F_Q) = -\operatorname{div}[A(x)\nabla F_Q] \in L^2(5Q)$$

$$(8.4) \quad \int_{5Q} |T(F_Q)|^2 dx \leq Cl_Q^{-2}|Q|$$

$$(8.5) \quad \sup_{Q \in \mathcal{Q}} \int_0^{l_Q} \int_Q |\gamma_\lambda(x)|^2 dx \frac{d\lambda}{\lambda} \leq C \sup_{Q \in \mathcal{Q}} \int_0^{l_Q} \int_Q |\gamma_\lambda(x) \cdot \Phi_\lambda \nabla F_Q(x)|^2 dx \frac{d\lambda}{\lambda},$$

où C désigne une constante.

Le «théorème $T(b)$ d'Auscher et Tchamitchian» est l'énoncé suivant [1].

THÉORÈME 8.1. — *Sous les hypothèses (8.1) à (8.5), la mesure*

$$d\mu(x, \lambda) = |\gamma_\lambda(x)|^2 dx \frac{d\lambda}{\lambda}$$

est une mesure de Carleson, ce qui implique la conjecture de Kato.

Ce résultat est capital, car il a défini la stratégie qui conduisit à la preuve complète de la conjecture de Kato. Aujourd'hui, la preuve du théorème 8.1 peut sembler assez naturelle. Nous devons estimer le membre de gauche de (8.5). Nous utilisons (8.3) et (7.11). Ceci, joint aux estimations de l'action «à longue portée» de l'opérateur $(1 + \lambda^2 T)^{-1}$, fournit l'estimation désirée.

Mais il reste à construire les «fonctions de test» F_Q , $Q \in \mathcal{Q}$. Cette construction repose sur de remarquables méthodes de temps d'arrêts (stopping time arguments) découvertes par S. Hofmann et J. Lewis [5] dans un contexte très différent. Cela suffisait pour démontrer la conjecture de Kato en dimension 2. La preuve finale [3] doit beaucoup aux améliorations apportées par M. Lacey.

BIBLIOGRAPHIE

- [1] P. Auscher and Ph. Tchamitchian – *Square root problem for divergence operators and related topics*, Astérisque Vol. 249 (1998), Société Mathématique de France.
- [2] P. Auscher, S. Hofmann, J. Lewis and Ph. Tchamitchian – *Extrapolation of Carleson measures and the analyticity of Kato's square root operator*, Acta Math., à paraître.

- [3] P. Auscher, S. Hofmann, M. Lacey, A. McIntosh, and Ph. Tchamitchian – *The solution of the Kato square root problem for second order elliptic operators on \mathfrak{R}^n* , à paraître aux Annals of Maths.
- [4] M. Christ and J.L.Journé – *Polynomial growth estimates for multilinear singular operators*, Acta Math. **159** (1987), 51-80.
- [5] S. Hofmann and J.L. Lewis – *The Dirichlet problem for parabolic operators with singular drift terms*, à paraître aux Memoirs of the Amer. Math. Soc.
- [6] J-L. Journé – *Remarks on the square root problem*, Pub. Math. **35** (1991), 299-321.
- [7] T. Kato – *Fractional powers of dissipative operators*, J. Math. Soc. Japan **13** (1961), 246-274.
- [8] J-L. Lions – *Espaces d'interpolation et domaines de puissances fractionnaires*, J. Math. Soc. Japan **14** (1962), 233-241.
- [9] V. G. Maz'ya, S. A. Nazarov and B. A. Plamenevskii – *Absence of the De Giorgi-type theorems for strongly elliptic equations with complex coefficients*, J. Math. Sov. **28** (1985), 726-739.
- [10] Y. Meyer et R. Coifman – *Wavelets, Calderón-Zygmund operators and multilinear operators*, Cambridge Studies in advanced mathematics **48** (1997).

Yves MEYER

École Normale Supérieure

C.M.L.A.

61 avenue du Président Wilson

F-94235 CACHAN Cedex

E-mail : Yves.Meyer@cmla.ens-cachan.fr

ON THE NEWTON STRATIFICATION

by Michael RAPOPORT

INTRODUCTION

This is a report on algebraic geometry in characteristic p . Let A/S be a family of abelian varieties over a base scheme of characteristic p . For any prime number $\ell \neq p$ the family of Tate modules $T_\ell(A_{\bar{s}})$ (\bar{s} ranging over the geometric points of S) defines a local system of \mathbf{Z}_ℓ -modules on S . The replacement for $\ell = p$ of the Tate module $T_\ell(A_{\bar{s}})$ is the *Dieudonné module* $M(A_{\bar{s}})$ which is an F -crystal. However, in contrast to the ℓ -adic case, the Dieudonné module is not locally constant as \bar{s} varies over the base. This leads to the *Newton stratification* of S into locally closed subsets where the isomorphism classes of the *rational Dieudonné modules* $M(A_{\bar{s}}) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p$ are constant. Recently de Jong and Oort proved some general qualitative facts on this stratification. These were applied by Oort to the universal family of abelian varieties over the Siegel moduli space. We formulate this result in imprecise terms as follows.

THEOREM 0.1. — *The Newton stratification of the moduli space of principally polarized abelian varieties of a fixed dimension g in characteristic p has the strong stratification property (the closure of a stratum is a union of strata). Furthermore, the jumps in this stratification all occur in codimension one.*

The main tools in the proof of this theorem are the purity theorem on families of F -isocrystals and deformation theory of p -divisible groups. The latter is based on the theory of displays for formal p -divisible groups, which has recently been completed by Zink.

The layout of the report is as follows. In section 1 we introduce the notion of an F -isocrystal over a base scheme S and of the corresponding Newton stratification. Section 2 is devoted to the general theorems on Newton stratifications, and section 3 to the particular case of the Siegel moduli space. In section 4 we give the main theorem of display theory. In the final section 5 we comment on other moduli spaces of abelian varieties.

I thank Th. Zink for his help with this report; the presentation in section 2 is largely based on his explanations. I also thank G. Laumon, F. Oort and T. Wedhorn for useful comments.

The subject matter of this report has deep historical roots, with contributions by many mathematicians. I apologize in advance for any oversights and misrepresentations, which are not intentional but rather due to my ignorance.

1. F -CRYSTALS

DEFINITION 1.1. — *Let k be a perfect field of characteristic p , with ring of Witt vectors $W(k)$. Let L be the fraction field of $W(k)$ and denote by σ the Frobenius automorphisms on k , $W(k)$ and L .*

a) *A (non-degenerate) F -crystal over $\text{Spec } k$ is a free $W(k)$ -module M of finite rank with a σ -linear endomorphism $F : M \rightarrow M$ such that $M/F(M)$ has finite length.*

b) *An F -isocrystal over $\text{Spec } k$ is a finite-dimensional L -vector space N with a σ -linear bijective endomorphism $F : N \rightarrow N$.*

Recall that $W(k)$ is the unique complete discrete valuation ring with residue field k and with p as uniformizer. An F -crystal (M, F) defines an F -isocrystal via $(N, F) := (M, F) \otimes_{W(k)} L$. Conversely, given an F -isocrystal (N, F) , the corresponding set of F -crystals is the set of $W(k)$ -lattices M in N such that $F(M) \subset M$ (such lattices need not exist).

The F -isocrystals over $\text{Spec } k$ form a category in the obvious way which is abelian \mathbb{Q}_p -linear and noetherian and artinian. If k' is a perfect field extension of k , then an F -crystal over $\text{Spec } k$ defines an F -crystal over $\text{Spec } k'$ via base extension $\otimes_{W(k)} W(k')$.

THEOREM 1.2 (Dieudonné). — *Let k be algebraically closed. Then the category of F -isocrystals is semi-simple. The simple objects are parametrized by the set of rational numbers. To $\lambda \in \mathbb{Q}$ corresponds the simple object E_λ defined as follows. If $\lambda = \frac{r}{s}$, with $s, r \in \mathbb{Z}$, $s > 0$, $(r, s) = 1$, then*

$$E_\lambda = \left(L^s, F = \begin{bmatrix} 0 & & & p^r \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix} \cdot \sigma \right) .$$

Furthermore

$$\text{End}(E_\lambda) = D_\lambda ,$$

where D_λ is the division algebra with center \mathbb{Q}_p and invariant equal to the image of λ in \mathbb{Q}/\mathbb{Z} . ▷

We may parametrize the F -isocrystals of rank n over the algebraically closed field k by their Newton polygons, or preferably but equivalently, by their *Newton vectors*.

COROLLARY 1.3. — *Let k be algebraically closed. Then there is an injection (the Newton map)*

$$\{\text{isomorphism classes of } F\text{-isocrystals of rank } n\} \rightarrow (\mathbb{Q}^n)_+, (N, F) \mapsto \nu(N, F).$$

Here $(\mathbb{Q}^n)_+ = \{(\nu_1, \dots, \nu_n) \in \mathbb{Q}^n, \nu_1 \geq \dots \geq \nu_n\}$. The Newton map sends (N, F) to $\nu(N, F) \in (\mathbb{Q}^n)_+$, where $\lambda \in \mathbb{Q}$ occurs in $\nu(N, F)$ with multiplicity equal to the dimension of the isotypical component of type λ . The image of the Newton map may be described as follows. Write $\nu \in (\mathbb{Q}^n)_+$ as

$$\nu = (\nu(1)^{m_1}, \dots, \nu(r)^{m_r}) \quad \text{with } \nu(1) > \dots > \nu(r) \quad .$$

Then ν lies in the image if and only if ν satisfies the integrality condition

$$\nu(i)m_i \in \mathbb{Z}, \quad \forall i = 1, \dots, r \quad . \quad \triangleright$$

The components of the *Newton vector* $\nu(N, F)$ (i.e. the types of the isotypical components occurring in (N, F)) are called the *slopes* of the F -isocrystal.

Let (N, F) be an F -isocrystal over a perfect field k . Then the Newton vector of the F -isocrystal $(N, F) \otimes_{W(k)} W(\bar{k})$ over $\text{Spec } \bar{k}$ is independent of the algebraically closed field \bar{k} containing k . We may therefore speak of the Newton vector of (N, F) .

Let X be a p -divisible group over a perfect field k of characteristic p . Then one may associate to X its (contravariant) Dieudonné module $(M(X), F)$, which is an F -crystal over $\text{Spec } k$, compare [D]. *In this way one obtains*

a) an *anti-equivalence* of the category of p -divisible groups over $\text{Spec } k$ and the full subcategory of the category of F -crystals over $\text{Spec } k$ consisting of those F -crystals (M, F) such that $pM \subset FM$,

b) an *anti-equivalence* of the category of p -divisible groups over $\text{Spec } k$ up to isogeny and the full subcategory of all F -isocrystals over $\text{Spec } k$ such that all slopes lie between 0 and 1.

Let S be a scheme of characteristic p . An F -crystal over S is a crystal \mathcal{E} of finite locally free $\mathcal{O}_{S_{\text{cris}}}$ -modules, with a morphism $F : \mathcal{E}^{(\sigma)} \rightarrow \mathcal{E}$ such that the kernel and cokernel of F are annihilated by a power of p . Here $\mathcal{O}_{S_{\text{cris}}}$ denotes the structure sheaf on the big crystalline site of S over \mathbb{Z}_p . We often write \mathcal{E} for the F -crystal (\mathcal{E}, F) . This notion makes precise the intuitive concept of a family of F -crystals parametrized by the (perfect closures of the residue fields of) points of S . The F -crystals over S form a \mathbb{Z}_p -linear category. A morphism of F -crystals $f : \mathcal{E} \rightarrow \mathcal{E}'$ is an *isogeny* if there exists locally on S a morphism $g : \mathcal{E}' \rightarrow \mathcal{E}$ such that $gf = p^n$ and $fg = p^n$ for some n . The category of F -isocrystals over S is obtained by formally inverting isogenies of F -crystals.

EXAMPLE 1.4. — Let X be a p -divisible group over S . Then crystalline Dieudonné theory [Me] associates to X an F -crystal over S . More precisely, the Lie algebra of the universal extension of X is a crystal, and its dual is an F -crystal where F is induced by the Frobenius $Fr : X \rightarrow X^{(\sigma)}$, compare [Me], IV.2.5.

For the sequel it is not essential to have mastered the notion of an F -crystal over a scheme, in order to understand the resulting statements for p -divisible groups (although some proofs in this special case are based on general F -crystals for which one can perform the usual linear algebra operations like tensor products etc.).

The most basic statement about families of F -crystals is the following semi-continuity theorem. Recall the usual dominance order on $(\mathbb{Q}^n)_+$, for which $(\nu_1, \dots, \nu_n) \leq (\nu'_1, \dots, \nu'_n)$ if and only if

$$\sum_{j=1}^r \nu_j \leq \sum_{j=1}^r \nu'_j, \quad \forall r = 1, \dots, n-1, \quad \text{and} \quad \sum_{j=1}^n \nu_j = \sum_{j=1}^n \nu'_j .$$

For $\nu = (\nu_1, \dots, \nu_n) \in \mathbb{Q}^n$ we set $\|\nu\| = \sum_{j=1}^n \nu_j$.

THEOREM 1.5 (Grothendieck [G]). — *Let (\mathcal{E}, F) be an F -isocrystal over a scheme S of characteristic p . Then the Newton vector of (\mathcal{E}_s, F_s) , for s ranging over the points of S , goes down under specialization. More precisely, let (\mathcal{E}, F) be of constant rank n . Then the function $s \mapsto \|\nu(\mathcal{E}_s, F_s)\|$ is locally constant on S and for any $\nu_0 \in (\mathbb{Q}^n)_+$ the set*

$$\{s \in S; \nu(\mathcal{E}_s, F_s) \leq \nu_0\}$$

is Zariski closed in S .

We note here that the Newton vector of the fiber of \mathcal{E} at a geometric point \bar{s} of S only depends on the underlying point $s \in S$.

The proof by Katz in [Ka] relies on the relation between the Newton vector of an F -isocrystal and the divisibility by p of F with respect to an underlying F -crystal. For (the F -isocrystal associated to) a p -divisible group over S a simple proof is contained in [D]. ▷

REMARK 1.6. — This theorem is reminiscent of a theorem on vector bundles on a compact Riemann surface. In this theory one associates to a vector bundle of rank n its *Harder-Narasimhan vector* in $(\mathbb{Q}^n)_+$, and it is a basic fact that the HN-vector *goes up* (!) under specialization [AB].

COROLLARY 1.7. — *Let (\mathcal{E}, F) be an F -isocrystal over a noetherian scheme S of characteristic p . Then the set of points of S where the Newton vector is constant is locally closed in S and this defines a finite decomposition of S .*

Proof. — We may assume S connected. Let us only consider the case when (\mathcal{E}, F) comes from a p -divisible group X over S . Then the height and the dimension of X are constant. The assertion then follows from the preceding theorem by the following two observations.

For any $\nu_0 \in (\mathbb{Q}^n)_+$ the set

$$\{\nu \in (\mathbb{Q}^n)_+ ; \nu \leq \nu_0 \text{ and } \nu \text{ satisfies the integrality condition in Cor. 1.3}\}$$

is finite.

If $\nu = \nu(M(X), F)$ for a p -divisible group X of dimension d and height n , then $\nu \leq (1^d, 0^{n-d})$. This is a consequence of Mazur's inequality between the Hodge vector of an F -crystal over a perfect field and the Newton vector of its underlying F -isocrystal, [Ka].

▷

Let (\mathcal{E}, F) be an F -isocrystal of rank n over a noetherian scheme S of characteristic p . Associating to a geometric point \bar{s} of S the Newton vector of $(\mathcal{E}_{\bar{s}}, F_{\bar{s}})$, we obtain a map

$$S \longrightarrow (\mathbb{Q}^n)_+ .$$

Let S_ν be the fiber of this map over $\nu \in (\mathbb{Q}^n)_+$ (with its reduced scheme structure). The corresponding disjoint decomposition of S , finite according to Corollary 1.7, is called the *Newton stratification of S associated to the F -isocrystal (\mathcal{E}, F)* . The subschemes S_ν are called the *Newton strata*.

We speak of a stratification *in the strong sense* if the closure of a stratum is a union of strata. In general the Newton stratification associated to an F -isocrystal is *not* a stratification in the strong sense.

2. PURITY OF THE NEWTON STRATIFICATION

Let (\mathcal{E}, F) be an F -isocrystal of rank n over a scheme S of characteristic p , with associated Newton stratification $(S_\nu)_{\nu \in (\mathbb{Q}^n)_+}$ of S . The purity theorem states that the jumps in this stratification all occur in codimension one. The corresponding statement for families of vector bundles on a Riemann surface (cf. Remark 1.6) is false.

THEOREM 2.1 (de Jong, Oort [JO]). — *Let (\mathcal{E}, F) be an F -isocrystal of rank n over a locally noetherian scheme S of characteristic p , with associated Newton stratification $(S_\nu)_{\nu \in (\mathbb{Q}^n)_+}$. Let $\nu \in \mathbb{Q}^n_+$. Let η be a generic point of the scheme $\bar{S}_\nu \setminus S_\nu$. Then*

$$\dim \mathcal{O}_{\bar{S}_\nu, \eta} = 1 .$$

An equivalent statement is the following.

THEOREM 2.2. — *Let (\mathcal{E}, F) be an F -isocrystal over a locally noetherian scheme S of characteristic p . Let U be an open subset of S such that $\text{codim}(S \setminus U) \geq 2$. If the Newton vector of (\mathcal{E}, F) is constant at all points of U , then it is constant on all of S .*

These theorems are referred to as *purity theorems* since they are reminiscent of the purity theorem of Nagata-Zariski on étale coverings. We shall be mainly interested in this statement when (\mathcal{E}, F) comes from a p -divisible group over S . The structure of a p -divisible group with constant Newton vector is addressed in the following two results.

THEOREM 2.3 (de Jong, Oort [JO]). — *Let $S = \text{Spec } A$, where A is a complete noetherian local ring of characteristic p with algebraically closed residue field k . Let X be a p -divisible group over S with constant scalar Newton vector, i.e., there is only one slope at all points of S (isoclinic case). Then X is isogenous to a constant p -divisible group, i.e., one of the form $X_0 \times_{\text{Spec } k} S$ for a p -divisible group X_0 over $\text{Spec } k$.*

The motivation for this theorem is the heuristic idea that a p -divisible group with constant scalar Newton vector is analogous to a local system over S . The hypotheses on S in Theorem 2.3 ensure that it behaves like a simply connected space. One may expect a similar result for general F -isocrystals with constant scalar Newton vector, compare [JO], Remark 2.18. The case where A is a complete discrete valuation ring with algebraically closed residue field is due to Katz, [Ka], Thm. 2.7.1. ▷

The constancy up to isogeny becomes false when the constant Newton vector has more than one slope, for there can be highly nontrivial extensions of constant F -isocrystals over $\text{Spec } k[[t]]$ (e.g. the p -divisible group of the universal deformation of an ordinary elliptic curve is a nontrivial extension of $\mathbb{Q}_p/\mathbb{Z}_p$ by $\hat{\mathbf{G}}_m$). When there is more than one slope, there is the following analogue of the Harder-Narasimhan filtration of vector bundles, cf. Remark 1.6.

THEOREM 2.4 (Zink [Z3]). — *Let S be a regular scheme of characteristic p . Let X be a p -divisible group over S with constant Newton vector $\nu \in (\mathbb{Q}^n)_+$. Then X is isogenous to a p -divisible group Y which admits a filtration by closed embeddings of p -divisible groups*

$$(0) = Y_0 \subset Y_1 \subset \dots \subset Y_r = Y \quad ,$$

such that the following condition is satisfied. Let $\nu = (\nu(1)^{m_1}, \dots, \nu(r)^{m_r})$ with $\nu(1) > \dots > \nu(r)$, cf. integrality condition in Cor. 1.3. Then there are natural numbers $r_i \geq 0$, $s_i > 0$ ($i = 1, \dots, r$) such that $\nu(i) = r_i/s_i$ and such that

$$p^{-r_i} \text{Fr}^{s_i} : Y_i \longrightarrow Y_i^{(\sigma^{s_i})} \quad \text{is an isogeny}$$

and

$$p^{-r_i} \text{Fr}^{s_i} : Y_i/Y_{i-1} \longrightarrow (Y_i/Y_{i-1})^{(\sigma^{s_i})} \quad \text{is an isomorphism ,}$$

$\forall i = 1, \dots, r$.

The degree of the isogeny between X and Y may be bounded in terms of the height of X . The heuristic idea behind this theorem is that the isotypic *direct sum decomposition* of an F -isocrystal over an algebraically closed field is replaced in the case of a more general base scheme by a filtration. In ongoing work of Oort and Zink, the regularity

hypothesis on S is weakened. The case where $S = \text{Spec } k$ for an arbitrary field k is due to Grothendieck [G]. ▷

The proof of Theorem 2.2 is based on the following result which is of independent interest.

THEOREM 2.5 (de Jong, Oort [JO]). — *Let $S = \text{Spec } A$, where A is a normal complete noetherian local ring of dimension 2 with algebraically closed residue field k . Let $U = S \setminus \{s\}$, where s denotes the closed point. Let $\pi : \tilde{S} \rightarrow S$ be a resolution of singularities, i.e. a proper morphism from a regular scheme which induces an isomorphism over U and such that $E = \pi^{-1}(s) = \bigcup_{i=1}^m E_i$ is a union of smooth divisors crossing each other normally. Identifying $\pi^{-1}(U)$ with U we have the restriction map*

$$H_{\text{et}}^1(\tilde{S}, \mathbf{Z}_p) \longrightarrow H_{\text{et}}^1(U, \mathbf{Z}_p) \ .$$

This map is an isomorphism.

In terms of the fundamental groups (w.r.t. some geometric point of U) the assertion is that

$$(1) \quad \text{Hom}(\pi_1(\tilde{S}), \mathbf{Z}_p) \xrightarrow{\sim} \text{Hom}(\pi_1(U), \mathbf{Z}_p) \ .$$

Since both $\pi_1(\tilde{S})$ and $\pi_1(U)$ are factor groups of the Galois group of the fraction field of S , the homomorphism $\pi_1(U) \rightarrow \pi_1(\tilde{S})$ is surjective. Therefore we have the injectivity of the map (1), and to prove the surjectivity we may replace \mathbf{Z}_p by \mathbf{Q}_p in (1). Topologically or when $p \neq \text{char } k$, this surjectivity is easy to see. Indeed, we need the injectivity of

$$(2) \quad H_E^2(\tilde{S}, \mathbf{Q}_p) \longrightarrow H^2(\tilde{S}, \mathbf{Q}_p) \ .$$

But $H_E^2(\tilde{S}, \mathbf{Q}_p)$ has the classes $\text{cls}(E_i)$, $i = 1, \dots, m$, as basis (purity). The image of $\text{cls}(E_i)$ under the composition

$$H_E^2(\tilde{S}, \mathbf{Q}_p) \longrightarrow H^2(\tilde{S}, \mathbf{Q}_p) \xrightarrow{\text{Res}} H^2(E_j, \mathbf{Q}_p)$$

is the intersection product $(E_i \cdot E_j)$. The injectivity of (2) follows therefore from the negative-definiteness of the intersection matrix $(E_i \cdot E_j)_{i,j=1,\dots,m}$.

When $p = \text{char } k$, the proof of Theorem 2.5 is much more difficult. (That the situation in this case is radically different is already apparent from the fact that $H_{\text{et}}^2(\tilde{S}, \mathbf{Z}_p) = (0)$ when $p = \text{char } k$. This is easily checked using Artin-Schreier theory.) Suppose that A has characteristic p , i.e. $k \subset A$. In this case, using de Jong's technique of alterations there is a reduction to the following situation. Let $\mathcal{C} \rightarrow \text{Spec } k[[t]]$ be a flat projective family of curves with smooth generic fiber and strict semistable reduction. Let \mathcal{C}' be the scheme obtained from \mathcal{C} by collapsing a proper union E of irreducible components of the special fiber to a point P . Then A is the complete local ring of P .

One now starts with an element $\alpha \in H_{\text{et}}^1(U, \mathbf{Z}_p)$ and first globalizes it into an element $\alpha_1 \in H_{\text{et}}^1(\mathcal{C}' \setminus \{P\}, \mathbf{Z}_p) = H_{\text{et}}^1(\mathcal{C} \setminus E, \mathbf{Z}_p)$. This element α_1 is then extended to

$\alpha_2 \in H_{\text{ét}}^1(\mathcal{C}, \mathbb{Z}_p)$ by using de Jong's extension theorem on homomorphisms of p -divisible groups [J1], [J2]. According to this theorem, any homomorphism between the generic fibers of p -divisible groups over $\text{Spec } k[[t]]$ extends. \triangleright

Let X be a p -divisible group of height h over a scheme S of characteristic p , for which there exists $r \geq 0, s > 0$ such that

$$(3) \quad p^{-r} F r^s : X \longrightarrow X^{(\sigma^s)} \text{ is an isomorphism.}$$

To X we associate the lisse p -adic sheaf of $W(\mathbb{F}_{p^s})$ -modules $C_X = \varprojlim C_{X,n}$ for the étale topology on S , such that for any affine S -scheme $\text{Spec } R$

$$(4) \quad C_{X,n}(\text{Spec } R) = \{x \in M/p^n M; p^{-r} F^s(x) = x\} .$$

Here M denotes the $W(R)$ -module defined by the Dieudonné crystal of X . The fibers of C_X are free $W(\mathbb{F}_{p^s})$ -modules of rank h . The formation of C_X is compatible with base change and defines a functor from the category of p -divisible groups over S with (3) to the category of lisse p -adic sheaves of $W(\mathbb{F}_{p^s})$ -modules on S . The corresponding $W(\mathbb{F}_{p^s}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ -adic sheaf only depends on the isogeny class of X and corresponds to a representation of the fundamental group,

$$(5) \quad \varrho_X : \pi_1(S) \longrightarrow GL_h(W(\mathbb{F}_{p^s}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p) .$$

Let R be a discrete valuation ring of characteristic p , with uniformizer π , residue field k and fraction field K . Let X be a p -divisible group over R . After replacing X_K by an isogenous p -divisible group Y over K we have integers $r_i \geq 0, s > 0$ and a filtration $(0) = Y_0 \subset Y_1 \subset \dots \subset Y_r = Y$ as in Theorem 2.4. Applying the preceding considerations to $S = \text{Spec } K$ and Y_i/Y_{i-1} we therefore obtain p -adic Galois representations

$$(6) \quad \varrho_i = \varrho_{Y_i/Y_{i-1}} : \text{Gal}(\bar{K}/K) \longrightarrow GL_{h_i}(W(\mathbb{F}_{p^s}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p) ,$$

where $h_i = \text{height}(Y_i/Y_{i-1})$. The proof of Theorem 2.2 is based on the following lemma.

LEMMA 2.6. — With the previous notation, the following conditions are equivalent.

- (i) $\wedge^{h_i} \varrho_i$ is unramified, $\forall i = 1, \dots, r$.
- (ii) The Newton vector of X is constant.

Under these conditions the representations ϱ_i are also unramified.

In the proof of this lemma, again, as in the proof of Theorem 2.5, de Jong's theorem on extension of homomorphisms of p -divisible groups (or rather the techniques entering into the proof) plays a key role. \triangleright

Proof of Theorem 2.2. — We limit ourselves to the case where (\mathcal{E}, F) comes from a p -divisible group X on S . An easy reduction allows us to assume that $S = \text{Spec } A$, where A is a complete normal noetherian local ring of dimension 2 and where $U = S \setminus \{s\}$, with s denoting the special point. In proving Theorem 2.2, we may replace A by an A -algebra A' of the same kind such that the special point s' of $\text{Spec } A'$ is the unique point mapping

to s . Since the Newton vector of X is constant on the regular scheme U , we obtain via Theorem 2.4 p -adic Galois representations,

$$(7) \quad \varrho_i : \pi_1(U) \longrightarrow GL_{h_i}(W(\mathbf{F}_{p^s}) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p) \quad ,$$

$i = 1, \dots, r$. The determinant representation of each ϱ_i is a character of $\pi_1(U)$ with values in $(W(\mathbf{F}_{p^s}) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p)^\times$. Since $(W(\mathbf{F}_{p^s}) \otimes_{\mathbf{Z}_p} \mathbf{Q}_p)^\times$ contains an open subgroup of the form \mathbf{Z}_p^s , we may assume by the initial remark that $\wedge^{h_i} \varrho_i$ is an s -tuple of homomorphisms from $\pi_1(U)$ to \mathbf{Z}_p . Let $\pi : \tilde{S} \rightarrow S$ be a resolution of singularities. Then by Theorem 2.5 this s -tuple of homomorphisms factors through $\pi_1(\tilde{S})$. Let E be an irreducible component of the exceptional fiber $\pi^{-1}(s)$. Applying Lemma 2.6 to the discrete valuation ring $\mathcal{O}_{\tilde{S}, E}$, the pullback of X to $\text{Spec } \mathcal{O}_{\tilde{S}, E}$ has constant Newton vector, as had to be shown. \triangleright

3. THE SIEGEL MODULI SPACE

We fix a positive integer g . For an auxiliary integer $m \geq 3$ prime to p , we denote by $\mathcal{M} = \mathcal{M}_g = \mathcal{M}_{g,m}$ the *Siegel moduli space of genus g* over $\text{Spec } \mathbf{F}_p$. It represents the functor which to a locally noetherian scheme S in characteristic p associates the set of isomorphism classes of triples (A, λ, η) , where A is an abelian scheme of relative dimension g over S , and $\lambda : A \rightarrow \hat{A}$ is a principal polarization, and η is a (full) level- m -structure on (A, λ) .

The universal abelian scheme over \mathcal{M} defines a p -divisible group X on \mathcal{M} . The existence of a polarization implies that the Newton vectors of the fibers of X lie in the subset $(\mathbf{Q}^{2g})_+^1$ of $(\mathbf{Q}^{2g})_+$,

$$(\mathbf{Q}^{2g})_+^1 = \{(\nu_1, \dots, \nu_{2g}) \in (\mathbf{Q}^{2g})_+; \nu_i + \nu_{2g-i+1} = 1, \forall i = 1, \dots, g, \\ 0 \leq \nu_i \leq 1, \forall i = 1, \dots, 2g\}.$$

Let B_g be the set of elements in $(\mathbf{Q}^{2g})_+^1$ which satisfy the integrality condition in Cor. 1.3. Then B_g is a finite partially ordered set (*poset*), which has a unique maximal element and a unique minimal element,

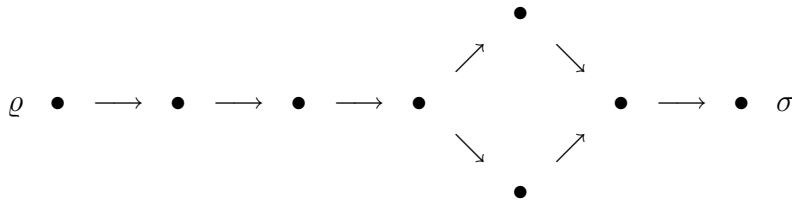
$$\begin{aligned} \varrho &= (1^g, 0^g) \text{ the ordinary Newton vector (maximal)} \\ \sigma &= ((\tfrac{1}{2})^{2g}) \text{ the supersingular Newton vector (minimal)}. \end{aligned}$$

For $\nu \in B_g$ we denote by \mathcal{S}_ν the corresponding Newton stratum in \mathcal{M} . By Grothendieck's semi-continuity theorem, Theorem 1.5, we have

$$(8) \quad \bar{\mathcal{S}}_\nu \subset \bigcup_{\nu' \leq \nu} \mathcal{S}_{\nu'} \quad .$$

In particular \mathcal{S}_σ is a closed subset and \mathcal{S}_ϱ is an open subset.

PICTURE OF B_4 : —



For $\nu \in B_g$, with $\nu = (\nu_1, \dots, \nu_{2g})$, let

$$(9) \quad \Delta(\nu) = \{(i, j) \in \mathbb{Z}^2; 0 \leq i \leq g, \sum_{\ell=1}^i \nu_{2g-\ell+1} \leq j < i\}$$

$$(10) \quad d(\nu) = \#\Delta(\nu)$$

EXAMPLE 3.1. —

$$d(\rho) = \frac{g(g+1)}{2}$$

$$d(\sigma) = \left\lfloor \frac{g^2}{4} \right\rfloor .$$

LEMMA 3.2. — Let $\nu, \nu' \in B_g$. Then

(i) $\nu \leq \nu'$ if and only if $\Delta(\nu) \subseteq \Delta(\nu')$.

(ii) If $\nu \leq \nu'$, then any shortest chain in the poset B_g starting at ν and ending at ν' has length $d(\nu') - d(\nu)$. In particular, B_g is a catenary poset. \triangleright

THEOREM 3.3 (Oort [O2]). — *Each Newton stratum \mathcal{S}_ν is equidimensional of dimension $d(\nu)$. The Newton stratification of \mathcal{M} has the strong stratification property,*

$$\bar{\mathcal{S}}_\nu = \bigcup_{\nu' \leq \nu} \mathcal{S}_{\nu'} .$$

REMARK 3.4. — The analogous statement for moduli spaces of vector bundles on a Riemann surface is false. More precisely, consider the stack $\mathcal{M} = \mathcal{M}_X(r, d)$ of holomorphic vector bundles of rank r and degree d over a Riemann surface X . Then \mathcal{M} is the disjoint union of its Harder-Narasimhan strata \mathcal{M}_ν , and we have

$$\bar{\mathcal{M}}_\nu \subset \bigcup_{\nu' \geq \nu} \mathcal{M}_{\nu'} ,$$

cf. Remark 1.6. If $g(X) = 0$ we have equality here and the same holds for $g(X) = 1$, according to a recent paper of Friedman and Morgan [FM]. This fails for $g(X) \geq 2$ [FM]. It is conceivable that

$$\nu' \geq \nu \iff \bar{\mathcal{M}}_\nu \cap \mathcal{M}_{\nu'} \neq \emptyset$$

and in loc. cit. this is proved, provided ν' and ν are adjacent. \triangleright

The proof of Theorem 3.3 is based on the following theorem.

THEOREM 3.5 (Oort [O2]). — *Let (X_0, λ_0) be a p -divisible group with principal polarization of dimension g and height $2g$ over an algebraically closed field k of characteristic p . Let $\nu_0 \in B_g$ be the Newton vector of X_0 , and let $\nu \in B_g$ with $\nu_0 \leq \nu$. Then there exists a principally polarized p -divisible group (X, λ) over $k[[t]]$ with special fiber (X_0, λ_0) and with Newton vector of the generic fiber of X equal to ν .*

Theorem 3.5 implies via the Serre-Tate theorem the second statement in Theorem 3.3. Since, as is easily seen, $\mathcal{S}_\sigma \neq \emptyset$ (the supersingular locus), Theorem 3.5 implies $\mathcal{S}_\nu \neq \emptyset$ for all $\nu \in B_g$. The first statement in Theorem 3.3 now follows from Lemma 3.2 and the purity theorem. \triangleright

The two extreme cases in the Newton stratification deserve a separate discussion.

THE SUPERSINGULAR STRATUM \mathcal{S}_σ : The fact that \mathcal{S}_σ is equidimensional of dimension $d(\sigma) = \left\lfloor \frac{g^2}{4} \right\rfloor$ (a special case of Theorem 3.3) was proved earlier by Li and Oort [LO], among other things. The supersingular Newton stratum is exceptional in several aspects. Every abelian variety occurring as a fiber of the universal abelian scheme at a point of \mathcal{S}_σ is isogenous to E^g , where E is a supersingular elliptic curve. Moreover, let $x \in \mathcal{S}_\sigma$ and let (A, λ) be the fiber of x of the universal object on \mathcal{M} . Then one can represent (A, λ) in an almost canonical way as the quotient of E^g by a finite group scheme. This leads to the dimension formula for this stratum and in fact much more. There is the hope for an *explicit synthetic* description of \mathcal{S}_σ , like the one of Kaiser [K] for $g = 2$ (compare also [KR], and [R] for $g = 3$). In [LO] the number of irreducible components of \mathcal{S}_σ is given.

THE ORDINARY STRATUM \mathcal{S}_ρ : In contrast to the supersingular stratum, the ordinary stratum is quite amorphous and nonlinear, and there is no hope of an explicit description of it. According to Theorem 3.3, \mathcal{S}_ρ is open and dense in \mathcal{M} . This fact has been known for a long time, by more direct and easier proofs: 1) There is the proof by Mumford [M], and Norman and Oort [NoO] (compare also Chai and Faltings [FC]) using Cartier theory to construct deformations. 2) There is the proof by Koblitz [Kob], compare also [Ill], App. 2, who investigated by deformation-theoretic arguments the stratification of \mathcal{M} by the p -rank of X . 3) There is the global proof using toroidal compactifications [FC]. 4) There is the proof of Ngo and Genestier [NG] who deduce the density result from a corresponding density result (of a combinatorial nature) in *bad reduction*. Furthermore, Chai [C1] has proved the much stronger assertion that the orbit of an arbitrary ordinary point under the Hecke correspondences of degree prime to p is dense in \mathcal{M} . \triangleright

The proof of Theorem 3.5 is rather round-about. For a p -divisible group X over an algebraically closed field k , let

$$(11) \quad a(X) = \text{Hom}_k(\alpha_p, X) = \dim_k M / (F(M) + V(M)) \quad .$$

Here (M, F) is the Dieudonné module of X and $V = pF^{-1}$. Hence $a(X) = 0$ if and only if $X \simeq \hat{\mathbf{G}}_m^d \times (\mathbf{Q}_p/\mathbf{Z}_p)^{d'}$. In a first step one proves Theorem 3.5 under the additional assumption $a(X_0) = 1$, cf. Oort [O1] (the case where $a(X_0) = 0$, where $\nu_0 = \varrho$, is trivial). The technical tool for this is the theory of displays, compare Section 4, which allows one to write down explicitly by display equations deformations of a (polarized) formal p -divisible group. The hypothesis $a(X_0) = 1$ is then needed to read off from these display equations the Newton vectors of the p -divisible groups occurring in the deformation.

In a second step, one shows that (X_0, λ_0) can be deformed into a principally polarized p -divisible group (X, λ) with the same Newton vector and with $a(X) = 1$. This in turn is reduced to the following statement which is of independent interest.

THEOREM 3.6 (de Jong, Oort [JO]). — *Let X_0 be a p -divisible group over an algebraically closed field k of characteristic p such that its F -isocrystal is irreducible. Then there exists an irreducible scheme T over k and a p -divisible group X over T together with an isogeny $X_0 \times_{\mathrm{Spec} k} T \rightarrow X$ over T , such that any p -divisible group over k isogenous to X_0 occurs as a fiber of X at a k -rational point of T .*

An equivalent formulation of the previous theorem is the following. Let X_0 be as in the previous theorem. By [RZ] the following functor on $(\mathrm{Sch}/\mathrm{Spec} k)$ is representable by a formal scheme \mathcal{S} locally formally of finite type over $\mathrm{Spec} k$,

$$S \longmapsto \left\{ \begin{array}{l} \text{isomorphism classes of pairs } (X, \varrho), \text{ where } X \text{ is a } p\text{-divisible group} \\ \text{over } S \text{ and } \varrho : X_0 \times_{\mathrm{Spec} k} S \longrightarrow X \text{ is a quasi-isogeny of height } 0 \end{array} \right\} .$$

Then \mathcal{S} is irreducible. ▷

COROLLARY 3.7. — *Let X_0 be as in the previous theorem. Then there exists a deformation of X_0 into a p -divisible group X isogenous to X_0 with $a(X) = 1$.*

Indeed, the locus in T where $a(X) = 1$ is open. It therefore suffices to produce one point $t \in T(k)$ where $a(X_t) = 1$. This is easy. ▷

CONJECTURE 3.8. — *Let $\nu, \nu' \in B_g$ with $\nu' \leq \nu$. The closure of each irreducible component of \mathcal{S}_ν meets $\mathcal{S}_{\nu'}$.*

Conjecture 3.8 would certainly hold if Oort's conjecture [O2] was true, according to which for $\nu \neq \sigma$ the intersection of \mathcal{S}_ν with any connected component of \mathcal{M} is irreducible.

4. DISPLAYS

Display equations for formal p -divisible groups were introduced by Mumford [M]. These techniques were applied to moduli problems of abelian varieties by Norman [No] and Norman and Oort [NoO]. We will follow here the recent formulation of the theory due to Zink [Z1].

Let R be a ring of characteristic p . We denote by $W(R)$ its ring of Witt vectors and by $x \mapsto {}^F x$ resp. $x \mapsto {}^V x$ its Frobenius resp. Verschiebung endomorphisms. Let $I_R \subset W(R)$ be the ideal of Witt vectors with trivial 0-component.

DEFINITION 4.1. — *A not necessarily nilpotent display (= 3n-display) over R is a quadruple (P, Q, F, V^{-1}) consisting of a finitely generated projective $W(R)$ -module P , a submodule $Q \subset P$ and F -linear maps $F : P \rightarrow P$ and $V^{-1} : Q \rightarrow P$. The following conditions are required:*

- (i) $I_R P \subset Q \subset P$ and the quotient P/Q is a projective R -module.
- (ii) $V^{-1} : Q \rightarrow P$ is a F -linear epimorphism.
- (iii) For $x \in P$ and $w \in W(R)$ we have ${}^V w \cdot x \in Q$ and we require that

$$V^{-1}({}^V w \cdot x) = w \cdot F(x) \quad .$$

We note that F is determined by the remaining data. There is no operator V . The reason for the notation comes from the following example.

EXAMPLE 4.2. — Let $R = k$ be a perfect field. Then an F -crystal (M, F) over k such that $pM \subset FM$ defines a 3n-display (M, VM, F, V^{-1}) . Here as usual $V = pF^{-1}$. This defines an equivalence of categories.

The notion of a display is obtained by imposing a nilpotency condition as follows. After localization in R there exists a $W(R)$ -basis e_1, \dots, e_n of P such that

$$Q = I_R e_1 \oplus \dots \oplus I_R e_d \oplus W(R)e_{d+1} \oplus \dots \oplus W(R)e_n \quad ,$$

for some d with $0 \leq d \leq n$. Then there exists an invertible matrix $(\alpha_{ij}) \in GL_n(W(R))$ such that

$$\begin{aligned} F e_j &= \sum_{i=1}^n \alpha_{ij} e_i \quad \text{for } j = 1, \dots, d \\ V^{-1} e_j &= \sum_{i=1}^n \alpha_{ij} e_i \quad \text{for } j = d + 1, \dots, n \quad . \end{aligned}$$

Conversely, any $(\alpha_{ij}) \in GL_n(W(R))$ defines a 3n-display. Let $(\beta_{k\ell})$ be the inverse of (α_{ij}) . Let $B \in M_{n-d}(R)$ be the image of $(\beta_{k\ell})_{k,\ell=d+1,\dots,n}$ under the 0-component map

$$M_{n-d}(W(R)) \longrightarrow M_{n-d}(R) \quad .$$

Let $B^{(p)}$ be the matrix obtained from B by raising its coefficients to the power p . The nilpotency condition can now be formulated: there exists N such that

$$B^{(p^N)} \cdot \dots \cdot B^{(p)} \cdot B = 0 \quad .$$

In the context of Example 4.2 the F -crystal (M, F) defines a display if and only if $pM \subset FM$ and if $V = pF^{-1}$ is topologically nilpotent on M .

THEOREM 4.3 (Zink [Z1]). — *We assume that the nilideal of R is nilpotent. Then there is a fully faithful functor BT from the category of displays over R to the category of formal*

p-divisible groups over R . This is an equivalence of categories if either R is an excellent local ring or an algebra of finite type over a field k .

It is quite likely that this equivalence of categories holds for any noetherian ring R of characteristic p . The functor BT has the following properties: 1) It commutes with arbitrary base change. 2) $\text{Lie } BT(P, Q, F, V^{-1}) = P/Q$. 3) P can be identified with the value at $W(R)$ of the crystal defined by the universal extension of $BT(P, Q, F, V^{-1})$, cf. Example 1.4. 4) The passage from a formal p -divisible group to its dual p -divisible group can be expressed in terms of displays, provided that the dual p -divisible group is a formal group, i.e. has trivial étale part.

The theory also works if p is only supposed to be nilpotent in R . For an extension of the theory to p -divisible groups with an étale part, compare [Z2].

5. OTHER MODULI SPACES OF ABELIAN VARIETIES

Let F be a finite-dimensional semisimple \mathbf{Q} -algebra equipped with a positive involution $*$ and let V be a finite F -module equipped with an alternating non-degenerate \mathbf{Q} -valued skew-hermitian pairing $\langle \ , \ \rangle$. The F -linear similitudes of $(V, \langle \ , \ \rangle)$ form an algebraic group G over \mathbf{Q} . We assume that G is a *connected* reductive algebraic group. We also fix a conjugacy class of algebraic homomorphisms $h : \mathbf{C}^\times \rightarrow G(\mathbf{R})$ satisfying the usual Riemann conditions. Let E be the corresponding Shimura field, i.e. the field of definition of the corresponding conjugacy class of cocharacters $\mu : \mathbf{G}_{m, \mathbf{C}} \rightarrow G_{\mathbf{C}}$. We assume that p is a prime of good reduction, in particular $G_{\mathbf{Q}_p}$ is unramified, and we choose a hyperspecial maximal compact subgroup K_p of $G(\mathbf{Q}_p)$. We choose a prime ideal of E over p with residue field κ . After a choice of some sufficiently small open compact subgroup $K^p \subset G(\mathbf{A}_f^p)$, Kottwitz [Kot1] has defined a moduli problem of abelian varieties which is representable by a smooth quasi-projective scheme $\mathcal{M} = \mathcal{M}(G, h)_K = \mathcal{M}(F, V, \langle \ , \ \rangle, h, K^p.K_p)$ over $\text{Spec } \kappa$.

Let L be the fraction field of $W(\overline{\mathbf{F}}_p)$ and let $B(G)$ be the set of σ -conjugacy classes in $G(L)$. By associating to a point $s \in \mathcal{M}$ the F -isocrystal with G -structure defined by the fiber at (a geometric point over) s of the universal abelian scheme over \mathcal{M} with its auxiliary structure (endomorphisms and polarization), we obtain a map

$$\mathcal{M} \longrightarrow B(G) \ .$$

The conjugacy class μ defines a finite subset $B(G, \mu)$ of $B(G)$ ([Kot2], §6). It is defined by the group-theoretic version of Mazur's inequality, compare the proof of Corollary 1.7. The image of the map above is contained in $B(G, \mu)$ [RR]. (In the case of the Siegel moduli space \mathcal{M}_g we have $G = GSp_{2g}$ and $B(G, \mu) = B_g$, cf. section 3.) Furthermore, $B(G, \mu)$ is partially ordered and the semicontinuity theorem 1.5 continues to hold in this

context [RR]. We therefore obtain the *generalized Newton stratification* of \mathcal{M} (by the locally closed subsets arising as inverse images of elements of $B(G, \mu)$),

$$\mathcal{M} = \bigcup_{b \in B(G, \mu)} \mathcal{M}_b .$$

Just as B_g , also $B(G, \mu)$ is a catenary poset [C2] with a unique minimal element b_0 (the μ -basic element) and a unique maximal element b_1 (the μ -ordinary element).

THEOREM 5.1 (Wedhorn[W]). — *The μ -ordinary locus \mathcal{M}_{b_1} is open and dense in \mathcal{M} . \triangleright*

This is about the only known general statement in direction of the following conjecture.

CONJECTURE 5.2. — (i) *The generalized Newton stratification of $\mathcal{M} = \mathcal{M}(G, h)_K$ has the strong stratification property.*

(ii) *The generalized Newton stratum corresponding to $b \in B(G, \mu)$ is equidimensional of dimension $d(b) = \dim \mathcal{M} - c(b)$, where $c(b)$ is the length of a chain joining b to b_1 .*

(iii) *Let $b, b' \in B(G, \mu)$ with $b' \leq b$. The closure of each irreducible component of \mathcal{M}_b meets $\mathcal{M}_{b'}$.*

We note that Chai [C2] has given a group theoretical formula for $d(b)$. When G is a group of unitary similitudes, there are results supporting (i) and (ii) of this conjecture:

THEOREM 5.3 (Oort). — *Let F be an imaginary quadratic field such that p splits in F . Then (i) and (ii) of Conjecture 5.2 hold true for $\mathcal{M} = \mathcal{M}(F, V, \langle \ , \ \rangle, h, K^p.K_p)$.*

The proof is analogous to the proof of Theorem 3.3 (which proves (i) and (ii) of Conjecture 5.2 for the Siegel moduli space). The analogue of Theorem 3.5 is the following statement which confirms a conjecture of Grothendieck [G]. Its proof is similar to that of Theorem 3.5, but simpler.

THEOREM 5.4 (Oort[O2]). — *Let X_0 be a p -divisible group of height n and dimension d over an algebraically closed field k of characteristic p , with Newton vector $\nu_0 \in (\mathbb{Q}^n)_+$. Let $\nu \in (\mathbb{Q}^n)_+$ such that ν satisfies the integrality condition of Corollary 1.3 and with $\nu_0 \leq \nu \leq (1^d, 0^{n-d})$. Then there exists a p -divisible group X over $k[[t]]$ with special fiber X_0 and with Newton vector of the generic fiber of X equal to ν . \triangleright*

Using the Serre-Tate theorem, Theorem 5.4 implies that property (i) holds in Theorem 5.3. Using Honda-Tate theory one shows that the μ -basic locus of \mathcal{M} is non-empty, compare [Z4]. Therefore as in the proof of Theorem 3.3, Theorem 5.4 implies that \mathcal{M}_b is non-empty for all $b \in B(G, \mu)$, and the purity theorem allows one now to deduce also property (ii) in Theorem 5.3 from Theorem 5.4. \triangleright

We mention that when F is an imaginary quadratic field such that $p \neq 2$ is inert, Bültel and Wedhorn [BW] have proved (i) and (ii) of Conjecture 5.2, provided that the signature of the skew-hermitian form $\langle \ , \ \rangle$ on V is of the form $(n - 1, 1)$. On the other hand, the conjecture seems to be open even for such classical moduli spaces as the Hilbert-Blumenthal varieties.

REFERENCES

- [AB] M. F. ATIYAH, R. BOTT – *The Yang-Mills equations over Riemann surfaces. Philos. Trans. Roy. Soc. London Ser. A* 308 (1983), 523–615.
- [BW] O. BÜLTEL, T. WEDHORN – *Congruence relations for Shimura varieties associated to some unitary groups. math.AG/0202026.*
- [C1] C. L. CHAI – *Every ordinary symplectic isogeny class in positive characteristic is dense in the moduli. Invent. Math.* 121 (1995), 439–479.
- [C2] C. L. CHAI – *Newton polygons as lattice points. Amer. J. Math.* 122 (2000), 967–990.
- [D] M. DEMAZURE – *Lectures on p -divisible groups.* Reprint of the 1972 original. *Lecture Notes in Mathematics* 302. Springer-Verlag, Berlin, 1986.
- [FC] G. FALTINGS, C. L. CHAI – *Degeneration of abelian varieties.* With an appendix by David Mumford. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*, 22. Springer-Verlag, Berlin, 1990. xii+316 pp.
- [FM] R. FRIEDMAN, J. W. MORGAN – *On the converse to a theorem of Atiyah and Bott. math.AG/0006086*
- [G] A. GROTHENDIECK – *Groupes de Barsotti-Tate et cristaux de Dieudonné. Sémin. Math. Sup.* 45, Presses de l’Univ. de Montreal, 1970.
- [Ill] L. ILLUSIE – *Déformations de Groupes de Barsotti-Tate*, in: Séminaire sur les pinceaux arithmétiques : la conjecture de Mordell, *Astérisque* 127 (1985), 151-198.
- [J1] A. J. DE JONG – *Barsotti-Tate groups and crystals. Proceedings of the International Congress of Mathematicians*, Vol. II (Berlin, 1998). Doc. Math. 1998, Extra Vol. II, 259–265
- [J2] A. J. DE JONG – *Homomorphisms of Barsotti-Tate groups and crystals in positive characteristic. Invent. Math.* 134 (1998), 301–333.
- [JO] A. J. DE JONG, F. OORT – *Purity of the stratification by Newton polygons. J. Amer. Math. Soc.* 13 (2000), 209–241.
- [K] C. KAISER – *Ein getwistetes fundamentales Lemma für die GSp_4 . Bonner Mathematische Schriften* 303. Universität Bonn, Mathematisches Institut, Bonn, 1997, 71 pp.
- [Ka] N. KATZ – *Slope filtration of F -crystals. Journées de géométrie algébrique de Rennes* (Rennes, 1978), Vol. I, *Astérisque* 63, Soc. Math. France, Paris (1979), 113–163.
- [Kob] N. KOBLITZ – *p -adic variation of the zeta function over families of varieties defined over finite fields. Compositio Math.* 31 (1975), 119–218.
- [Kot1] R. KOTTWITZ – *Points on some Shimura varieties over finite fields. J. Amer. Math. Soc.* 5 (1992), 373–444.

- [Kot2] R. KOTTWITZ – *Isocrystals with additional structure. II. Compositio Math.* 109 (1997), 255–339.
- [KR] S. KUDLA, M. RAPOPORT – *Cycles on Siegel threefolds and derivatives of Eisenstein series. Ann. Sci. École Norm. Sup. (4)* 33 (2000), 695–756.
- [LO] K. Z. LI, F. OORT – *Moduli of supersingular abelian varieties. Lecture Notes in Mathematics*, 1680. Springer-Verlag, Berlin, 1998. iv+116 pp.
- [Me] W. MESSING – *The crystals associated to Barsotti-Tate groups: with applications to abelian schemes. Lecture Notes in Mathematics*, Vol. 264. Springer-Verlag, Berlin-New York, 1972. iii+190 pp.
- [M] D. MUMFORD – *Bi-extensions of formal groups. Algebraic Geometry, Bombay Colloquium 1968*, 307–322, Tata Inst. Fund. Research and Oxford Univ. Press, 1969.
- [NG] B. C. NGO, A. GENESTIER – *Alcoves et p -rang des variétés abéliennes.* math.AG/0107223 (2001).
- [No] P. NORMAN – *An algorithm for computing local moduli of abelian varieties. Ann. of Math. (2)* 101 (1975), 499–509.
- [NoO] P. NORMAN, F. OORT – *Moduli of abelian varieties. Ann. of Math. (2)* 112 (1980), 413–439.
- [O1] F. OORT – *Newton polygons and formal groups: conjectures by Manin and Grothendieck. Ann. of Math. (2)* 152 (2000), 183–206.
- [O2] F. OORT – *Newton polygon strata in the moduli space of abelian varieties. Moduli of abelian varieties (Texel Island, 1999)*, 417–440, Progr. Math., 195, Birkhäuser, Basel, 2001.
- [RR] M. RAPOPORT, M. RICHARTZ – *On the classification and specialization of F -isocrystals with additional structure. Compositio Math.* 103 (1996), 153–181.
- [RZ] M. RAPOPORT, T. ZINK – *Period spaces for p -divisible groups. Annals of Mathematics Studies* 141. Princeton University Press, Princeton, NJ, 1996. xxii+324 pp.
- [R] M. RICHARTZ – *Klassifikation von selbstdualen Dieudonné-Gittern in einem dreidimensionalen polarisierten supersingulären Isokristall*, Diss. Universität Bonn, 1998.
- [W] T. WEDHORN – *Ordinariness in good reductions of Shimura varieties of PEL-type. Ann. Sci. Ecole Norm. Sup. (4)* 32 (1999), 575–618.
- [Z1] T. ZINK – *The display of a formal p -divisible group.* Preprint 98-017 SFB 343, Universität Bielefeld 1998, to appear in Astérisque.
- [Z2] T. ZINK – *A Dieudonné theory for p -divisible groups. Advanced Studies in Pure Mathematics* 30 (2001), 139–160.
- [Z3] T. ZINK – *On the slope filtration. Duke Math. J.* 109 (2001), 79–95.

- [Z4] T. ZINK – *Isogenieklassen von Punkten von Shimuramannigfaltigkeiten mit Werten in einem endlichen Körper*. *Math. Nachr.* 112 (1983), 103–124.

Michael RAPOPORT

Mathematisches Institut
der Universität zu Köln

Weyertal 86-90

D-50931 Köln

(Germany)

E-mail : `rapoport@mi.uni-koeln.de`

CATALAN'S CONJECTURE
[after Mihăilescu]

by **Yuri F. BILU**

TABLE OF CONTENTS

1. Introduction	1
2. Cassels' relations and lower estimates for $ x $ and $ y $	3
3. Algebraic criteria	5
4. Logarithmic forms, Tijdeman's argument and the relation $p \not\equiv 1 \pmod{q}$	6
5. Generalities	11
6. Overview of the proof	14
7. Proof of Theorem 6.3.2	17
8. Proof of Theorem 6.3.3	19
9. Proof of Theorem 6.3.4	23
REFERENCES	24

To E. W.

1. INTRODUCTION

In 1844 Crelle's journal published the following note [9].

Note

extraite d'une lettre adressée à l'éditeur par Mr. *E. Catalan*, Répétiteur à l'école polytechnique de Paris.

Je vous prie, Monsieur, de vouloir bien énoncer, dans votre recueil, le théorème suivant, que je crois vrai, bien que je n'aie pas encore réussi à le démontrer complètement : d'autres seront peut-être plus heureux :

Deux nombres entiers consécutifs, autres que 8 et 9, ne peuvent être des puissances exactes ; autrement dit : l'équation $x^m - y^n = 1$, dans laquelle les inconnues sont entières et positives, n'admet qu'une seule solution.

Thus, we have the following conjecture.

CONJECTURE 1.1 (Catalan). — Equation $x^u - y^v = 1$ has no solutions in integers $x, y, u, v > 1$ other than $3^2 - 2^3 = 1$.

Now, 158 years after, the conjecture is completely proved. Let us briefly review the most important events which lead to the solution of this celebrated problem. This is **not** a comprehensive historical account of Catalan's problem; the latter can be found in Ribenboim's book [29] and Mignotte's survey [22].

Seven years after Catalan's note appeared, Lebesgue [17] proved that equation $x^m - y^2 = 1$ has no solutions in positive integers x, y, m with $m > 1$. In 1965 Ko Chao [14] showed that equation $x^2 - y^n = 1$ has no solutions in positive integers x, y, n with $n > 1$ other than $3^2 - 2^3 = 1$. These two results reduce Catalan's conjecture to the following assertion.

CONJECTURE 1.2. — Equation

$$(1) \quad x^p - y^q = 1$$

has no solutions in non-zero integers x, y and odd primes p, q .

Notice that we no longer assume x and y positive. It is convenient, because now the problem is symmetric: if (x, y, p, q) is a solution, then so is $(-y, -x, q, p)$. This will be repeatedly used in the sequel.

From now on Conjecture 1.2 will be referred to as *Catalan's conjecture* and (1) as Catalan's equation.

Cassels [8] discovered important arithmetical properties of solutions of Catalan's equation. His results (see Proposition 2.1) are indispensable in most of the subsequent works on Catalan's equation.

In 1976 Tijdeman [32] made a breakthrough. Using Baker's theory, he proved that the exponents p and q are bounded by an explicit absolute constant. Together with the classical result of Baker [3] this implies that $|x|$ and $|y|$ are bounded by an explicit absolute constant as well, and Catalan's problem is thereby decidable.

In a different direction, Inkeri [12, 13] and others obtained algebraic criteria of solubility of (1) in terms of the exponents p and q . In nineties, Mignotte and Roy used Inkeri-type criteria, Tijdeman's argument and electronic computations to obtain tight lower and upper bounds for p and q . (Upper bounds were also obtained by Blass *et al.* [6] and O'Neil [27].) By 2000, it was proved that p and q lie between 10^7 and 10^{18} . See [25] for more precise results and a survey of this period.

In 1999 Preda Mihăilescu enters the scene. In his first paper [25] he drastically refined Inkeri's criterion. And quite recently, after several unsuccessful attempts, he finally settled [26] Catalan's conjecture:

THEOREM 1.3 (Mihăilescu). — Conjecture 1.2 is true.

The present paper contains a reasonably self-contained proof of this result.

Plan of the paper : In Section 2 we recall Cassels' relations and derive their immediate consequence, in particular, Hyrö's lower bounds for $|x|$ and $|y|$. In Section 3 we very

briefly review algebraic criteria for Catalan's equation in terms of p and q , and prove Mihăilescu's "double Wieferich" criterion. In Section 4 we use binary logarithmic forms, Tijdeman's argument, and computations by Mignotte and Roy to show that $p \not\equiv 1 \pmod{q}$. Section 5 contains general lemmas. In Section 6 Theorem 1.3 is reduced to three more technical statements, which are proved in the three final section.

Acknowledgements : My deepest gratitude goes to Hendrik W. Lenstra and Yann Bugeaud, who carefully read the manuscript and suggested numerous corrections and improvements. I am indebted to Yann Bugeaud, Andrew Glass, Guillaume Hanrot, Maurice Mignotte and Preda Mihăilescu for explaining to me various results from Sections 2 and 4 and other useful discussions. I also thank Bruno Anglès, John Coates, Gabi Hecke, Shanta Laishram, Hendrik W. Lenstra, Tauno Metsänkylä and Gisbert Wüstholz, who detected inaccuracies in previous versions of this note. Finally, I thank Denis Benois and Leonid Positselski for a tutorial in commutative algebra.

1.1. Notation

In the sequel we assume, unless the contrary is indicated explicitly, that x, y are non-zero integers and p, q are odd prime numbers satisfying

$$(2) \quad x^p - y^q = 1.$$

As we had already noticed, (2) implies that $(-y)^q - (-x)^p = 1$, and all the statements below remain true with x, y, p, q replaced by $-y, -x, q, p$.

We denote by ζ a primitive p -th root of unity and put

$$K = \mathbf{Q}(\zeta), \quad G = \text{Gal}(K/\mathbf{Q}).$$

The principal ideal $(1 - \zeta)$ will be denoted by \mathfrak{p} . Recall that it is a prime ideal of K and that $(p) = \mathfrak{p}^{p-1}$.

More specific notation will be introduced at the appropriate places.

2. CASSELS' RELATIONS AND LOWER ESTIMATES FOR $|x|$ AND $|y|$

Cassels [8] proved that $q|x$ and $p|y$. More precisely, he established the following relations.

PROPOSITION 2.1 (Cassels). — There exist a non-zero integer a and a positive integer v such that

$$(3) \quad x - 1 = p^{q-1}a^q, \quad y = pav,$$

$$(4) \quad \frac{x^p - 1}{x - 1} = pv^q,$$

and, symmetrically, there exist a non-zero integer b and a positive integer u such that

$$(5) \quad y + 1 = q^{p-1}b^p, \quad x = qub,$$

$$(6) \quad \frac{y^q + 1}{y + 1} = qu^p.$$

□

The following consequence is crucial.

COROLLARY 2.2. — The number $\lambda := (x - \zeta)/(1 - \zeta)$ is an algebraic integer. The principal ideal (λ) is a q -th power of an ideal of the field K .

Proof Since $p|(x - 1)$ by (3), the prime ideal $\mathfrak{p} = (1 - \zeta)$ divides $x - \zeta$, but \mathfrak{p}^2 does not. Hence λ is an algebraic integer, not divisible by \mathfrak{p} , and the same is true for its conjugates λ^σ , where $\sigma \in G$. Identity $(1 - \zeta^\sigma)\lambda^\sigma - (1 - \zeta^\tau)\lambda^\tau = \zeta^\tau - \zeta^\sigma$ implies that for distinct $\sigma, \tau \in G$, the greatest common divisor of λ^σ and λ^τ divides $(\zeta^\tau - \zeta^\sigma) = \mathfrak{p}$. Hence the numbers λ^σ are pairwise co-prime.

Now rewrite (4) as $\prod_{\sigma \in G} \lambda^\sigma = v^q$. Since the factors are pairwise co-prime, each principal ideal (λ^σ) is a q -th power of an ideal. □

Cassels' relations imply various lower estimates for the variables x and y in terms of p and q . For instance, (3) and (5) immediately yield

$$(7) \quad |x| \geq p^{q-1} - 1,$$

$$(8) \quad |y| \geq q^{p-1} - 1,$$

and this can be refined without much effort.

Hyyrö [11] obtained an estimate of a different kind: $|x| \geq q(2p + 1)(2q^{p-1} + 1)$ (and similarly for $|y|$). Since Hyyrö's paper is not easily available, I prove below a slightly weaker estimate, which is totally sufficient for our purposes. It is an easy consequence of the following proposition.

PROPOSITION 2.3. — If p does not divide $q - 1$ then $q^{p-2} | (u - 1)$.

Proof : Rewriting (6) as

$$((-y)^{q-1} - 1) + ((-y)^{q-2} - 1) + \cdots + (-y - 1) = q(u^p - 1),$$

we deduce that $(y + 1) | (q(u^p - 1))$. Now (5) implies that $u^p \equiv 1 \pmod{q^{p-2}}$. Since p does not divide the order $q^{p-3}(q - 1)$ of the multiplicative group $\text{mod } q^{p-2}$, this implies that $u \equiv 1 \pmod{q^{p-2}}$. □

COROLLARY 2.4. — We have $|x| \geq q^{p-1}$.

Proof : If $p|(q - 1)$ then $p < q$ and the result follows from (7). If p does not divide $q - 1$ then $q^{p-2} | (u - 1)$, and, since u is positive, this implies $u \geq q^{p-2} + 1$. Since $x = qub$, we have $|x| \geq qu \geq q^{p-1} + q$, better than wanted. □

REMARK 2.5. — This version of Hyyrö's argument is due to Mignotte and Bugeaud. It was kindly communicated to me by Yann Bugeaud. Using more advanced tools, Mihăilescu [26, Appendix A] obtained a much sharper estimate $|x| \geq (q^{2p-2}/2)^4$.

3. ALGEBRAIC CRITERIA

Using Cassels' relations and some algebraic number theory, one may get various algebraic criteria of solvability of Catalan's equation with given exponents p and q . The most famous criterion is due to Inkeri [12, 13]:

THEOREM 3.1 (Inkeri). — With the notation of Subsection 1.1, put $K_p = \mathbf{Q}(\sqrt{-p})$ if $p \equiv 3 \pmod{4}$ and $K_p = K$ if $p \equiv 1 \pmod{4}$. Then either $p^{q-1} \equiv 1 \pmod{q^2}$ or q divides the class number of the field K_p . \square

It will be explained in Subsection 4.4 how algebraic criteria of this kind, together with electronic computations, allow one to obtain lower bounds for p and q .

Refinements of and supplements for Inkeri's criterion were suggested by Mignotte [20], Schwarz [30] and others; see [22] for a survey of these results. I would especially mention the paper by Bugeaud and Hanrot [7], which strongly influenced Mihăilescu's work.

Verification of Inkeri's criterion for a given pair (p, q) requires computing certain class numbers, which seriously affects its computational efficiency. Mihăilescu [25] made a major step forward, showing that the class number condition can be omitted.

THEOREM 3.2 (Mihăilescu). — For any solution of (x, y, p, q) of (2) we have $q^2|x$ and

$$(9) \quad p^{q-1} \equiv 1 \pmod{q^2}.$$

Congruence (9) (called *Wieferich's relation*) will be used in Section 4 to prove that $p \not\equiv 1 \pmod{q}$. Relation $q^2|x$ is crucial in the proof of Theorem 6.3.2.

By symmetry, one has $q^{p-1} \equiv 1 \pmod{p^2}$. Pairs (p, q) , satisfying this and (9) are called *double Wieferich pairs*. Only six such pairs are currently known:

$$(2, 1093), (3, 1006003), (5, 1645333507), (83, 4871), (911, 318917), (2903, 18787).$$

I sketch the proof of Theorem 3.2, because it is very instructive and can serve as a good model of the much more involved proof of Theorem 1.3. See [19, 28] for different proofs.

3.1. Proof of Theorem 3.2

For $a \in \{1, 2, \dots, p-1\}$ let σ_a be the element of $G = \text{Gal}(K/\mathbf{Q})$ be defined by $\zeta \mapsto \zeta^a$. In the group ring $\mathbf{Z}[G]$ consider elements

$$\Theta_c = \sum_{a=1}^{p-1} [ac/p] \sigma_a^{-1} \quad (c = 1, 2, \dots, p-1).$$

In particular, $\Theta_1 = 0$ and $\Theta_2 = \sigma_{(p+1)/2} + \cdots + \sigma_{p-1}$. Ideal $\mathcal{I} = (\Theta_1, \Theta_2, \dots, \Theta_{p-1})$ of $\mathbf{z}[G]$ is called the *Stickelberger ideal*. Its main property is the *Stickelberger theorem*: any $\Theta \in \mathcal{I}$ annihilates the class group of K . That is, for any ideal \mathfrak{a} of K and any $\Theta \in \mathcal{I}$, the ideal \mathfrak{a}^Θ is principal. See [34, Section 6.2] for the details.

Let $\iota = \sigma_{p-1}$ be the complex conjugation. Mihăilescu proves the following assertion.

PROPOSITION 3.1.1. — For any $\Theta \in (1 - \iota)\mathcal{I}$, the element $(x - \zeta)^\Theta$ is a q -th power in K .

Proof : Write $\Theta = (1 - \iota)\Theta'$, where $\Theta' \in \mathcal{I}$. Put $\lambda := (x - \zeta)/(1 - \zeta)$. By Corollary 2.2 the principal ideal (λ) is a q -th power: $(\lambda) = \mathfrak{a}^q$. By the Stickelberger theorem $\mathfrak{a}^{\Theta'}$ is a principal ideal, say, (α) . It follows that $(\lambda^{\Theta'}) = (\alpha)^q$, or $\lambda^{\Theta'} = \eta\alpha^q$, where η is a unit of K . We obtain

$$(10) \quad (x - \zeta)^\Theta = \left(\frac{1 - \zeta}{1 - \bar{\zeta}} \right)^{\Theta'} \frac{\eta}{\bar{\eta}} \left(\frac{\alpha}{\bar{\alpha}} \right)^q.$$

Since η is a unit, $\eta/\bar{\eta}$ is a root of unity¹. The quotient $(1 - \zeta)/(1 - \bar{\zeta})$ is a root of unity as well. Thus, $(x - \zeta)^\Theta$ is a q -th power times a root of unity. Since any root of unity in K is a q -th power, so is $(x - \zeta)^\Theta$. \square

Proof of $q^2|x$: Since $(1 - \zeta x)^\Theta$ is equal to $(x - \zeta)^\Theta$ times a root of unity, it is a q -th power as well. On the other hand, $q|x$ implies that $(1 - \zeta x)^\Theta \equiv 1 \pmod{q}$. Since q is unramified in K , this implies that $(1 - \zeta x)^\Theta \equiv 1 \pmod{q^2}$ (cf. Proposition 5.3.1 below).

However, if $\Theta = \sum_{\sigma \in G} n_\sigma \sigma$, then a quick calculation shows that

$$(1 - \zeta x)^\Theta \equiv 1 - x \sum_{\sigma \in G} n_\sigma \zeta^\sigma \pmod{q^2}.$$

It follows that either $q^2|x$ or $q|\sum_{\sigma \in G} n_\sigma \zeta^\sigma$. In the latter case $q|n_\sigma$ for any $\sigma \in G$. However, this is not true if, for instance,

$$\Theta = (1 - \iota)\Theta_2 = -\sigma_1^{-1} - \cdots - \sigma_{(p-1)/2}^{-1} + \sigma_{(p+1)/2}^{-1} + \cdots + \sigma_{p-1}^{-1}.$$

Thus, $q^2|x$. \square

Proof of (9) : This is just an elementary exercise. Since $q^2|x$, the first equality in (3) implies that

$$(11) \quad p^{q-1}a^q \equiv -1 \pmod{q^2}.$$

Since $p^{q-1} \equiv 1 \pmod{q}$, we have $a^q \equiv -1 \pmod{q}$, which implies $a^q \equiv -1 \pmod{q^2}$, which, together with (11), implies (9). \square

¹It is an algebraic integer, and for any $\sigma \in G$ we have $|(\eta/\bar{\eta})^\sigma| = 1$

4. LOGARITHMIC FORMS, TIJDEMAN'S ARGUMENT AND THE RELATION $p \not\equiv 1 \pmod{q}$

As I mentioned in the introduction, Tijdeman [32] applied Baker's theory of logarithmic form to establish an effective upper bound for the solutions, reducing the problem to a finite computation. In this section we use Tijdeman's argument and electronic computations due to Mignotte and Roy to prove the following important theorem.

THEOREM 4.1. — Let (x, y, p, q) be a solution of (2). Then $p \not\equiv 1 \pmod{q}$.

The relation $p \not\equiv 1 \pmod{q}$ is indispensable for Mihăilescu's proof. It is repeatedly used in Section 6 and in the proof of Theorem 6.3.2. A reader ready to take Theorem 4.1 for granted may skip the rest of this section.

When writing this section, I profited a lot from helpful explanations and suggestions of Maurice Mignotte and Andrew Glass.

4.1. Logarithmic forms

In this subsection we recall Baker's lower bound for logarithmic forms

$$\Lambda = b_1 \log \alpha_1 + \cdots + b_n \log \alpha_n.$$

Here b_1, \dots, b_n are non-zero integers and $\alpha_1, \dots, \alpha_n$ are usually algebraic numbers. To avoid unnecessary technicalities, we shall assume that $\alpha_1, \dots, \alpha_n$ are **positive rational numbers, distinct from 1**. This is totally sufficient for applications in Catalan's problem.

Define the height of a rational number $\alpha = \mu/\nu$ (where μ and ν are relatively prime integers) by $h(\alpha) = \log \max\{|\mu|, |\nu|\}$. Assume that $\Lambda \neq 0$. Then it is rather easy to bound $|\Lambda|$ from below. Indeed, $e^\Lambda - 1$ is a non-zero rational number with denominator bounded by $e^{(h(\alpha_1) + \cdots + h(\alpha_n))B}$, where

$$B = \max\{|b_1|, \dots, |b_n|\}.$$

It follows that

$$(12) \quad |\Lambda| \gg e^{-(h(\alpha_1) + \cdots + h(\alpha_n))B},$$

where here and below in this subsection the positive constants implied by $O(\cdot)$, \ll and \gg are absolute and effective.

However, (12) is too weak for applications: one needs $o(B)$ in the exponent. Such an estimate was obtained by Gelfond [10] for $n = 2$ and by Baker [2] in the general case. Baker's inequality belongs to the top arithmetical results of the twentieth century.

The modern estimate [4, 18, 33] is of the form

$$(13) \quad |\Lambda| \geq e^{-c(n)h(\alpha_1) \cdots h(\alpha_n) \log B}$$

(provided $\Lambda \neq 0$.) See the recent volume [35] for the history of the subject and the present state of art.

When one wants to be explicit, the numerical value of the constant $c(n)$ becomes vital. For growing n , the best result is due to Matveev [18], who showed that one may take $c(n) = c^n$ with an explicit absolute constant c .

However, in Catalan's problem one uses (13) only with $n = 2$ and $n = 3$. Therefore it is practical to have special bounds for these two cases, which are numerically sharper than the general bound (13). Such bounds were obtained by Laurent, Mignotte and Nesterenko [16] for binary forms and by Bennett *et al.* [5] for ternary forms. Here is a simplified form of the Laurent-Mignotte-Nesterenko result (see Corollary 2 from [16, Section 2]), to be used in Subsection 4.3 below.

PROPOSITION 4.1.1. — Let α_1, α_2 be multiplicatively independent positive rational numbers and b_1, b_2 positive integers. Let A_1, A_2 be real numbers satisfying $A_i \geq \max\{h(\alpha_i), 1\}$ for $i = 1, 2$. Put $B = b_1/A_2 + b_2/A_1$ and $\Lambda = b_1 \log \alpha_1 - b_2 \log \alpha_2$. Then

$$(14) \quad \log |\Lambda| \geq -24.34 (\max\{\log B + 0.14, 21\})^2 A_1 A_2.$$

This is asymptotically weaker than (13) when B grows (because $\log B$ is replaced by $(\log B)^2$), but for small B inequality (14) is very sharp numerically.

I do not formulate the result of [5], because it is very involved and will not be used here.

4.2. An informal introduction to Tijdeman's argument

In this subsection we assume that $p > q$.

In Catalan's problem, the most obvious logarithmic form to try is $\Lambda = p \log |x| - q \log |y|$. The upper estimate is obvious: $|\Lambda| \leq |x|^{-p}$. The lower estimate coming from (13) is $|\Lambda| \geq e^{-O(p \log |x| \log |y|)}$, and comparing the two estimates does not yield any interesting consequence.

Tijdeman's [32] brilliant idea was to use $\Lambda = q \log |y + 1| - p \log |x|$. Upper bound is now slightly worse: $|\Lambda| \ll q |y|^{-1}$. For the lower bound, we use Cassels' relations (6) to obtain $\Lambda = p \log \alpha - q \log q$, where $\alpha = (q|b|)^{q-1} u^{-1}$ (recall that $u > 0$). It is easy to show (see Subsection 4.3) that $h(\alpha) = \log |u| + O(1) \leq (q/p) \log |y| + O(1)$. Now (13) implies that $|\Lambda| \geq e^{-O((q/p) \log |y| \log q \log p)}$, which, compared with the lower estimate, implies that

$$(15) \quad p \ll q \log q \log p.$$

If (14) is used instead of (13), then one obtains the slightly weaker inequality

$$(16) \quad p \ll q \log q (\log p)^2.$$

Similarly, using

$$\Lambda = q \log |y + 1| - p \log |x - 1| = pq \log \beta - q \log q + p \log p$$

with $\beta = bq/ap$, one obtains the estimate

$$(17) \quad q \ll (\log p)^2 \log q.$$

Together with (15) this implies an effective upper bound for p , as wanted.

As I already mentioned in Subsection 4.1, Tijdeman's argument does not require the full strength of Baker's inequality. One needs a lower bound for binary logarithmic forms to obtain (15) and a lower bound for ternary logarithmic form to obtain (17).

Langevin [15] made Tijdeman's work explicit by proving that $p, q \leq 10^{110}$. This bound has been refined several times until O'Neil [27] (see also [6]) proved that $p \leq 3.2 \cdot 10^{17}$ and $q \leq 2.6 \cdot 10^{12}$, and Mignotte [25] announced that $p \leq 7.8 \cdot 10^{16}$ and $q \leq 7.2 \cdot 10^{11}$. Mignotte used the already mentioned bounds for binary and ternary logarithmic forms from [16] and [5], respectively.

4.3. Explicit Tijdeman's inequality

In this subsection we apply Proposition 4.1.1 to obtain an explicit analogue of (16).

PROPOSITION 4.3.1. — For any solution of (2) we have

$$(18) \quad p \leq 24.34q \left(\max \left\{ \log \frac{p+1}{\log q} + 0.14, 21 \right\} \right)^2 \log q.$$

Inequality (18) will be used in Subsection 4.5. It is less sharp than the corresponding results from [23] and [6], but easier to prove and sufficient for our purposes.

Proof of Proposition 4.3.1 : We may assume that

$$(19) \quad p \geq 10000q \log q,$$

and, in particular, $p > q$, since otherwise (18) holds trivially.

As indicated in Subsection 4.2, we will compare upper and lower estimates for the quantity

$$\Lambda = q \log |y+1| - p \log |x| = p \log \alpha - q \log q,$$

with $\alpha = (q|b|)^{q-1}u^{-1}$, where $b \in \mathbf{Z}$ and $u \in \mathbf{Z}_{>0}$ are defined in Proposition 2.1.

The upper estimate is trivial. Rewriting Catalan's equation (2) as

$$p \log |x| = q \log |y| + \log(1 + y^{-q}),$$

we obtain

$$(20) \quad \Lambda = q \log(1 + y^{-1}) - \log(1 + y^{-q}).$$

Since $|\log(1+t)| \leq 2|t|$ for $|t| \leq 1/2$, this implies that

$$(21) \quad |\Lambda| \leq 2|y|^{-q} + 2q|y|^{-1} \leq 3q|y|^{-1},$$

and $|\Lambda| < 1$ by (8). Equality (20) implies also that $\Lambda \neq 0$: the first term always dominates over the second one.

For the lower bound, let us estimate $h(\alpha)$. Since

$$\log((q|b|)^{q-1}) = \log u + (\Lambda + q \log q)/p \leq \log u + 1,$$

we have $h(\alpha) \leq \log u + 1$. Also, q and α are multiplicatively independent: otherwise, Λ would have been a multiple of $\log q$, contradicting the previously established inequality $0 < |\Lambda| < 1$.

Thus, we are in a position to use Proposition 4.1.1. We obtain

$$(22) \quad \log |\Lambda| \geq -24.34 (\max\{\log B + 0.14, 21\})^2 (\log u + 1) \log q$$

with $B = p/\log q + q/(\log u + 1)$. Proposition 2.3 and (19) imply that

$$(23) \quad u \geq q^{p-2} \geq e^{9000q(\log q)^2}.$$

Hence $B \leq (p+1)/\log q$. Substituting this into (22) and combining the resulting inequality with (21), we obtain

$$(24) \quad \frac{\log |y|}{\log u} \leq 24.34 \left(\max\left\{ \log \frac{p+1}{\log q} + 0.14, 21 \right\} \right)^2 \log q \left(1 + \frac{1}{\log u} \right) + \frac{\log(3q)}{\log u}.$$

Further, (6) implies that $q|y|^{q-1} \geq qu^p$, whence

$$(25) \quad p \leq (q-1) \frac{\log |y|}{\log u} \leq 24.34(q-1) \left(\max\left\{ \log \frac{p+1}{\log q} + 0.14, 21 \right\} \right)^2 \log q \left(1 + \frac{1}{\log u} \right) + \frac{(q-1) \log(3q)}{\log u}.$$

Using (19) and (23), one easily shows that the right-hand side of (25) does not exceed the right-hand side of (18). The proposition is proved. \square

4.4. Lower bounds for p and q

One can bound exponents p and q from below, using algebraic criteria (see Section 3) and electronic computations. This has been realized by Mignotte and Roy [23, 24, 25]. To show that $q \geq Q_0$, one has to verify an algebraic criterion (Inkeri's or other), for all pairs (p, q) satisfying $q \leq Q_0$, $p > q$ and (18). Actually, Mignotte and Roy used sharper, than (18), inequalities.

With Inkeri-type criteria, Mignotte and Roy managed to prove that

$$(26) \quad \min\{p, q\} \geq 10^5,$$

using several months of computations. With Mihăilescu's criterion (Theorem 3.2) this required only a few hours of computations, and with one month of computations they managed to prove that $\min\{p, q\} \geq 10^7$. I am aware about the computations of Grantham and Wheeler showing that $\min\{p, q\} \geq 3.2 \cdot 10^8$ but I have never seen this result announced in print.

Inequality (26) will be used in Subsection 4.5.

4.5. Proof of Theorem 4.1

First of all, we deduce from Proposition 4.3.1 the following consequence.

PROPOSITION 4.5.1. — If $q \geq 28000$ then $p \leq 4q^2$.

Proof : Assume first that

$$\log \frac{p+1}{\log q} + 0.14 \leq 21.$$

Then (18) reads $p \leq 10734q \log q$, and $p \geq 4q^2$ would imply $q \leq 2683.5 \log q$, which is wrong for $q \geq 28000$.

Now assume that

$$\log \frac{p+1}{\log q} + 0.14 \geq 21.$$

Then (18) reads

$$p \leq 24.34q \left(\log \frac{p+1}{\log q} + 0.14 \right)^2 \log q.$$

Since $0.14 - \log \log q \leq 0.14 - \log \log 28000 \leq -2.18$, this implies that

$$(27) \quad \frac{p}{(\log(p+1) - 2.18)^2} \leq 24.34q \log q.$$

It is easy to show, calculating the derivative, that the left hand-side of (27), viewed as a function in p , increases when $p \geq 67$. Hence, assuming that $p \geq 4q^2$, we may replace in (27) p by $4q^2$, which would result in the inequality $q \leq 6.085 (\log(4q^2 + 1) - 2.18)^2 \log q$. Since $\log(4q^2 + 1) - 2.27 \leq \log q^2$, we obtain the inequality $q \leq 24.34(\log q)^3$, which is contradictory for $q \geq 28000$. \square

Proof of Theorem 4.1 : Assume that $p \equiv 1 \pmod{q}$. Wieferich's relation (9) implies that $p \equiv 1 \pmod{q^2}$. Since p is odd, it cannot be equal to $q^2 + 1$ or $3q^2 + 1$. Also, $p \neq 2q^2 + 1$, because the latter number is divisible by 3. (This simple, but important observation is due to Mignotte.) Thus, $p \geq 4q^2 + 1$. On the other hand (26) and Proposition 4.5.1 imply that $p \leq 4q^2$, a contradiction.

REMARK 4.2. — Inequality (26) is the only result, used by Mihăilescu, that depends on electronic computations. One can avoid using it, showing instead that

there exist no pairs (p, q) satisfying $q < 28000$,

$$1 + 4q^2 \leq p \leq 24.34q \left(\max \left\{ \log \frac{p+1}{\log q} + 0.14, 21 \right\} \right)^2 \log q,$$

$p \equiv 1 \pmod{q^2}$ and $q^{p-1} \equiv 1 \pmod{p^2}$.

The running time of the corresponding PARI-script (written by Preda Mihăilescu at my request) is about 1 minute on a modern computer.

Still, it would be very interesting to find a purely algebraic proof of $p \not\equiv 1 \pmod{q}$, or, at least, a proof independent of electronic computations. Recently Mihăilescu announced that he has such a proof.

5. GENERALITIES

In this section we recall some simple results about modules over commutative rings and several other facts to be used in the proof. They are certainly well-known, but it was easier for me to supply direct proofs than to look for suitable references.

All rings in this section are **commutative and with unity**. An ideal \mathfrak{a} of a ring R is *radical* if R/\mathfrak{a} has no non-zero nilpotent elements.

5.1. Rings and modules

Let R be a ring and M an R -module. Given a subset $S \subseteq M$, we denote by $\text{ann}_R(S)$ the ideal of annihilators of S in R . When no confusion is possible, we omit the index and write $\text{ann}(S)$. In this subsection *isomorphic* means R -isomorphic. For instance, a cyclic R -module M is (non-canonically) isomorphic to $R/\text{ann}(M)$.

The following property of cyclic modules is immediate.

PROPOSITION 5.1.1. — Let M be a cyclic R -module. Then any quotient of M is cyclic. If R is a principal ideal ring, then any submodule of R is cyclic as well. \square

PROPOSITION 5.1.2. — Let R be a ring and M a finitely generated R -module. Let \mathfrak{b} be an ideal of R such that $\mathfrak{b} + \text{ann}_R(M)$ is a radical ideal of R . Then $\text{ann}_{R/\mathfrak{b}}(M/\mathfrak{b}M)$ is the image of $\text{ann}_R(M)$ in R/\mathfrak{b} .

Proof : We have to prove that for any $\alpha \in R$ one has

$$\alpha M \subseteq \mathfrak{b}M \iff \alpha \in \mathfrak{b} + \text{ann}_R(M).$$

Implication “ \Leftarrow ” is obvious, so we are left with “ \Rightarrow ”. Let φ be an endomorphism of M such that $\varphi(M) \subseteq \mathfrak{b}M$. Then, according to [1, Proposition 2.4], there exist a positive integer n and $\beta_1, \dots, \beta_n \in \mathfrak{b}$ such that $\varphi^n + \beta_1\varphi^{n-1} + \dots + \beta_n = 0$. For φ equal to multiplication by α this means that $\alpha^n + \beta_1\alpha^{n-1} + \dots + \beta_n \in \text{ann}(M)$. Thus, $\alpha^n \in \mathfrak{b} + \text{ann}(M)$. Since the latter ideal is radical, we obtain $\alpha \in \mathfrak{b} + \text{ann}(M)$, which proves “ \Rightarrow ”. \square

PROPOSITION 5.1.3. — Let R be a direct product of finitely many fields: $R = \prod_{\alpha \in A} K_\alpha$, where each K_α is a field. Then we have the following.

- (1) If $B \subseteq A$ then the set $\mathcal{I}(B) := \{(x_\alpha)_{\alpha \in A} : x_\alpha = 0 \text{ for } \alpha \in B\}$ is an ideal of R , and all ideals are of this form. In particular, any quotient of R is itself a direct product of fields.
- (2) For any ideals $\mathcal{I}, \mathcal{I}' \trianglelefteq R$ one has $\mathcal{I}\mathcal{I}' = \mathcal{I} \cap \mathcal{I}'$. Moreover, for any $b \in \mathcal{I}\mathcal{I}'$ there exist $a \in \mathcal{I}$ and $a' \in \mathcal{I}'$ such that $b = aa'$. In particular, $\mathcal{I}^2 = \mathcal{I}$, and for any $a \in \mathcal{I}$ there exist $a_1, a_2 \in \mathcal{I}$ such that $a = a_1a_2$.
- (3) For any ideal $\mathcal{I} \trianglelefteq R$ there is a uniquely defined ideal $\mathcal{I}^\perp \trianglelefteq R$ such that $\mathcal{I}\mathcal{I}^\perp = (0)$ and $\mathcal{I} + \mathcal{I}^\perp = R$.

(4) For any ideals $\mathcal{I}, \mathcal{I}' \trianglelefteq R$ one has

$$(28) \quad (\mathcal{I}\mathcal{I}')^\perp = \mathcal{I}^\perp + \mathcal{I}'^\perp, \quad (\mathcal{I} + \mathcal{I}')^\perp = \mathcal{I}^\perp \mathcal{I}'^\perp.$$

Also, $\mathcal{I}\mathcal{I}' = (0)$ if and only if $\mathcal{I}' \subseteq \mathcal{I}^\perp$.

(5) Let M a cyclic R -module and M' is a submodule M . Then

$$\text{ann}(M') + \text{ann}(M/M') = R \quad \text{and} \quad \text{ann}(M')\text{ann}(M/M') = \text{ann}(M).$$

(6) Let M be an R -module. Then there exists $a \in M$ such that $\text{ann}(a) = \text{ann}(M)$. In other words, M has a submodule isomorphic to $R/\text{ann}(M)$. In particular, if R is finite then $|M| \geq |R/\text{ann}(M)|$, with equality if and only if M is cyclic.

Proof : **Part 1** is obvious, and **parts 2–5** are its immediate consequences.

In the sequel we write $\mathcal{I}(\alpha)$ for $\mathcal{I}(\{\alpha\})$. For $\beta \in A$ denote by $\mathbf{1}_\beta$ the element $(x_\alpha)_{\alpha \in A} \in R$ such that $x_\beta = 1$ and $x_\alpha = 0$ for $\alpha \neq \beta$. For any $x \in R \setminus \mathcal{I}(\beta)$ there exists $y \in R$ such that $yx = \mathbf{1}_\beta$.

After this preparation we are ready to prove **part 6**. Let $B \subseteq A$ be such that $\text{ann}(M) = \mathcal{I}(B)$. I claim that for any $\beta \in B$ there exists $b_\beta \in M$ such that $\text{ann}(b_\beta) \subseteq \mathcal{I}(\beta)$. Indeed, assume that for any $b \in M$ there exists $x \in R \setminus \mathcal{I}(\beta)$ such that $xb = 0$. Then, as follows from the previous paragraph, $\mathbf{1}_\beta b = 0$ for any $b \in M$, which is a contradiction because $\mathbf{1}_\beta \notin \text{ann}(M)$.

Now put $a = \sum_{\beta \in B} \mathbf{1}_\beta b_\beta$. If $x = (x_\alpha)_{\alpha \in A} \in \text{ann}(a)$, then for any $\beta \in B$ one has $0 = \mathbf{1}_\beta xa = \mathbf{1}_\beta xb_\beta$. Hence $\mathbf{1}_\beta x \in \mathcal{I}(\beta)$ by the choice of b_β , or, in other words, $x_\beta = 0$.

Thus, $x_\beta = 0$ for any $\beta \in B$. Hence $x \in \mathcal{I}(B)$, which proves part 6. \square

5.2. Group rings

Let A be a commutative ring and G a finite abelian group. Consider the group ring $A[G]$. Define the *weight* of $\Theta = \sum_{g \in G} n_g g \in A[G]$ by $w(\Theta) = \sum_{g \in G} n_g \in A$. The weight function is additive and multiplicative, defining thereby a ring homomorphism $A[G] \xrightarrow{w} A$. The kernel of this homomorphism is called *the augmentation ideal* of the group ring $A[G]$. It is generated over A by the elements of the form $\sigma - \tau$, where $\sigma, \tau \in G$.

The following proposition is true for any finite abelian groups, but we formulate it only for cyclic groups, which is sufficient for our purposes.

PROPOSITION 5.2.1. — Let G be a finite cyclic group of order n . Then we have the following.

- (1) Let K be a field of characteristic not dividing n . Then the group ring $K[G]$ is a direct product of finitely many fields.
- (2) An ideal of the ring $\mathbf{z}[G]$ containing a prime number not dividing n is a radical ideal of $\mathbf{z}[G]$.

Proof : **Part 1** follows by observing that $K[G] = K[x]/(x^n - 1)$. Since the characteristic does not divide n , the polynomial $x^n - 1$ is separable over K , which means that $K[x]/(x^n - 1)$ is a direct product of several finite extensions of K .

To prove **part 2**, let \mathfrak{a} be an ideal of $\mathbf{Z}[G]$ containing a prime number q not dividing n . Then $\mathbf{Z}[G]/\mathfrak{a} = \mathbf{F}_q[G]/\mathfrak{a}'$, where \mathbf{F}_q is the field of q elements and \mathfrak{a}' is the image of \mathfrak{a} in $\mathbf{F}_q[G]$. Part 1 implies that $\mathbf{F}_q[G]$ is a direct product of fields, and hence so is $\mathbf{F}_q[G]/\mathfrak{a}'$ (see Proposition 5.1.3:1). Thus, $\mathbf{Z}[G]/\mathfrak{a}$ has no non-zero nilpotents, as wanted. \square

5.3. Miscellaneous

PROPOSITION 5.3.1. — Let R be a ring and q a prime number such that the principal ideal (q) is radical. (In particular, the assumption is satisfied if K is a number field, q a prime number unramified in K and $R = S^{-1}\mathcal{O}_K$, where $S \subset \mathcal{O}_K$ consists of elements co-prime with q .) Let $\alpha, \beta \in R$ satisfy $\alpha^q \equiv \beta^q \pmod{q}$. Then $\alpha^q \equiv \beta^q \pmod{q^2}$.

Proof : We have $(\alpha - \beta)^q \equiv \alpha^q - \beta^q \equiv 0 \pmod{q}$. Since (q) is radical, this implies $\alpha \equiv \beta \pmod{q}$, which, in turn, yields $\alpha^q \equiv \beta^q \pmod{q^2}$. \square

PROPOSITION 5.3.2. — Let R be an integral domain and K its quotient field. Let

$$\sum_{k=0}^{\infty} \frac{a_k}{k!} T^k, \quad \sum_{k=0}^{\infty} \frac{b_k}{k!} T^k \in K[[T]]$$

be formal power series with the following properties:

$$a_k, b_k \in R, \quad a_k \equiv a^k \pmod{\mathfrak{a}}, \quad b_k \equiv b^k \pmod{\mathfrak{a}} \quad (k = 0, 1, \dots)$$

for some $a, b \in R$ and an ideal $\mathfrak{a} \trianglelefteq R$. Then

$$\left(\sum_{k=0}^{\infty} \frac{a_k}{k!} T^k \right) \left(\sum_{k=0}^{\infty} \frac{b_k}{k!} T^k \right) = \sum_{k=0}^{\infty} \frac{c_k}{k!} T^k$$

with $c_k \in R$ satisfying $c_k \equiv (a + b)^k \pmod{\mathfrak{a}}$.

Proof : We have $c_k = \sum_{i=0}^k \binom{k}{i} a_i b_{k-i} \equiv \sum_{i=0}^k \binom{k}{i} a^i b^{k-i} = (a + b)^k$, as wanted. \square

PROPOSITION 5.3.3. — Let m be a non-negative integer and α a rational number with denominator b . Then for a sufficiently large positive integer N one has $b^N \binom{\alpha}{m} \in \mathbf{Z}$.

Proof : Write $\alpha = a/b$. For any prime number p not dividing b we have

$$\text{ord}_p(a(a-b) \cdots (a - (m-1)b)) \geq \lfloor m/p \rfloor + \lfloor m/p^2 \rfloor + \cdots = \text{ord}_p(m!),$$

whence the result. \square

6. OVERVIEW OF THE PROOF

In this section I give a general overview of the proof of Theorem 1.3 and show how it reduces to three more technical statements.

6.1. **Three steps**

The structure of the proof of Theorem 1.3 resembles that of the proof of the statement $q^2|x$ in Theorem 3.2. Recall that the latter argument consisted of three steps.

- (1) Find “many” $\Theta \in \mathbf{z}[G]$ such that $(x - \zeta)^\Theta$ is a q -th power.
- (2) Show that $(x - \zeta)^\Theta$ is a q -th power only if either $q|\Theta$ or $q^2|x$.
- (3) Show that not all Θ from step (1) are divisible by q .

The proof of Theorem 1.3 has steps (1) and (3), but step (2) has to be replaced by the following much more difficult task:

- (2*) Show that $(x - \zeta)^\Theta$ is a q -th power only if $q|\Theta$.

Unfortunately, we are able to verify step (2*) only if Θ is even, that is, $(1 + \iota)|\Theta$, where $\iota \in G$ is the complex conjugation. This creates several serious problems.

First of all, there are too few even elements in the Stickelberger ideal (see [34, Example (a) after Theorem 6.10]). Hence we cannot use Stickelberger’s theorem anymore, and have to find a substitute. Fortunately, such a substitute is available: it is the famous theorem of Thaine [31], who gave a (partial) analogue of Stickelberger’s theorem for real abelian fields.

Second, now we have $\Theta = (1 + \iota)\Theta'$ rather than $\Theta = (1 - \iota)\Theta'$. Hence instead of $((1 - \bar{\zeta})/(1 - \zeta))^{\Theta'} (\eta/\bar{\eta})$, as in (10), which is a root of unity, we have $((1 - \bar{\zeta})(1 - \zeta))^{\Theta'} \eta\bar{\eta}$, which is usually not a root of unity and *a priori* has no reasons to be a q -th power.

In Subsection 6.3 we reduce Theorem 2 to three statements, corresponding to steps (1), (2*) and (3) above. But before, we need some preparations. It will be more convenient to work mod q . In the next subsection we introduce certain modules over the ring $F_q[G]$ which will play vital role in the sequel.

6.2. **The ring $R = F_q[G]$ and some R -modules**

In this subsection p and q are distinct odd prime numbers satisfying

$$(29) \quad p \not\equiv 1 \pmod{q}.$$

As usual, ζ is a primitive p -th root of unity, $K = \mathbf{Q}(\zeta)$ and $G = \text{Gal}(K/\mathbf{Q})$.

Consider the group ring $R = F_q[G]$. Relation (29) and Proposition 5.2.1:1 imply that R is a direct product of fields, and, in particular, **Proposition 5.1.3 applies to this ring**. This will be repeatedly used in the sequel.

By Proposition 5.1.3, for any ideal $\mathcal{I} \trianglelefteq R$ there is a uniquely defined $\mathcal{I}^\perp \trianglelefteq R$ such that $\mathcal{I} + \mathcal{I}^\perp = R$ and $\mathcal{I}\mathcal{I}^\perp = 0$. For instance, $(1 + \iota)^\perp = (1 - \iota)$, where $\iota \in G$ is the complex conjugation, and $(\mathcal{N})^\perp$ is the augmentation ideal (see Subsection 5.2), where

$$\mathcal{N} = \sum_{\sigma \in G} \sigma \in R$$

is the “norm” element.

PROPOSITION 6.2.1. — Let E be the group of units of K . Then E/E^q is a cyclic R -module, and, in the notation of Subsection 5.1, we have

$$(30) \quad \text{ann}(E/E^q) = (\mathcal{N}, 1 - \iota).$$

Proof Let Ω the group of roots of unity from K and put $\bar{E} = E/\Omega$. Since the roots of unity in K are q -th powers, E/E^q is G -isomorphic to \bar{E}/\bar{E}^q .

In every pair of complex conjugate elements of G pick a representative. Denote by $\tilde{\mathcal{N}}'$ the sum of chosen representatives in $\mathbf{z}[G]$ and by \mathcal{N}' its image in $R = \mathbf{F}_q[G]$, so that $\mathcal{N} = \mathcal{N}'(1 + \iota)$. Then the annihilator of the $\mathbf{z}[G]$ -module \bar{E} is $(\tilde{\mathcal{N}}', 1 - \iota)$. Proposition 5.2.1:2 implies that $(\tilde{\mathcal{N}}', 1 - \iota, q)$ is a radical ideal of $\mathbf{z}[G]$, and Proposition 5.1.2 implies that $\text{ann}_{\mathbf{F}_q[G]}(\bar{E}/\bar{E}^q) = (\mathcal{N}', 1 - \iota)$. Since

$$\mathcal{N}' = \frac{1}{2}(\mathcal{N} + (1 - \iota)\mathcal{N}') \in (\mathcal{N}, 1 - \iota),$$

we have $(\mathcal{N}', 1 - \iota) = (\mathcal{N}, 1 - \iota)$, which proves (30).

Further, since $1 - \iota$ belongs to the augmentation ideal $(\mathcal{N})^\perp$, we have $\mathcal{N} \cap (1 - \iota) = \mathcal{N}(1 - \iota) = (0)$, which implies that

$$|(\mathcal{N}, 1 - \iota)| = |(\mathcal{N})| \cdot |(1 - \iota)| = q \cdot q^{(p-1)/2} = q^{(p+1)/2}.$$

We obtain $|R/(\mathcal{N}, 1 - \iota)| = q^{(p-3)/2} = |\bar{E}/\bar{E}^q|$, and the R -module \bar{E}/\bar{E}^q is cyclic by Proposition 5.1.3:6. □

DEFINITION 6.1. — We say that $\beta \in \mathcal{O}_K$ is q -primary if there exists $\gamma \in \mathcal{O}_K$ such that $\beta \equiv \gamma^q \pmod{q^2}$.

Denote by C and C_q the groups of cyclotomic units and of q -primary cyclotomic units of K , respectively. Recall C is, by definition, the group generated by $-\zeta$ and units of the form $(1 - \zeta^k)/(1 - \zeta)$. It is a full rank subgroup of E .

The R -modules E/CE^q , C/C_q and $C_q/(C_q \cap E^q)$ and their annihilators play a central role in Mihăilescu’s work. Since $C/C_q \cong CE^q/C_qE^q$ and $C_q/(C_q \cap E^q) \cong C_qE^q/E^q$, all three are cyclic R -modules by Proposition 5.1.1. Moreover, Proposition 5.1.3:5 and equality (30) imply the following.

PROPOSITION 6.2.2. — The three ideals

$$(31) \quad \mathcal{I}_1 = \text{ann}(E/CE^q), \quad \mathcal{I}_2 = \text{ann}(C/C_q), \quad \mathcal{I}_3 = \text{ann}(C_q/(C_q \cap E^q))$$

are pairwise coprime and satisfy

$$(32) \quad \mathcal{I}_1\mathcal{I}_2\mathcal{I}_3 = (\mathcal{N}, 1 - \iota).$$

□

6.3. The three main theorems

In this subsection we reduce Theorem 1.3 to three statements, corresponding to steps (1), (2*) and (3) from Subsection 6.1. We use the notation of Subsection 6.2.

REMARK 6.3.1. — In the sequel, for $\gamma \in K^*$ and $\Theta \in R$ we define γ^Θ as $\gamma^{\tilde{\Theta}}$, where $\tilde{\Theta}$ is a lifting of Θ to $\mathbf{z}[G]$. Of course, γ^Θ is well-defined only up to multiplication by a q -th power. This, however, will never be confusing, since any statement involving terms like γ^Θ will include the q -th power of an (unspecified) element of K^* .

In the first two theorems, x, y, p, q is a solution of the Catalan equation (2). In particular, (29) is satisfied, as follows from Theorem 4.1.

THEOREM 6.3.2. — For any $\Theta \in (\mathcal{N})^\perp(1 + \iota)\mathcal{I}_1\mathcal{I}_3$ we have $(x - \zeta)^\Theta \in (K^*)^q$.

THEOREM 6.3.3. — Assume that $q \geq 7$. If for $\Theta \in (\mathcal{N})^\perp(1 + \iota)$ we have $(x - \zeta)^\Theta \in (K^*)^q$, then $\Theta = 0$.

The third theorem is a general fact, independent of Catalan's condition; in fact, even (29) is not required.

THEOREM 6.3.4. — If $p > q$ then $C_q \neq C$.

Proof of Theorem 1.3 (assuming Theorems 6.3.2, 6.3.3 and 6.3.4) : Let (x, y, p, q) be a solution. Replacing it, if necessary, by $(-y, -x, q, p)$, we may assume that $p > q$. We may also assume that $q \geq 7$ by (26). Thus, the assumptions of Theorems 6.3.2–6.3.4 are verified.

Theorems 6.3.2 and 6.3.3 imply that $(1 + \iota)(\mathcal{N})^\perp\mathcal{I}_1\mathcal{I}_3 = (0)$, which, by Proposition 5.1.3:4 and (32), implies that

$$\mathcal{I}_1\mathcal{I}_3 \subseteq ((1 + \iota)(\mathcal{N})^\perp)^\perp = (1 + \iota)^\perp + (\mathcal{N}) = (1 - \iota) + (\mathcal{N}) = \mathcal{I}_1\mathcal{I}_2\mathcal{I}_3.$$

On the other hand \mathcal{I}_2 and $\mathcal{I}_1\mathcal{I}_3$ are co-prime by Proposition 6.2.2. Hence

$$1 \in \mathcal{I}_2 + \mathcal{I}_1\mathcal{I}_3 \subseteq \mathcal{I}_2 + \mathcal{I}_1\mathcal{I}_2\mathcal{I}_3 = \mathcal{I}_2,$$

that is, $\mathcal{I}_2 = (1)$. Since $\mathcal{I}_2 = \text{ann}(C/C_q)$, this means that $C = C_q$, contradicting Theorem 6.3.4. \square

Theorems 6.3.2, 6.3.3 and 6.3.4 are proved in the next three sections. Theorem 6.3.2 is purely algebraic and relies on the already mentioned result of Thaine about cyclotomic fields. The proof of Theorem 6.3.3 is a beautiful Runge-type diophantine argument, while that of Theorem 6.3.4 is short and elementary.

7. PROOF OF THEOREM 6.3.2

Thus, let (x, y, p, q) be a solution of Catalan's equation (2). Theorem 4.1 implies that $p \not\equiv 1 \pmod q$. In particular, Proposition 5.1.3 applies to the group ring $R = F_q[G]$, to be repeatedly used in the sequel.

Let H be the class group of the number field $K = \mathbf{Q}(\zeta)$ and H^+ the “plus-part” of H (it consists of the classes stable with respect to the complex conjugation). Recall that $\Theta \in \mathbf{Z}[G]$ is called *even* if it is divisible by $1 + \iota$. The following is a particular case of [34, Theorem 15.2].

THEOREM 7.1 (Thaine). — Let an even $\Theta \in \mathbf{Z}[G]$ annihilate the q -part of the group E/C . Then Θ annihilates the q -part of H^+ as well.

(By the q -part we mean the q -Sylow subgroup.)

REMARK 7.2. — Thaine's result is more general. Let L be a real abelian field, and denote by E_L, C_L, H_L and G_L , the groups of units, of cyclotomic units, the class group and the Galois group of L , respectively. Let q be an odd prime number not dividing $[L : \mathbf{Q}]$. Then any $\Theta \in \mathbf{Z}[G_L]$, annihilating the q -part of the group E_L/C_L , annihilates the q -part of H_L as well. For $q = 2$ a slightly weaker statement holds.

In our case $L = \mathbf{Q}(\zeta + \bar{\zeta})$ and the condition “ q does not divide $[L : \mathbf{Q}]$ ” is ensured by $p \not\equiv 1 \pmod q$.

We shall use the following consequence of Theorem 7.1.

PROPOSITION 7.3. — Any $\Theta \in (1 + \iota)\mathcal{I}_1$ has a lifting $\tilde{\Theta} \in \mathbf{Z}[G]$ annihilating the q -part of H .

Proof : Let q^m be the order of the q -part of E/C . By Proposition 5.1.3:2, there exist $\Theta_1, \dots, \Theta_m \in \mathcal{I}_1$ such that $\Theta = (1 + \iota)^2 \Theta_1 \cdots \Theta_m$. Pick liftings $\tilde{\Theta}_1, \dots, \tilde{\Theta}_m$ for $\Theta_1, \dots, \Theta_m$, respectively, and put $\tilde{\Theta}' = (1 + \iota)\tilde{\Theta}_1 \cdots \tilde{\Theta}_m$ and $\tilde{\Theta} = (1 + \iota)\tilde{\Theta}'$. Since every Θ_i annihilates E/CE^q , we have $E^{\Theta_i} \subseteq CE^q$, which implies $E^{\Theta'} \subseteq CE^{q^m}$. By the definition of m this means that $\tilde{\Theta}'$ annihilates the q -part of E/C . By Thaine's theorem, it annihilates the q -part of H^+ as well. Since $H^{1+\iota} \subseteq H^+$, the q -part of H is annihilated by $\tilde{\Theta} = (1 + \iota)\tilde{\Theta}'$. \square

PROPOSITION 7.4. — For any $\Theta \in (1 + \iota)(\mathcal{N})^\perp \mathcal{I}_1$ we have $(x - \zeta)^\Theta \in E(K^*)^q$.

Proof Put $\lambda = (x - \zeta)/(1 - \zeta)$. By Corollary 2.2, there exists an ideal \mathfrak{a} of K such that $(\lambda) = \mathfrak{a}^q$. The class of the ideal \mathfrak{a} belongs to the q -part of the class group of K . Since the statement of the proposition does not depend on the choice of the lifting $\tilde{\Theta}$ used to define (cf. Remark 6.3.1) $(x - \zeta)^\Theta$, we may select $\tilde{\Theta}$ in the most suitable way. Thus, let $\tilde{\Theta}$ be a lifting which annihilates the q -part of the class group, which exists by Proposition 7.3.

Then \mathfrak{a}^Θ is a principal ideal. Thus, the principal ideal (λ^Θ) is a q -th power of another principal ideal, that is, $\lambda^\Theta \in E(K^*)^q$.

On the other hand, since Θ belongs to the augmentation ideal $(\mathcal{N})^\perp$, we have $(1 - \zeta)^\Theta \in C(K^*)^q \subseteq E(K^*)^q$. (Indeed, the augmentation ideal is generated by the elements of the form $\sigma - \tau$, where $\sigma, \tau \in G$; and $(1 - \zeta)^{\sigma - \tau}$ is a cyclotomic unit.) Thus, $(x - \zeta)^\Theta = \lambda^\Theta(1 - \zeta)^\Theta \in E(K^*)^q$, as wanted. \square

Next, we use Mihăilescu's Theorem 3.2 to refine Proposition 7.4. Recall that C_q stands for the group of q -primary cyclotomic units.

PROPOSITION 7.5. — For any $\Theta \in (1 + \iota)(\mathcal{N})^\perp \mathcal{I}_1$ we have $(x - \zeta)^\Theta \in C_q(K^*)^q$.

Proof : By Proposition 5.1.3:2 we have $\Theta = \Theta_1 \Theta_2$ with $\Theta_1 \in (1 + \iota)(\mathcal{N})^\perp \mathcal{I}_1$ and $\Theta_2 \in \mathcal{I}_1$. Proposition 7.4 implies that $(x - \zeta)^{\Theta_1} \in E(K^*)^q$. Since $\Theta_2 \in \mathcal{I}_1 = \text{ann}(E/CE^q)$, we have

$$(x - \zeta)^\Theta = (x - \zeta)^{\Theta_1 \Theta_2} \in E^{\Theta_2}(K^*)^q \subseteq C(K^*)^q.$$

Write $(x - \zeta)^\Theta = \eta \alpha^q$ with $\eta \in C$ and $\alpha \in K^*$. Since $q^2 | x$ by Theorem 3.2, we have $\eta \alpha^q \equiv (-\zeta)^\Theta \pmod{q^2}$. Since $-\zeta$ is a q -th power, η is q -primary, and the proposition follows. \square

We are ready to prove Theorem 6.3.2. Let $\Theta \in (1 + \iota)(\mathcal{N})^\perp \mathcal{I}_1 \mathcal{I}_3$. By Proposition 5.1.3:2 we have $\Theta = \Theta_1 \Theta_2$ with $\Theta_1 \in (1 + \iota)(\mathcal{N})^\perp \mathcal{I}_1$ and $\Theta_2 \in \mathcal{I}_3 = \text{ann}(C_q/(C_q \cap E^q))$. Now

$$(x - \zeta)^\Theta = (x - \zeta)^{\Theta_1 \Theta_2} \in C_q^{\Theta_2}(K^*)^q \subseteq (K^*)^q.$$

Theorem 6.3.2 is proved. \square

8. PROOF OF THEOREM 6.3.3

8.1. A reformulation

In the proof of Theorem 6.3.3 it is more practical to work in the ring $\mathbf{Z}[G]$ rather than $\mathbf{F}_q[G]$. Thus, we have to find a suitable lifting of $\Theta \in \mathbf{F}_q[G]$ to $\mathbf{Z}[G]$. Since $(x - \zeta)^\Theta$ is a q -th power if and only if $(x - \zeta)^{-\Theta}$ is, we may choose between lifting Θ or $-\Theta$.

Recall that an element Θ of $\mathbf{F}_q[G]$ or $\mathbf{Z}[G]$ is *even* if it is divisible by $1 + \iota$. Equivalently, $\Theta = \sum_{\sigma \in G} n_\sigma \sigma$ is even if for any $\sigma \in G$ we have $n_\sigma = n_{\bar{\sigma}}$, where $\bar{\sigma} = \iota \sigma$ is the complex conjugate of σ .

We say that $\Theta = \sum_{\sigma \in G} n_\sigma \sigma \in \mathbf{Z}[G]$ is *non-negative* if $n_\sigma \geq 0$ for any $\sigma \in G$. We say that $\Theta \in \mathbf{Z}[G]$ is *positive* if it is non-negative and distinct from 0.

PROPOSITION 8.1.1. — Let $\Theta \in \mathbf{F}_q[G]$. Then either Θ or $-\Theta$ has a non-negative lifting $\tilde{\Theta} \in \mathbf{Z}[G]$ such that $w(\tilde{\Theta}) \leq q(p - 1)/2$. If Θ belongs to the augmentation ideal of $\mathbf{F}_q[G]$ then $q | w(\tilde{\Theta})$. If Θ is even then so is $\tilde{\Theta}$.

Proof : Let $\tilde{\Theta}_1$ be the smallest non-negative lifting of Θ . That is, $\tilde{\Theta}_1 = \sum_{\sigma \in G} \tilde{n}_\sigma \sigma$ with $\tilde{n}_\sigma \in \{0, 1, \dots, q-1\}$. Further, put $\tilde{\Theta}_2 = q \sum_{\sigma \in G} \sigma - \tilde{\Theta}_1$, so that $\tilde{\Theta}_2$ is a non-negative lifting of $-\Theta$. Obviously, both $\tilde{\Theta}_1$ and $\tilde{\Theta}_2$ are even if Θ is, and both the weights $w(\tilde{\Theta}_1)$ and $w(\tilde{\Theta}_2)$ are divisible by q if Θ is belongs to the augmentation ideal.

Since $w(\tilde{\Theta}_1) + w(\tilde{\Theta}_2) = q(p-1)$, one of the weights $w(\tilde{\Theta}_1)$ and $w(\tilde{\Theta}_2)$ does not exceed $q(p-1)/2$. The proposition is proved. \square

By this proposition, Theorems 6.3.3 is equivalent to the following statement.

Theorem 6.3.3'. Let x, y, p, q be a solution of the Catalan equation with $q \geq 7$. Let Θ be an even positive element of $\mathbf{z}[G]$ satisfying $q|w(\Theta)$ and $w(\Theta) \leq q(p-1)/2$. Assume that $(x - \zeta)^\Theta$ is a q -th power in K . Then $q|\Theta$.

This theorem will be proved in Subsection 8.3, after some preparations in Subsection 8.2.

8.2. The power series $(1 - \zeta T)^{\Theta/q}$

In this section we investigate the properties of a special power series introduced by Mihăilescu. Everywhere below capital T stands for an independent variable, while small letters t, z etc. denote complex numbers. For instance, $(1 + T)^r$ denotes the binomial series $\sum_{k=0}^{\infty} \binom{r}{k} T^k$, while, for $|t| < 1$, the expression $(1 + t)^r$ is the complex number, equal to the sum of the binomial series at $T = t$. In particular, $(1 + t)^r$ is a positive real number if $r \in \mathbf{R}$ and $t \in (-1, 1)$.

Fix a $\Theta = \sum_{\sigma \in G} n_\sigma \sigma \in \mathbf{z}[G]$. The series we are interested in is

$$(33) \quad (1 - \zeta T)^{\Theta/q} = \prod_{\sigma \in G} (1 - \zeta^\sigma T)^{n_\sigma/q}.$$

Its convergence radius is 1. Let us estimate its remainder term. Write

$$(34) \quad (1 - \zeta T)^{\Theta/q} = \sum_{k=0}^{\infty} \alpha_k(\Theta) T^k,$$

and denote by $S_m(T) = \sum_{k=0}^m \alpha_k(\Theta) T^k$ the m -th partial sum.

PROPOSITION 8.2.1. — Let $\Theta \in \mathbf{z}[G]$ be non-negative. Then for $|z| < 1$ one has

$$(35) \quad \left| (1 - \zeta z)^{\Theta/q} - S_m(z) \right| \leq \binom{w(\Theta)/q + m}{m+1} (1 - |z|)^{-w(\Theta)/q - m - 1} |z|^{m+1}.$$

Proof : A power series $\sum_{k=0}^{\infty} a_k T^k$ with complex coefficients is *dominated* by the series $\sum_{k=0}^{\infty} b_k T^k$ with non-negative real coefficients if $|a_k| \leq b_k$ for $k = 0, 1, \dots$. The relation of dominance is preserved by addition and multiplication of power series.

Let $r > 0$ be a positive real number, and χ a complex number satisfying $|\chi| \leq 1$. Then the binomial series $(1 + \chi T)^r = \sum_{k=0}^{\infty} \binom{r}{k} \chi^k T^k$ is dominated by $(1 - T)^{-r} = \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} T^k$. Indeed, the coefficients of the latter series are positive and $\left| \binom{r}{k} \right| \leq \left| \binom{-r}{k} \right|$.

It follows that $(1 - \zeta T)^{\Theta/q}$ is dominated by $(1 - T)^{-\nu}$, where $\nu = w(\Theta)/q$. Denoting by $\bar{S}_m(T)$ the m -th partial sum of the series $(1 - T)^{-\nu}$, we obtain the following:

$$\begin{aligned} \left| (1 - \zeta z)^{\Theta/q} - S_m(z) \right| &\leq \left| (1 - |z|)^{-\nu} - \bar{S}_m(|z|) \right| \\ &\leq \sup_{0 \leq \xi \leq |z|} \left| \left(\frac{d^{m+1}(1 - T)^{-\nu}}{dT^{m+1}} \right) \Big|_{T=\xi} \right| \frac{|z|^{m+1}}{(m + 1)!} \\ &= \binom{\nu + m}{m + 1} (1 - |z|)^{-\nu - m - 1} |z|^{m+1}, \end{aligned}$$

as wanted. □

Next, we investigate the arithmetic of the coefficients of Mihăilescu’s series. Say that $\alpha \in K$ is a q -integer if $q^N \alpha \in \mathbf{Z}[\zeta]$ for a sufficiently large positive integer N .

PROPOSITION 8.2.2. — The coefficients $\alpha_0(\Theta), \alpha_1(\Theta), \dots$ of Mihăilescu’s series $(1 - \zeta T)^{\Theta/q}$ are q -integers. Write

$$(36) \quad (1 - \zeta T)^{\Theta/q} = \sum_{k=0}^{\infty} \frac{a_k(\Theta)}{q^k k!} T^k,$$

(so that $\alpha_k(\Theta) = a_k(\Theta)/q^k k!$). Then

$$(37) \quad a_k(\Theta) \in \mathbf{Z}[\zeta] \quad \text{and} \quad a_k(\Theta) \equiv \left(- \sum_{\sigma \in G} n_{\sigma} \zeta^{\sigma} \right)^k \pmod{q} \quad (k = 0, 1, \dots).$$

Proof : As follows from Proposition 5.3.3, for every $n \in \mathbf{Z}$ the coefficients of the series $(1 - \zeta T)^{n/q}$ are q -integers. Hence so are the coefficients of $(1 - \zeta T)^{\Theta/q}$.

Further, $(1 - \zeta q T)^{n/q} = \sum_{k=0}^{\infty} (b_k/k!) T^k$ with

$$b_k = n(n - q) \cdots (n - (k - 1)q) (-\zeta)^k \equiv (-n\zeta)^k \pmod{q}.$$

Now, applying Proposition 5.3.2 to the equality

$$\sum_{k=0}^{\infty} \frac{a_k(\Theta)}{k!} T^k = \prod_{\sigma \in G} (1 - \zeta^{\sigma} q T)^{n_{\sigma}/q},$$

we obtain (37). □

We arrived to the most delicate part of Mihăilescu’s argument. The G -action extends to the ring of power series $K[T]$ by $(\sum_{k=0}^{\infty} a_k T^k)^{\sigma} = \sum_{k=0}^{\infty} a_k^{\sigma} T^k$, and we have the “compatibility relation”

$$(38) \quad \left((1 - \zeta T)^{\Theta/q} \right)^{\sigma} = (1 - \zeta T)^{\sigma \Theta/q}.$$

However, since the Galois action is not continuous in the complex topology, this relation **does not**, in general, extend to the *values* of power series, even if the convergence is

ensured. For instance, if $t \in \mathbf{Q}$ satisfies $|t| < 1$ then we need not have

$$(39) \quad \left((1 - \zeta t)^{\Theta/q} \right)^\sigma = (1 - \zeta t)^{\sigma\Theta/q}.$$

In fact, the left-hand side is even not well-defined, because $(1 - \zeta t)^{\Theta/q}$ need not belong to the field K .

Nevertheless, under some additional assumptions (39) may hold.

PROPOSITION 8.2.3. — Assume that Θ is even. Let $t \in \mathbf{Q}$ satisfy $|t| < 1$, and assume that $(1 - \zeta t)^{\Theta/q} \in K$. Then (39) is true for any $\sigma \in G$.

Proof : Since Θ is even, the series $(1 - \zeta T)^{\Theta/q}$ has real coefficients. It follows that

$$\alpha := (1 - \zeta t)^{\Theta/q} \in \mathbf{R}.$$

Thus, α belongs to the real cyclotomic field $\mathbf{Q}(\zeta + \bar{\zeta})$, which implies that $\alpha^\sigma \in \mathbf{R}$ for any $\sigma \in G$.

Now fix $\sigma \in G$. Then $\sigma\Theta$ is also even, which implies that $\beta := (1 - \zeta t)^{\sigma\Theta/q} \in \mathbf{R}$ as well.

On the other hand, $(\alpha^\sigma)^q = (\alpha^q)^\sigma = \left((1 - \zeta t)^\Theta \right)^\sigma = (1 - \zeta t)^{\sigma\Theta}$. Hence α^σ is equal to the real q -th root of $(1 - \zeta t)^{\sigma\Theta}$, which is β . The proposition is proved. \square

8.3. Proof of Theorem 6.3.3'

8.3.1. *The number $(1 - \zeta/x)^{\Theta/q}$.* By the assumption, $(x - \zeta)^\Theta$ has a q -th root in the field K . Moreover, it has exactly one q -th root in K , because this field does not contain q -th roots of unity (other than 1).

Since Θ is even, $(x - \zeta)^\Theta$ is a positive real number. It follows that the real q -th root of $(x - \zeta)^\Theta$ belongs to K . This real root is equal to $|x|^{w(\Theta)/q} (1 - \zeta/x)^{\Theta/q}$, where $(1 - \zeta/x)^{\Theta/q}$ is defined as the sum of Mihăilescu series

$$(40) \quad (1 - \zeta T)^{\Theta/q} = \sum_{k=0}^{\infty} \alpha_k(\Theta) T^k$$

at $T = 1/x$.

So far, everything was true for any even Θ . Now recall the assumption $q|w(\Theta)$, that is, $w(\Theta) = mq$ with $m \in 2\mathbf{Z}$. We have just proved that $x^m (1 - \zeta/x)^{\Theta/q} \in K$. Hence $(1 - \zeta/x)^{\Theta/q} \in K$, and proposition 8.2.3 implies that

$$(41) \quad \left((1 - \zeta/x)^{\Theta/q} \right)^\sigma = (1 - \zeta/x)^{\sigma\Theta/q} \quad (\sigma \in G).$$

8.3.2. *The polynomial $P(T)$.* For $k = 1, 2, \dots$ put $E(k) = k + \text{ord}_q(k!)$. Then

$$(42) \quad E(k+1) \geq E(k) + 1,$$

$$(43) \quad E(k) \leq kq/(q-1).$$

Since Θ is positive, we have $m > 0$. Consider the polynomial

$$(44) \quad P(T) = q^{E(m)} (\alpha_0(\Theta)T^m + \alpha_1(\Theta)T^{m-1} + \dots + \alpha_m(\Theta)),$$

where $\alpha_k(\Theta)$ are the coefficients of the Mihăilescu series (40). Proposition 8.2.2 implies that $q^{E(k)}\alpha_k(\Theta) \in \mathbf{z}[\zeta]$. It follows that $P(T) \in \mathbf{z}[\zeta][T]$, and (42) implies that

$$(45) \quad P(T) \in q^{E(m)}\alpha_m(\Theta) + q\mathbf{z}[\zeta][T].$$

Also, (38) implies that for $\sigma \in G$

$$(46) \quad P^\sigma(T) = q^{E(m)}(\alpha_0(\sigma\Theta)T^m + \alpha_1(\sigma\Theta)T^{m-1} + \cdots + \alpha_m(\sigma\Theta)).$$

8.3.3. *The number β and its conjugates.* Since Θ is non-negative, the number $(x - \zeta)^\Theta$ is an algebraic integer. Therefore its q -th root $x^m(1 - \zeta/x)^{\Theta/q}$ is an algebraic integer as well. Hence so is

$$\beta := q^{E(m)}x^m(1 - \zeta/x)^{\Theta/q} - P(x).$$

Relations (41) and (46) imply that

$$(47) \quad \beta^\sigma = q^{E(m)}x^m \left((1 - \zeta/x)^{\sigma\Theta/q} - \sum_{k=0}^m \alpha_k(\sigma\Theta)x^{-k} \right) \quad (\sigma \in G).$$

Now estimate $|\beta^\sigma|$ using Proposition 8.2.1 (with $\sigma\Theta$ instead of Θ). We obtain

$$(48) \quad |\beta^\sigma| \leq q^{E(m)} \binom{2m}{m+1} (1 - |x|^{-1})^{-2m-1} |x|^{-1} = A|x|^{-1}.$$

Now recall that $|x| \geq q^{p-1}$ by Corollary 2.4, and, in particular, $|x| \geq 49$ because $q \geq 7$. Using (43), estimating $\binom{2m}{m+1} \leq 2^{2m}$ and using that $|x| \geq 49$, we obtain $A < q^{mq/(q-1)} 2.05^{2m}$.

Further, the assumption $w(\Theta) \leq q(p-1)/2$ implies that $m \leq (p-1)/2$, and we obtain $A < (2.05q^{7/12})^{p-1} < q^{p-1}$ (we again use the assumption $q \geq 7$). Thus, $A < |x|$, which implies that $|\beta^\sigma| < 1$ for all $\sigma \in G$. Since β is an algebraic integer, this is only possible if $\beta = 0$.

8.3.4. *Finishing the proof.* Thus, $P(x) = q^{E(m)}x^m(1 - \zeta/x)^{\Theta/q}$. Since $x^m(1 - \zeta/x)^{\Theta/q}$ is an algebraic integer, (45) implies that

$$q^{E(m)}\alpha_m(\Theta) \equiv 0 \pmod{q}.$$

By Proposition 8.2.2, this is possible only if $q \mid (\sum_{\sigma \in G} n_\sigma \zeta^\sigma)^m$. Since q is unramified in K , this implies that $q \mid \sum_{\sigma \in G} n_\sigma \zeta^\sigma$, that is, $q \mid n_\sigma$ for all $\sigma \in G$. Thus, $q \mid \Theta$, and this completes the proof of Theorem 6.3.3'.

9. PROOF OF THEOREM 6.3.4

To begin with, introduce the polynomial

$$(49) \quad f(T) = ((1+T)^q - 1 - T^q)/q \in \mathbf{z}[T].$$

It is a non-zero monic polynomial of degree $q-1$.

Now assume that all cyclotomic units of K are q -primary. In particular, so is $1 + \zeta^q = (1 - \zeta^{2q})/(1 - \zeta^q)$. Thus, there exists $\nu \in \mathbf{Z}[\zeta]$ such that $1 + \zeta^q \equiv \nu^q \pmod{q^2}$. Then $(1 + \zeta)^q \equiv 1 + \zeta^q \equiv \nu^q \pmod{q}$. Proposition 5.3.1 implies that $(1 + \zeta)^q \equiv \nu^q \pmod{q^2}$.

Thus, $(1 + \zeta)^q \equiv 1 + \zeta^q \pmod{q^2}$. This can be rewritten as $f(\zeta) \equiv 0 \pmod{q}$, where $f(T)$ is the polynomial defined in (49).

Applying the Galois conjugation, we obtain $f(\zeta^\sigma) \equiv 0 \pmod{q}$ for any $\sigma \in G$. Let now \mathfrak{q} be a prime ideal of K dividing q . Then we have $p - 1$ congruences

$$(50) \quad f(\zeta^\sigma) \equiv 0 \pmod{\mathfrak{q}} \quad (\sigma \in G).$$

Since $\zeta^\sigma \not\equiv \zeta^\tau \pmod{\mathfrak{q}}$ for distinct $\sigma, \tau \in G$, congruences (50) imply that $p - 1 \leq \deg f = q - 1$, which contradicts our assumption $p > q$. The theorem is proved.

REFERENCES

- [1] M.F. ATIYAH, I.G. MACDONALD, *Introduction to Commutative Algebra*, Addison-Wesley, 1969.
- [2] A. BAKER, Linear forms in the logarithms of algebraic numbers I–IV *Mathematika* **13** (1966), 204–216; **14** (1967), 102–107; 220–224; IV, *ibid.* **15** (1968), 204–216.
- [3] A. BAKER, Bounds for solutions of hyperelliptic equations, *Proc. Cambridge Phil. Soc.* **65** (1969), 439–444.
- [4] A. BAKER, G. WÜSTHOLZ, Logarithmic forms and group varieties, *J. reine angew. Math.* **442** (1993), 19–62.
- [5] C.D. BENNETT, J. BLASS, A.M.W. GLASS, D.B. MERONK, R.P. STEINER, Linear forms in the logarithms of three positive rational numbers, *J. Th. Nombres Bordeaux* **9** (1997), 97–136.
- [6] J. BLASS, A.M.W. GLASS, T.W. O’NEIL, Catalan’s conjecture and linear forms in logarithms, *Ulam Q. J.*, accepted, but never appeared in print.
- [7] Y. BUGEAUD, G. HANROT, Un nouveau critère pour l’équation de Catalan, *Mathematika* **47** (2000), 63–73.
- [8] J.W.S. CASSELS, On the equation $a^x - b^y = 1$, II, *Proc. Cambridge Society* **56** (1960), 97–103.
- [9] E. CATALAN, Note extraite d’une lettre adressée à l’éditeur, *J. reine angew. Math.* **27** (1844), 192.
- [10] A.O. GELFOND, *Transcendent and Algebraic Numbers* (Russian), Moscow 1952; English trans.: New York, Dover, 1960.
- [11] S. HYYRÖ, Über das Catalan’sche Problem, *Ann. Univ. Turku Ser. AI* **79** (1964), 3–10.
- [12] K. INKERI, On Catalan’s problem, *Acta Arith.* **9** (1964), 285–290.
- [13] K. INKERI, On Catalan’s conjecture, *J. Number Th.* **34** (1990), 142–152.

- [14] KO CHAO, On the diophantine equation $x^2 = y^n + 1$, $xy \neq 0$, *Sci. Sinica* **14** (1965), 457–460.
- [15] M. LANGEVIN, Quelques applications de nouveaux résultats de Van der Poorten, *Sém. Delange-Pisot-Poitou* **17** (1975/76), Paris, 1977.
- [16] M. LAURENT, M. MIGNOTTE, Y. NESTERENKO, Formes linéaires en deux logarithmes et déterminants d'interpolation, *J. Number Theory* **55** (1995), 285–321.
- [17] V.A. LEBESGUE, Sur l'impossibilité en nombres entiers de l'équation $x^m = y^2 + 1$, *Nouv. Ann. Math.* **9** (1850), 178–181.
- [18] E. MATVEEV, An explicit lower bound for a homogeneous rational linear form in logarithms of algebraic numbers I,II (Russian), I *Izv. Ross. Akad. Nauk Ser. Mat.* **62** (1998), 81–136; **64** (2000), 125–180; English trans.: *Izv. Math.* **62** (1998), 723–772; **64** (2000), 1217–1269.
- [19] T. METSÄNKYLÄ, Catalan's equation with a quadratic exponent, *C. R. Math. Acad. Sci. Soc. R. Can.* **23** (2001), 28–32.
- [20] M. MIGNOTTE, Un critère élémentaire pour l'équation de Catalan, *C. R. Math. Rep. Acad. Sci. Canada*, **15** (1993), 199–200.
- [21] M. MIGNOTTE, A criterion on Catalan's equation, *J. Number Th.* **52** (1995), 280–284.
- [22] M. MIGNOTTE, Catalan's equation just before 2000, *Number theory (Turku, 1999)*, de Gruyter, Berlin, 2001, pp. 247–254.
- [23] M. MIGNOTTE, Y. ROY, Catalan's equation has no new solutions with either exponent less than 10651, *Experimental Math.* **4** (1995), 259–268.
- [24] M. MIGNOTTE, Y. ROY, Minorations pour l'équation de Catalan, *C. R. Acad. Sci. Paris* **324** (1997), 377–380.
- [25] P. MIHĂILESCU, A class number free criterion for Catalan's conjecture, *J. Number Th.*, to appear.
- [26] P. MIHĂILESCU, Primary cyclotomic units and a proof of Catalan's conjecture, submitted.
- [27] T.W. O'NEIL, Improved upper bounds on the exponents in Catalan's equation, a manuscript (1995).
- [28] J.-C. PUCHTA, On a criterion for Catalan's conjecture, *Ramanujan J.* **5** (2001), 405–407.
- [29] P. RIBENBOIM, *Catalan's Conjecture*, Acad. Press, Boston, 1994.
- [30] W. SCHWARZ, A note on Catalan's equation, *Acta Arith.* **72** (1995), 277–279.
- [31] F. THAINE, On the ideal class groups of real abelian number fields, *Ann. of Math.* **128** (1988), 1–18.
- [32] R. TIJDEMAN, On the equation of Catalan, *Acta Arith.* **29** (1976), 197–209.
- [33] M. WALDSCHMIDT, Minorations de combinaisons linéaires de logarithmes de nombres algébriques, *Canadian J. Math.* **45** (1993), 176–224.

- [34] L. WASHINGTON, *Introduction to cyclotomic fields*, second edition, Graduate Texts in Math. **83**, Springer, New York, 1997.
- [35] G. WÜSTHOLZ (ed.), *A Panorama of Number Theory or The View from Bakers Garden*, Cambridge University Press 2002.

Yuri F. BILU

A2X, Université Bordeaux 1

351, cours de la Libération,

F-33405 TALENCE Cedex

E-mail : `yuri@math.u-bordeaux.fr`

IRRATIONALITÉ DE VALEURS DE ZÊTA

[d'après Apéry, Rivoal, ...]

par Stéphane FISCHLER

INTRODUCTION

Cet exposé est consacré aux valeurs aux entiers $s \geq 2$ de la fonction zêta de Riemann, définie par $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$. Quand $s = 2k$ est pair, on sait que $\zeta(2k)\pi^{-2k}$ est un nombre rationnel, lié aux nombres de Bernoulli. Comme π est transcendant (voir l'appendice de [La] pour une preuve), $\zeta(2k)$ l'est aussi pour tout $k \geq 1$. La nature arithmétique des $\zeta(2k + 1)$ est beaucoup moins bien connue. D'un point de vue conjectural, la situation est simple :

CONJECTURE 0.1. — *Les nombres π , $\zeta(3)$, $\zeta(5)$, $\zeta(7)$, ... sont algébriquement indépendants sur \mathbb{Q} .*

Cette conjecture est un cas particulier d'une conjecture diophantienne sur les polyzêtas (voir [Wa] ou [Ca2]). Elle implique que les $\zeta(2k + 1)$ sont tous transcendants, donc irrationnels, et linéairement indépendants sur \mathbb{Q} .

Très peu de résultats sont connus en direction de la conjecture 0.1. Le premier d'entre eux a été annoncé par Apéry lors des Journées Arithmétiques de Luminy, en 1978 :

THÉORÈME 0.2 ([Ap1]). — *$\zeta(3)$ est irrationnel.*

Apéry lui-même n'a donné lors de son exposé (voir [Me]), et n'a publié [Ap1], qu'une esquisse de sa preuve. Les détails (qui sont loin d'être triviaux) ont été publiés par Van Der Poorten [Po1] (voir aussi [Coh1] et [Re1]), grce à des contributions de Cohen et Zagier. Par la suite, plusieurs autres démonstrations du théorème d'Apéry sont parues. La première partie de ce texte est consacrée à une synthèse des différents points de vue qu'on peut adopter pour le démontrer.

La grande percée suivante date de 2000 :

THÉORÈME 0.3 ([Ri1], [BR]). — *Le \mathbb{Q} -espace vectoriel engendré par 1 , $\zeta(3)$, $\zeta(5)$, $\zeta(7)$, ... est de dimension infinie.*

En conséquence, il existe une infinité de k tels que $\zeta(2k + 1)$ soit irrationnel. On peut donner des versions effectives de ce dernier énoncé : Rivoal a démontré [Ri3] que parmi

les neuf nombres $\zeta(5), \zeta(7), \dots, \zeta(21)$, l'un au moins est irrationnel. Ce résultat a été amélioré par Zudilin :

THÉORÈME 0.4 ([Zu1], [Zu4]). — *L'un au moins des quatre nombres $\zeta(5), \zeta(7), \zeta(9), \zeta(11)$ est irrationnel.*

Malgré ces développements récents, il n'existe aucun entier $s \geq 5$ impair pour lequel on sache si $\zeta(s)$ est rationnel ou non.

Ce texte est divisé en trois parties. La première est une synthèse des méthodes connues pour démontrer l'irrationalité de $\zeta(3)$; l'intérêt des différentes approches est qu'elles se généralisent plus ou moins facilement à d'autres situations. La deuxième partie fournit une preuve du théorème 0.3, et de résultats voisins. La troisième est consacrée à des résultats "quantitatifs" : mesure d'irrationalité de $\zeta(3)$ et théorème 0.4.

REMERCIEMENTS : Je remercie toutes les personnes qui m'ont aidé dans la préparation de ce texte, notamment F. Amoroso, V. Bosser, N. Brisebarre, P. Cartier, G. Christol, P. Colmez, P. Grinspan, L. Habsieger, M. Huttner, C. Krattenthaler, C. Maclean, F. Martin, Yu. Nesterenko, F. Pellarin, A. Pulita, E. Royer, M. Waldschmidt, D. Zagier et W. Zudilin. Je remercie tout particulièrement T. Rivoal pour les nombreuses discussions très instructives que nous avons eues.

1. IRRATIONALITÉ DE $\zeta(3)$

Toutes les preuves connues de l'irrationalité de $\zeta(3)$ ont la même structure. On construit, pour tout $n \geq 0$, des nombres rationnels u_n et v_n ayant les propriétés suivantes :

(1) La forme linéaire $I_n = u_n \zeta(3) - v_n$ vérifie

$$\limsup_{n \rightarrow \infty} |I_n|^{1/n} \leq (\sqrt{2} - 1)^4 = 0,0294372 \dots$$

(2) En notant d_n le p.p.c.m. des entiers compris entre 1 et n , les coefficients u_n et v_n vérifient :

$$u_n \in \mathbf{Z} \text{ et } 2d_n^3 v_n \in \mathbf{Z}.$$

(3) Pour une infinité d'entiers n , on a $I_n \neq 0$.

La conclusion est alors immédiate : si $\zeta(3)$ était un nombre rationnel p/q , alors $2qd_n^3 I_n$ serait un entier pour tout n , et tendrait vers zéro quand n tend vers l'infini (car $(\sqrt{2} - 1)^4 e^3 < 1$, en utilisant [Ing] le théorème des nombres premiers sous la forme $\lim_{n \rightarrow \infty} \frac{\log(d_n)}{n} = 1$) : cela contredit la troisième assertion.

REMARQUE 1.1. — Comme $(\sqrt{2} - 1)^4 \cdot 3 \cdot 23^3 < 1$, le théorème des nombres premiers peut être remplacé par l'assertion plus faible $d_n < 3 \cdot 23^n$ pour n assez grand, qui se démontre en utilisant des arguments élémentaires à la Tchebychev ([NZM], §8.1 ; [Ing], p. 15).

Dans la suite, on donne plusieurs constructions (§ 1.1 à 1.10) de u_n , v_n et I_n , à chaque fois notées $u_{i,n}$, $v_{i,n}$ et $I_{i,n}$ (l'indice $i \in \{\mathbf{R}, \mathbf{E}, \mathbf{R}, \Sigma, \mathbf{C}, \mathbf{P}, \mathbf{TB}, \mathbf{M}\}$ fait référence à la construction utilisée). En fait, on construit toujours les mêmes formes linéaires : *a posteriori* on s'aperçoit que $u_{i,n}$, $v_{i,n}$ et $I_{i,n}$ ne dépendent pas de i . La preuve de cette indépendance est le plus souvent directe. Parfois, on montre simplement que $I_{i,n} = I_{j,n}$; les deux autres égalités en découlent en utilisant l'irrationalité de $\zeta(3)$.

Les premières valeurs de u_n et v_n sont :

$$\begin{aligned}(u_n)_{n \geq 0} &= 1, 5, 73, 1445, 33001, 819005, \dots \\ (v_n)_{n \geq 0} &= 0, 6, \frac{351}{4}, \frac{62531}{36}, \frac{11424695}{288}, \dots\end{aligned}$$

Cette partie contient l'esquisse de plusieurs preuves de l'irrationalité de $\zeta(3)$, notamment celles d'Apéry [Ap1] (§1.1 et 1.2), de Beukers [Be1] par les intégrales multiples (§1.3) ou [Be6] par les formes modulaires (§1.10), de Prevost [Pr1] (§1.1 et 1.2), de Nesterenko [Ne2] (§1.4 et 1.5), de Sorokin [So3] (§1.8), et de nombreuses variantes. Certaines preuves sont obtenues en montrant que deux constructions différentes fournissent les mêmes formes linéaires, puis en prouvant le point (2) à l'aide de l'une et les points (1) et (3) à l'aide de l'autre (par exemple en montrant que $\lim_{n \rightarrow \infty} |I_n|^{1/n} = (\sqrt{2} - 1)^4$).

La plupart des méthodes connues pour démontrer des résultats d'irrationalité sur les valeurs de ζ sont liées aux polylogarithmes, définis pour tout entier $k \geq 1$ par :

$$\operatorname{Li}_k(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^k},$$

avec $|z| < 1$ si $k = 1$ et $|z| \leq 1$ si $k \geq 2$. L'idée est de construire des formes linéaires en polylogarithmes, à coefficients polynomiaux, puis de spécialiser en $z = 1$. C'est la méthode employée dans les paragraphes 1.3 à 1.9. Les formes linéaires en polylogarithmes $I_{i,n}(z)$ qu'on utilise ne sont pas toujours les mêmes, mais elles coïncident en $z = 1$, pour donner les formes linéaires d'Apéry.

Les polylogarithmes s'insèrent dans la famille des séries hypergéométriques ${}_qF_q$ (avec $q \geq 1$), définies par :

$${}_qF_q \left(\begin{matrix} \alpha_0, & \alpha_1, & \dots, & \alpha_q \\ & \beta_1, & \dots, & \beta_q \end{matrix} \middle| z \right) = \sum_{k=0}^{\infty} \frac{(\alpha_0)_k (\alpha_1)_k \cdots (\alpha_q)_k}{k! (\beta_1)_k \cdots (\beta_q)_k} z^k,$$

où le symbole de Pochhammer est $(\alpha)_k = \alpha(\alpha + 1) \cdots (\alpha + k - 1)$. Dans cet exposé, les α_j et les β_j seront des entiers, les β_j étant positifs, et z sera un nombre complexe avec $|z| \leq 1$. On adopte les définitions suivantes ([AAR], §3.3 et 3.4) :

- ${}_qF_q$ est dite *bien équilibrée* si $\alpha_0 + 1 = \alpha_1 + \beta_1 = \cdots = \alpha_q + \beta_q$;
- ${}_qF_q$ est dite *très bien équilibrée* si elle est bien équilibrée et $\alpha_1 = \frac{1}{2}\alpha_0 + 1$.

1.1. Récurrence linéaire

DÉFINITION 1.2. — Soient $(u_{R,n})_{n \geq 0}$ et $(v_{R,n})_{n \geq 0}$ les suites définies par la relation de récurrence

$$(1) \quad (n+1)^3 y_{n+1} - (34n^3 + 51n^2 + 27n + 5)y_n + n^3 y_{n-1} = 0$$

et les conditions initiales

$$u_{R,0} = 1, \quad u_{R,1} = 5, \quad v_{R,0} = 0, \quad v_{R,1} = 6.$$

Une récurrence immédiate montre que les suites $(u_{R,n})$ et $(v_{R,n})$ sont croissantes et à termes rationnels, avec $n!^3 u_{R,n} \in \mathbf{Z}$ et $n!^3 v_{R,n} \in \mathbf{Z}$. En fait on verra qu'on peut remplacer $n!^3$ par d_n^3 .

Les propriétés asymptotiques des suites vérifiant la récurrence (1) sont faciles à déterminer (voir par exemple [Gel], Chapitre 5). L'équation caractéristique associée est $X^2 - 34X + 1$; elle a deux racines simples, $(\sqrt{2} + 1)^4$ et $(\sqrt{2} - 1)^4$. L'espace vectoriel des solutions de (1) est de dimension deux, et admet une base formée de suites $(y_n^{(0)})_{n \geq 0}$ et $(y_n^{(1)})_{n \geq 0}$ avec $\lim_{n \rightarrow +\infty} \frac{\log |y_n^{(0)}|}{n} = \log((\sqrt{2} + 1)^4)$ et $\lim_{n \rightarrow +\infty} \frac{\log |y_n^{(1)}|}{n} = \log((\sqrt{2} - 1)^4)$. La suite $(y_n^{(1)})$ est uniquement déterminée (à proportionnalité près) par son comportement asymptotique ; toutes les autres solutions de (1) se comportent comme $(y_n^{(0)})$. Comme $(u_{R,n})$ et $(v_{R,n})$ sont croissantes, on a :

$$(2) \quad \lim_{n \rightarrow \infty} u_{R,n}^{1/n} = \lim_{n \rightarrow \infty} v_{R,n}^{1/n} = (\sqrt{2} + 1)^4 = 33,9705627 \dots$$

Quand on adopte ce point de vue, on a intérêt [Po1] à considérer $\Delta_n = \begin{vmatrix} v_{R,n} & v_{R,n-1} \\ u_{R,n} & u_{R,n-1} \end{vmatrix}$ pour $n \geq 1$. La relation de récurrence montre qu'on a $\Delta_n = \frac{6}{n^3}$ pour tout n , ce qui signifie $\frac{v_{R,n}}{u_{R,n}} - \frac{v_{R,n-1}}{u_{R,n-1}} = \frac{6}{n^3 u_{R,n} u_{R,n-1}}$. Donc la suite $(\frac{v_{R,n}}{u_{R,n}})$ est strictement croissante et tend vers une limite finie ℓ , avec $u_{R,n} \ell - v_{R,n} = \sum_{k=n+1}^{\infty} \frac{6 u_{R,n}}{k^3 u_{R,k} u_{R,k-1}}$. Ceci prouve que $u_{R,n} \ell - v_{R,n}$ est une solution de (1) qui tend vers zéro quand n tend vers l'infini : son comportement asymptotique est nécessairement donné par

$$\lim_{n \rightarrow +\infty} \frac{\log |u_{R,n} \ell - v_{R,n}|}{n} = \log((\sqrt{2} - 1)^4).$$

Avec cette définition de $u_{R,n}$ et $v_{R,n}$, il n'est pas évident de démontrer que $\ell = \zeta(3)$, et de borner par d_n^3 les dénominateurs de $u_{R,n}$ et $v_{R,n}$. Pour ceci, une possibilité est de faire le lien avec le paragraphe 1.2 : c'est la méthode employée dans les premières preuves détaillées de l'irrationalité de $\zeta(3)$, qui sont parues peu après l'exposé d'Apéry ([Re1], [Po1], [Coh1]).

REMARQUE 1.3. — Le raisonnement ci-dessus montre que $\frac{v_{R,n}}{u_{R,n}}$ est la n -ième somme partielle de la série $\zeta(3) = \sum_{k=1}^{\infty} \frac{6}{k^3 u_{R,k} u_{R,k-1}}$.

La définition 1.2 s'interprète en termes de fractions continues généralisées. En effet, considérons la récurrence linéaire

$$(3) \quad Y_{n+1} - (34n^3 + 51n^2 + 27n + 5)Y_n + n^6Y_{n-1} = 0.$$

On passe d'une solution de (1) à une solution de (3), et réciproquement, en posant $Y_n = n!^3 y_n$. Si $U_{R,n}$ et $V_{R,n}$ sont ainsi associées à $u_{R,n}$ et $v_{R,n}$, alors $\frac{V_{R,n}}{U_{R,n}} = \frac{v_{R,n}}{u_{R,n}}$ est la n -ième réduite de la fraction continue généralisée

$$\zeta(3) = \cfrac{6}{5} - \cfrac{1}{117} - \cfrac{64}{535} - \dots - \cfrac{n^6}{34n^3 + 51n^2 + 27n + 5} - \dots.$$

On peut trouver cette formule grâce à un procédé général ([Ap2], [BO], [Ze2]) qui accélère la convergence d'un développement en fraction continue généralisée. Ce procédé s'applique, en particulier, au développement dont les réduites sont les sommes partielles de la série $\sum_{n=1}^{\infty} \frac{1}{f(n)}$, où f est un polynôme sans zéro parmi les entiers strictement positifs.

En utilisant cette méthode d'accélération de convergence, André-Jeannin a démontré [AnJ] que la somme des inverses des nombres de Fibonacci est irrationnelle (voir aussi [BV] et [Pr2]).

1.2. Formules explicites

DÉFINITION 1.4. — Soient $(u_{E,n})$ et $(v_{E,n})$ les suites définies par les formules suivantes :

$$\begin{aligned} u_{E,n} &= \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}^2 \\ v_{E,n} &= \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}^2 \left(\sum_{m=1}^n \frac{1}{m^3} + \sum_{m=1}^k \frac{(-1)^{m-1}}{2m^3 \binom{n}{m} \binom{n+m}{m}} \right) \end{aligned}$$

Sous cette forme, il est clair que $u_{E,n} \in \mathbf{Z}$ et que $\frac{v_{E,n}}{u_{E,n}}$ tend vers $\zeta(3)$. Pour démontrer ([Po1], [Coh1], [Re1]) que $2d_n^3 v_{E,n} \in \mathbf{Z}$, il suffit de démontrer que, pour $1 \leq m \leq k \leq n$,

$$(4) \quad \frac{\binom{n+k}{k} d_n^3}{m^3 \binom{n}{m} \binom{n+m}{m}} = \frac{\binom{n+k}{k-m} d_n^3}{m^3 \binom{n}{m} \binom{k}{m}}$$

est entier. Soit p un nombre premier ; la valuation p -adique $v_p(n!)$ de $n!$ vaut $\sum_{i=1}^{\alpha} [\frac{n}{p^i}]$ avec $\alpha = [\frac{\log(n)}{\log(p)}] = v_p(d_n)$. Pour $1 \leq i \leq v_p(m)$ on a $[\frac{n}{p^i}] = [\frac{n-m}{p^i}] + [\frac{m}{p^i}]$ et pour $v_p(m) < i \leq v_p(d_n)$ on a $[\frac{n}{p^i}] \leq [\frac{n-m}{p^i}] + [\frac{m}{p^i}] + 1$. On en déduit $v_p(\binom{n}{m}) \leq v_p(d_n) - v_p(m)$ et $v_p(\binom{k}{m}) \leq v_p(d_k) - v_p(m)$. Il en résulte que $\frac{d_n^3}{m^3 \binom{n}{m} \binom{k}{m}}$ est un entier, et le quotient (4) aussi.

Montrons maintenant ([Po1], [Coh1]) que les suites $(u_{E,n})$ et $(v_{E,n})$ vérifient la récurrence (1). On pose $\lambda_{n,k} = \binom{n}{k}^2 \binom{n+k}{k}^2$ pour $k, n \in \mathbf{Z}$, et

$$\mathbf{A}_{n,k} = 4(2n+1)(k(2k+1) - (2n+1)^2)\lambda_{n,k},$$

avec les conventions habituelles (i.e. $\lambda_{n,k} = 0$ si $k < 0$ ou $k > n$). On a alors

$$\mathbf{A}_{n,k} - \mathbf{A}_{n,k-1} = (n+1)^3 \lambda_{n+1,k} - (34n^3 + 51n^2 + 27n + 5) \lambda_{n,k} + n^3 \lambda_{n-1,k}.$$

En sommant sur k , on obtient que la suite $(u_{E,n})$ satisfait à la récurrence (1). Pour la suite $(v_{E,n})$, on peut faire de même en utilisant la suite double

$$\mathbf{B}_{n,k} = \mathbf{A}_{n,k} \left(\sum_{m=1}^n \frac{1}{m^3} + \sum_{m=1}^k \frac{(-1)^{m-1}}{2m^3 \binom{n}{m} \binom{n+m}{m}} \right) + \frac{5(2n+1)k(-1)^{k-1}}{n(n+1)} \binom{n}{k} \binom{n+k}{k}.$$

Ceci démontre qu'on a $u_{E,n} = u_{R,n}$ et $v_{E,n} = v_{R,n}$ pour tout $n \geq 0$. Compte tenu des résultats démontrés au paragraphe 1.1, on obtient une preuve de l'irrationalité de $\zeta(3)$.

La démonstration donnée ci-dessus que $(u_{E,n})$ et $(v_{E,n})$ vérifient la récurrence (1) n'est qu'une simple vérification, à condition d'être capable d'exhiber les suites doubles $\mathbf{A}_{n,k}$ et $\mathbf{B}_{n,k}$, ce qui n'a pas été une tâche facile (voir [Po1], §7). Motivés par ce problème, plusieurs auteurs (notamment Zeilberger) ont ensuite mis au point des algorithmes permettant d'exhiber de telles suites doubles. On a ainsi un moyen automatique de produire des preuves d'identités (voir [Ca1], [Ze1], [PWZ]). De plus, ces preuves sont immédiatement vérifiables à la main.

Dans les formules ci-dessus, un rôle central est joué par la suite double $c_{n,k} = \sum_{m=1}^n \frac{1}{m^3} + \sum_{m=1}^k \frac{(-1)^{m-1}}{2m^3 \binom{n}{m} \binom{n+m}{m}}$ (définie pour $0 \leq k \leq n$). Elle tend vers $\zeta(3)$ quand n tend vers l'infini, uniformément en k . On a $c_{n,n} - c_{n-1,n-1} = \frac{5}{2} \frac{(-1)^{n-1}}{n^3 \binom{2n}{n}}$ et $\lim_{n \rightarrow \infty} c_{n,n} = \zeta(3)$ donc :

$$(5) \quad \zeta(3) = \frac{5}{2} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^3 \binom{2n}{n}}.$$

Cette série n'est pas utilisée dans la preuve de l'irrationalité de $\zeta(3)$, mais elle a un intérêt non négligeable puisque les $c_{n,k}$ sont au cœur des formules explicites définissant $u_{E,n}$ et $v_{E,n}$. C'est pourquoi plusieurs auteurs ont cherché des généralisations de (5) (voir par exemple [Po1], [Po3], [Coh2], [Ko], [Le], [BB], [AG]), parmi lesquelles $\zeta(5) = \frac{5}{2} \sum_{n \geq 1} \frac{(-1)^n}{n^3 \binom{2n}{n}} \left(\sum_{j=1}^{n-1} \frac{1}{j^2} - \frac{4}{5n^2} \right)$. Mais aucune de ces généralisations n'a permis d'obtenir de nouveau résultat d'irrationalité : la croissance des dénominateurs est trop rapide par rapport à la convergence.

Prévost a montré [Pr1] comment interpréter les formules explicites données dans ce paragraphe en termes d'approximants de Padé. Posons $\varphi(x) = \sum_{k \geq 1} \frac{1}{(k+x)^3}$, c'est-à-dire $\zeta(3, 1+x)$ où ζ est la fonction zêta d'Hurwitz (voir [WW], Chapitre XIII). Pour tout

$n \geq 1$, considérons les polynômes suivants :

$$P_n(x) = \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} \binom{x}{k} \binom{x+k}{k} = {}_4F_3 \left(\begin{matrix} -n, & -x, & n+1, & x+1 \\ & 1, & 1, & 1 \end{matrix} \middle| 1 \right)$$

et $Q_n(x) = \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} \binom{x}{k} \binom{x+k}{k} \sum_{m=1}^k \frac{(-1)^{m-1}}{2m^3 \binom{x}{m} \binom{x+m}{m}}$.

Alors P_n est de degré $2n$, Q_n de degré $2n-2$, et on a $P_n(x)\varphi(x) - Q_n(x) = O(x^{-2n-1})$ quand x tend vers l'infini. Cela signifie que P_n et Q_n sont des approximants de Padé de la fonction φ . Quand x est un entier n , on a $\varphi(n) = \zeta(3) - \sum_{m=1}^n \frac{1}{m^3}$ d'où $P_n(n)\varphi(n) - Q_n(n) = u_{E,n}\zeta(3) - v_{E,n}$. On peut en déduire [Pr1] la majoration $|u_{E,n}\zeta(3) - v_{E,n}| \leq \frac{4\pi^2}{(2n+1)^2 u_{E,n}}$. Pour conclure, on a besoin d'une minoration asymptotique de $u_{E,n}$ comme celle de la formule (2). Il suffit donc de vérifier que $u_{E,n}$ satisfait à la récurrence (1). On peut utiliser $\mathbf{A}_{n,k}$ comme ci-dessous ; une autre méthode [AW] est d'utiliser des relations de contiguïté entre séries hypergéométriques balancées.

En effet, $u_{E,n}$ s'écrit ${}_4F_3 \left(\begin{matrix} -n, & -n, & n+1, & n+1 \\ & 1, & 1, & 1 \end{matrix} \middle| 1 \right)$. Une série hypergéométrique ${}_4F_3 \left(\begin{matrix} \alpha_0, & \alpha_1, & \alpha_2, & \alpha_3 \\ \beta_1, & \beta_2, & \beta_3 \end{matrix} \middle| z \right)$ est dite ([Sl], §2.1.1) *balancée* (ou *Saalschütziennne*) si $1 + \sum_{i=0}^3 \alpha_i = \sum_{j=1}^3 \beta_j$. Si on modifie deux des sept paramètres d'une série balancée, en ajoutant ou en retranchant 1 à chacun des deux, on peut obtenir à nouveau une série balancée. Si c'est le cas, on dit que ces deux séries sont *contiguës*. Il y a $2 \cdot \binom{7}{2} = 42$ séries balancées qui sont contiguës à une série balancée donnée. Quand α_0 est un entier négatif (ce qui signifie que la série hypergéométrique est en fait un polynôme), il existe des relations linéaires entre les valeurs en 1 de ces 42 séries, dont les coefficients sont des polynômes en les paramètres α_0, \dots, β_3 (voir [AAR], §3.7). On peut [AW] déduire de ces relations de contiguïté que la suite $u_{E,n}$ vérifie la récurrence (1).

1.3. Intégrale triple réelle

Considérons l'intégrale suivante, qui a été introduite par Beukers [Be1] (voir aussi [Be3]) :

$$I_{\mathcal{R},n}(z) = \int_0^1 \int_0^1 \int_0^1 \frac{u^n(1-u)^n v^n(1-v)^n w^n(1-w)^n}{((1-w)z + uvw)^{n+1}} du dv dw.$$

Cette intégrale converge pour tout $z \in \mathbf{C} \setminus]-\infty, 0]$. Voici une esquisse de preuve de l'irrationalité de $\zeta(3)$ qui utilise $I_{\mathcal{R},n}(1)$. Les détails se trouvent dans [Be1].

Comme le maximum de la fonction $\frac{u(1-u)v(1-v)w(1-w)}{1-w(1-uv)}$ sur le cube unité vaut $(\sqrt{2}-1)^4$, on a :

$$\lim_{n \rightarrow +\infty} \frac{\log(I_{\mathcal{R},n}(1))}{n} = \log((\sqrt{2}-1)^4).$$

Par ailleurs, si on intègre n fois par parties par rapport à v , qu'on change w en $\frac{1-w}{1-w(1-uv)}$, et enfin qu'on intègre n fois par parties par rapport à u , on obtient :

$$I_{\mathcal{R},n}(1) = \int_0^1 \int_0^1 \int_0^1 \frac{P_n(u)P_n(v)}{1-w(1-uv)} du dv dw,$$

où $P_n(X) = \frac{1}{n!}(X^n(1-X)^n)^{(n)}$ est le n -ième polynôme de Legendre. En intégrant par rapport à w , il vient $I_{\mathcal{R},n}(1) = \int_0^1 \int_0^1 \frac{-\log(uv)}{1-uv} P_n(u)P_n(v) du dv$. Or pour tous $k, l \in \{0, \dots, n\}$ on peut écrire $\int_0^1 \int_0^1 \frac{-\log(uv)}{1-uv} u^k v^l du dv = 2a_{k,l}\zeta(3) + b_{k,l}$ avec $a_{k,l} \in \mathbf{Z}$ et $d_n^3 b_{k,l} \in \mathbf{Z}$. On a donc :

$$I_{\mathcal{R},n}(1) = 2(u_{\mathcal{R},n}\zeta(3) - v_{\mathcal{R},n}) \text{ avec } u_{\mathcal{R},n} \in \mathbf{Z} \text{ et } 2d_n^3 v_{\mathcal{R},n} \in \mathbf{Z}.$$

Cela termine la preuve de l'irrationalité de $\zeta(3)$.

1.4. Série de type hypergéométrique

Posons

$$(6) \quad R_n(X) = \frac{(X-1)^2 \dots (X-n)^2}{X^2(X+1)^2 \dots (X+n)^2} = \frac{(X-n)_n^2}{(X)_{n+1}^2} = \frac{\Gamma(X)^4}{\Gamma(X-n)^2 \Gamma(X+n+1)^2},$$

où Γ est la fonction Gamma d'Euler, qui vérifie $\Gamma(s+1) = s\Gamma(s)$. En outre, pour $|z| \geq 1$ on pose :

$$(7) \quad I_{\Sigma,n}(z) = - \sum_{k=1}^{\infty} R'_n(k) z^{-k}.$$

En suivant [Be2], [Gu2] et [Ne2] on développe la fraction rationnelle R_n en éléments simples :

$$(8) \quad R_n(X) = \sum_{i=0}^n \left(\frac{\alpha_i}{(X+i)^2} + \frac{\beta_i}{X+i} \right),$$

avec $\alpha_i = \binom{n}{i}^2 \binom{n+i}{i}^2$ et $\beta_i = 2(-1)^i \binom{n}{i} \binom{n+i}{i} \sum_{j \in \{0, \dots, n\}, j \neq i} \frac{(-1)^j \binom{n}{j} \binom{n+j}{j}}{j-i}$ pour $i \in \{0, \dots, n\}$ (ces formules s'obtiennent en remarquant que $R_n(X) = \left(\frac{(X-n)_n}{(X)_{n+1}}\right)^2$; voir la démonstration du lemme 2.12 ci-dessous, ou bien [Col], [Hab] ou [Zu5]). En utilisant (8) pour exprimer (7) il vient :

$$(9) \quad \begin{aligned} I_{\Sigma,n}(z) &= 2 \sum_{i=0}^n \alpha_i z^i \sum_{k \geq 1} \frac{z^{-(k+i)}}{(k+i)^3} + \sum_{i=0}^n \beta_i z^i \sum_{k \geq 1} \frac{z^{-(k+i)}}{(k+i)^2} \\ &= 2A_n(z)\text{Li}_3(1/z) + B_n(z)\text{Li}_2(1/z) + C_n(z) \end{aligned}$$

où les polynômes A_n , B_n et C_n sont définis par :

$$\begin{aligned} A_n(z) &= \sum_{i=0}^n \alpha_i z^i = {}_4F_3 \left(\begin{matrix} -n, & -n, & n+1, & n+1 \\ & 1, & 1, & 1 \end{matrix} \middle| z \right) \\ B_n(z) &= \sum_{i=0}^n \beta_i z^i \\ C_n(z) &= -\sum_{t=0}^{n-1} z^t \sum_{i=t+1}^n \left(\frac{2\alpha_i}{(i-t)^3} + \frac{\beta_i}{(i-t)^2} \right) \end{aligned}$$

Il est clair que les polynômes $A_n(z)$, $d_n B_n(z)$ et $d_n^3 C_n(z)$ sont à coefficients entiers. On a $B_n(1) = 0$ car R_n n'a pas de résidu à l'infini. En posant $u_{\Sigma,n} = A_n(1)$ et $v_{\Sigma,n} = -C_n(1)/2$ il vient :

$$(10) \quad I_{\Sigma,n}(1) = 2(u_{\Sigma,n} \zeta(3) - v_{\Sigma,n}) \text{ avec } u_{\Sigma,n} \in \mathbf{Z} \text{ et } 2d_n^3 v_{\Sigma,n} \in \mathbf{Z}.$$

Pour démontrer l'irrationalité de $\zeta(3)$, il ne reste plus qu'à estimer $I_{\Sigma,n}(1)$. On peut le faire en transformant $I_{\Sigma,n}(1)$ en une intégrale complexe (voir le paragraphe 1.5) ; c'est ainsi que Nesterenko démontre [Ne2] le théorème d'Apéry.

On peut démontrer, en utilisant [Zu5] l'algorithme de "creative telescoping" ([PWZ], Chapitre 6), que $I_{\Sigma,n}(1)$, $u_{\Sigma,n}$ et $v_{\Sigma,n}$ satisfont à la relation de récurrence (1). Cela démontre en particulier l'identité $v_{\Sigma,n} = v_{E,n}$.

1.5. Intégrale complexe

Soit c un réel, avec $0 < c < n+1$. Pour $z \neq 0$, choisissons une détermination de $\arg(z)$ strictement comprise entre -2π et 2π , et considérons l'intégrale suivante, le long de la droite verticale $\operatorname{Re}(s) = c$ dans \mathbf{C} , orientée de bas en haut :

$$\begin{aligned} I_{c,n}(z) &= \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \left(\frac{\pi}{\sin(\pi s)} \right)^2 R_n(s) z^{-s} ds \\ (11) \quad &= \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(n+1-s)^2 \Gamma(s)^4}{\Gamma(n+1+s)^2} z^{-s} ds, \end{aligned}$$

cette dernière égalité provenant directement de (6) et de la formule classique $\frac{\pi}{\sin(\pi s)} = \Gamma(s)\Gamma(1-s)$. La valeur de $I_{c,n}(z)$ ne dépend pas du choix de c d'après le théorème des résidus. L'intégrale (11) est un exemple de G -fonction de Meijer (voir [Lu], §5.2) :

$$I_{c,n}(z) = G_{4,4}^{4,2} \left(\begin{matrix} -n, & -n, & n+1, & n+1 \\ 0, & 0, & 0, & 0 \end{matrix} \middle| z \right).$$

La méthode du col (voir par exemple [Di], Chapitre IX) permet [Ne2] d'obtenir une estimation asymptotique très précise :

$$I_{c,n}(1) = \frac{\pi^{3/2} 2^{3/4}}{n^{3/2}} (\sqrt{2} - 1)^{4n+2} (1 + O(n^{-1})).$$

Quand on déplace le contour d'intégration vers la droite pour faire apparaître les pôles $n + 1, n + 2, \dots$, le théorème des résidus donne ([Gu1], [Gu2]), puisque $(\frac{\pi}{\sin(\pi s)})^2 = \frac{1}{(s-k)^2} + O(1)$ quand s tend vers un entier k :

$$(12) \quad I_{c,n}(z) = I_{\Sigma,n}(z) + \log(z) \sum_{k=1}^{\infty} R_n(k) z^{-k}.$$

En particulier pour $z = 1$ on obtient $I_{c,n}(1) = I_{\Sigma,n}(1)$.

Par ailleurs, Nesterenko a démontré [Ne3] un théorème général qui relie une intégrale multiple réelle à une intégrale complexe ; dans notre cas particulier, ce théorème donne $I_{c,n}(z) = I_{\mathcal{R},n}(z)$.

On peut démontrer [Ne2] que $I_{c,n}(1)$ vérifie la récurrence (1) en utilisant les relations de contiguïté sur les G -fonctions de Meijer. C'est en fait une preuve parallèle à celle du paragraphe 1.2, où on utilisait la contiguïté entre des ${}_4F_3$. En effet ([Lu], §5.8), ces ${}_4F_3$ satisfont aux mêmes équations différentielles que les G -fonctions de Meijer correspondantes, donc aux mêmes relations de contiguïté.

1.6. Un problème d'approximation de Padé

Considérons [Be2] le problème suivant : trouver quatre polynômes A_n, B_n, C_n et D_n , à coefficients rationnels, de degré au plus n , tels que :

$$(13) \quad \begin{cases} F_n(z) := A_n(z)\text{Li}_2(1/z) + B_n(z)\text{Li}_1(1/z) + D_n(z) = O(z^{-n-1}) \text{ quand } z \rightarrow \infty \\ G_n(z) := 2A_n(z)\text{Li}_3(1/z) + B_n(z)\text{Li}_2(1/z) + C_n(z) = O(z^{-n-1}) \text{ quand } z \rightarrow \infty \\ B_n(1) = 0 \end{cases}$$

Une solution à ce problème de Padé est donnée par les polynômes A_n, B_n et C_n du paragraphe 1.4 (et un polynôme D_n convenable). On a alors :

$$\begin{cases} F_n(z) = \sum_{k=1}^{\infty} R_n(k) z^{-k} = \frac{n!^4}{(2n+1)!^2} z^{-n-1} {}_4F_3 \left(\begin{matrix} n+1, & n+1, & n+1, & n+1 \\ & 2n+2, & 2n+2, & 1 \end{matrix} \middle| z^{-1} \right) \\ G_n(z) = I_{\Sigma,n}(z) = - \sum_{k=1}^{\infty} R'_n(k) z^{-k} \end{cases}$$

En effet, la seconde égalité est simplement une réécriture de (7) et (9). La première se démontre de manière analogue à (9), mais sans dériver (8).

L'équation différentielle hypergéométrique sous-jacente aux constructions des paragraphes 1.4 et 1.5 s'écrit $Ly = 0$, en posant

$$L = z(\delta + n + 1)^2(\delta - n)^2 - \delta^4 \text{ avec } \delta = z \frac{d}{dz}.$$

Elle admet au voisinage de l'infini quatre solutions linéairement indépendantes : $F_n(z)$, $I_{c,n}(z) = G_n(z) + F_n(z) \log(z)$, $A_n(z)$ et $B_n(z) - A_n(z) \log(z)$ (voir [Lu], §5.1 et 5.8, [Hu1] et [Gu2]). Ces solutions sont reliées par la monodromie : en prolongeant analytiquement

F_n le long d'un lacet qui entoure le point 1, on fait apparaître $B_n(z) + A_n(z) \log(1/z)$, puis en faisant le tour de l'infini, on obtient $A_n(z)$ (voir [Oe] pour la monodromie des polylogarithmes).

Ce point de vue permet de démontrer [Hu1] que le problème de Padé (13) a une solution unique (à proportionnalité près). En effet, en partant d'une solution A_n, B_n, C_n, D_n , on montre que F_n vérifie une équation différentielle linéaire fuchsienne d'ordre 4 qu'on détermine explicitement (en calculant ses exposants, et en utilisant la relation de Fuchs) : on trouve que c'est $Ly = 0$.

Pour démontrer l'unicité de la solution de ce problème de Padé, on peut aussi suivre [Be2]. On part d'une solution quelconque, avec des polynômes A_n, B_n, C_n, D_n et des fonctions F_n et G_n . On note α_i et β_i les coefficients de A_n et B_n , et on leur associe la fraction rationnelle R_n définie par (8). On voit alors que $F_n(z) = \sum_{k=1}^{\infty} R_n(k)z^{-k}$ et $G_n(z) = -\sum_{k=1}^{\infty} R'_n(k)z^{-k}$, donc les deux premières contraintes de (13) signifient que R_n et sa dérivée s'annulent aux points $1, 2, \dots, n$. En outre, le résidu à l'infini de R_n est alors $B_n(1) = 0$: la fraction rationnelle R_n est nécessairement donnée, à constante multiplicative près, par (6).

1.7. Polynômes orthogonaux

Considérons ([BE], [As]) le problème suivant : trouver deux polynômes \tilde{A}_n et \tilde{B}_n , de degré au plus n , tels que :

$$(14) \quad \begin{cases} \int_0^1 \left(\tilde{B}_n(x) - \tilde{A}_n(x) \log(x) \right) x^k dx = 0 \text{ pour tout } k \in \{0, \dots, n-1\} \\ \int_0^1 \left(\tilde{B}_n(x) - \tilde{A}_n(x) \log(x) \right) x^k \log(x) dx = 0 \text{ pour tout } k \in \{0, \dots, n-1\} \\ \tilde{B}_n(1) = 0 \end{cases}$$

Une solution à ce problème est donnée par les polynômes \tilde{A}_n et \tilde{B}_n définis par :

$$(15) \quad \tilde{B}_n(x) - \tilde{A}_n(x) \log(x) = \int_x^1 P_n\left(\frac{x}{t}\right) P_n(t) \frac{dt}{t},$$

où P_n est le n -ième polynôme de Legendre (comme au paragraphe 1.3). En effet, on a alors $\int_0^1 \left(\tilde{B}_n(x) - \tilde{A}_n(x) \log(x) \right) x^k dx = \left(\int_0^1 P_n(u) u^k du \right)^2$ en posant $u = \frac{x}{t}$. La première condition de (14) en découle immédiatement ; la deuxième s'obtient après dérivation par rapport à k .

Comme on a $\text{Li}_j(1/z) = \frac{(-1)^{j-1}}{(j-1)!} \int_0^1 \log^{j-1}(x) \frac{dx}{z-x}$ pour tout entier $j \geq 1$, il vient :

$$(16) \quad 2\tilde{A}_n(z) \text{Li}_3(1/z) + \tilde{B}_n(z) \text{Li}_2(1/z) = - \int_0^1 \left(\tilde{B}_n(z) - \tilde{A}_n(z) \log(x) \right) \frac{\log(x) dx}{z-x}.$$

On définit un polynôme $\tilde{C}_n(z)$ par :

$$\tilde{C}_n(z) = \int_0^1 \frac{\tilde{B}_n(z) - \tilde{B}_n(x)}{z-x} \log(x) dx - \int_0^1 \frac{\tilde{A}_n(z) - \tilde{A}_n(x)}{z-x} \log^2(x) dx.$$

Grâce à (15) on peut obtenir des formules explicites pour \tilde{A}_n , \tilde{B}_n et \tilde{C}_n ; on trouve les mêmes que pour A_n , B_n et C_n respectivement au paragraphe 1.4. Donc \tilde{A}_n , $d_n\tilde{B}_n$ et $d_n^3\tilde{C}_n$ sont à coefficients entiers. On obtient aussi ([BE], Corollaire A.2.3) que tous les zéros de $\tilde{A}_n(z)$ et de $\frac{\tilde{B}_n(z)}{z-1}$ sont réels négatifs, et entrelacés. Par ailleurs, on a :

$$2\tilde{A}_n(z)\text{Li}_3(1/z) + \tilde{B}_n(z)\text{Li}_2(1/z) + \tilde{C}_n(z) = - \int_0^1 \left(\tilde{B}_n(x) - \tilde{A}_n(x) \log(x) \right) \frac{\log(x)dx}{z-x}.$$

Quand $z = 1$, le membre de droite se transforme (en utilisant (15) et en posant $u = t$, $v = \frac{x}{t}$) en $I_{\mathcal{R},n}(1) = - \int_0^1 \int_0^1 \frac{\log(uv)}{1-uv} P_n(u)P_n(v)du dv$. En appliquant l'estimation asymptotique de $I_{\mathcal{R},n}(1)$ obtenue au paragraphe 1.3, on obtient une démonstration de l'irrationalité de $\zeta(3)$.

En fait un couple $(\tilde{A}_n, \tilde{B}_n)$ vérifie (14) si, et seulement si, il existe C_n et D_n tels que $(\tilde{A}_n, \tilde{B}_n, C_n, D_n)$ soit une solution du problème de Padé (13). Plus précisément, la première (resp. la deuxième) assertion de (13) équivaut à la première (resp. la deuxième) assertion de (14) (il s'agit d'un fait général : voir par exemple [NS], Chapitre 4, §3.4). Démontrons-le pour la deuxième. Soient Γ un chemin qui entoure le segment $[0, 1]$ dans sens direct, et $k \in \{0, \dots, n-1\}$. On a :

$$\frac{1}{2i\pi} \int_{\Gamma} z^k \left(2\tilde{A}_n(z)\text{Li}_3(1/z) + \tilde{B}_n(z)\text{Li}_2(1/z) \right) dz = - \int_0^1 \left(\tilde{B}_n(x) - \tilde{A}_n(x) \log(x) \right) x^k \log(x) dx,$$

d'après (16), en intervertissant les deux signes d'intégration et en appliquant le théorème des résidus.

Il découle de ceci que le problème (14) a une solution unique (à proportionnalité près), donnée par $\tilde{A}_n = A_n$ et $\tilde{B}_n = B_n$.

1.8. D'autres problèmes d'approximation de Padé

Sorokin [So3] considère le problème de Padé suivant : pour $n \geq 0$, trouver des polynômes T_n , U_n , V_n , W_n de degré au plus n tels qu'on ait :

$$\begin{cases} I_{\mathcal{P},n}(z) := T_n(z)\text{Le}_{2,1}(1/z) + U_n(z)\text{Le}_{1,1}(1/z) + V_n(z)\text{Li}_1(1/z) + W_n(z) = O(z^{-n-1}) & \text{quand } z \rightarrow \infty \\ T_n(z)\text{Li}_2(1-z) + V_n(z) = O((1-z)^{n+1}) \text{ quand } z \rightarrow 1 \\ T_n(z)\text{Li}_1(1-z) + U_n(z) = O((1-z)^{n+1}) \text{ quand } z \rightarrow 1, \end{cases}$$

où pour $s_1, \dots, s_k \geq 1$ on définit le polylogarithme multiple

$$\text{Le}_{s_1, \dots, s_k}(z) = \sum_{n_1 \geq \dots \geq n_k \geq 1} \frac{z^{n_1}}{n_1^{s_1} \dots n_k^{s_k}},$$

qui vérifie $\text{Le}_{2,1}(1) = 2\zeta(3)$ (voir [Wa]).

Sorokin démontre que ce problème de Padé admet une solution unique, et qu'à proportionnalité près elle vérifie (pour $z \in \mathbf{C} \setminus [0, 1]$) :

$$I_{P,n}(z) = z^{n+1} \int_0^1 \int_0^1 \int_0^1 \frac{u^n(1-u)^n v^n(1-v)^n w^n(1-w)^n}{(z-uv)^{n+1}(z-uvw)^{n+1}} du dv dw.$$

Avec cette normalisation, T_n est à coefficients entiers (donc aussi $d_n U_n$, $d_n^2 V_n$ et $d_n^3 W_n$), d'où :

$$I_{P,n}(1) = 2(u_{P,n} \zeta(3) - v_{P,n}) \text{ avec } u_{P,n} \in \mathbf{Z} \text{ et } 2d_n^3 v_{P,n} \in \mathbf{Z}.$$

De plus l'expression intégrale donne facilement l'estimation asymptotique de $I_{P,n}(1)$; c'est ainsi que Sorokin démontre l'irrationalité de $\zeta(3)$.

Un théorème général de Zlobin [Zl] montre qu'on a

$$I_{P,n}(z) = \int_0^1 \int_0^1 \int_0^1 \frac{u^n(1-u)^n v^n(1-v)^n w^n(1-w)^n}{(z-w(1-uv))^{n+1}} du dv dw,$$

d'où $I_{P,n}(1) = I_{r,n}(1)$. On peut obtenir directement ce résultat en appliquant le changement de variables ([Fi1], §2) défini par $U = 1 - w$, $V = \frac{(1-u)v}{1-uv}$ et $W = u$ (et qui vérifie $1 - W(1 - UV) = \frac{(1-u)(1-uvw)}{1-uv}$).

Il existe plusieurs autres problèmes de Padé liés à $\zeta(3)$; l'un d'entre eux [So1] fait apparaître l'intégrale suivante :

$$\int_0^1 \int_0^1 \int_0^1 \frac{u^n(1-u)^n v^n(1-v)^n w^n(1-w)^n}{(z(1-u+uv) - uvw)^{n+1}} du dv dw.$$

Le changement de variables qui fixe u et w et change v en $\frac{v}{1-u(1-v)}$ transforme cette intégrale en

$$\int_0^1 \int_0^1 \int_0^1 \frac{u^n(1-u)^n v^n(1-v)^n w^n(1-w)^n}{(1-uv)^{n+1}(z-uvw)^{n+1}} du dv dw.$$

Ces différents problèmes de Padé fournissent tous les formes linéaires d'Apéry en 1 et $\zeta(3)$, mais ils correspondent à des combinaisons linéaires différentes de polylogarithmes.

1.9. Série hypergéométrique très bien équilibrée

On pose :

$$\begin{aligned} H_n(X) &= n!^2(2X+n) \frac{(X-1)\dots(X-n)(X+n+1)\dots(X+2n)}{X^4(X+1)^4\dots(X+n)^4} \\ &= n!^2(2X+n) \frac{(X-n)_n(X+n+1)_n}{(X)_{n+1}^4} \end{aligned}$$

et

$$I_{TB,n}(z) = \sum_{k=1}^{\infty} H_n(k) z^{-k}.$$

La série $I_{TB,n}(1)$ a été introduite par K. Ball (voir [Ri4]) dans le but de répondre à une question de Nesterenko [Ne2] : trouver une preuve de l'irrationalité de $\zeta(3)$ analogue

à celle de Fourier ([FN], Chapitre 2, §1.1) pour l'irrationalité de e . En effet, on peut estimer $I_{\text{TB},n}(1)$ de manière élémentaire ([Zu5], Lemme 4 ; [Ri2], §5.1 ; voir aussi la seconde démonstration du lemme 3 de [BR]) :

$$\lim_{n \rightarrow +\infty} \frac{\log(I_{\text{TB},n}(1))}{n} = \log((\sqrt{2} - 1)^4),$$

ou bien (voir le paragraphe 2.3) déduire cette estimation d'une représentation intégrale de $I_{\text{TB},n}(z)$ vue comme série hypergéométrique très bien équilibrée :

$$I_{\text{TB},n}(z) = z^{-n-1} \frac{n!^7 (3n+2)!}{(2n+1)!^5} {}_7F_6 \left(\begin{matrix} 3n+2, & \frac{3}{2}n+2, & n+1, & \dots, & n+1 \\ \frac{3}{2}n+1, & 2n+2, & \dots, & 2n+2 \end{matrix} \middle| z^{-1} \right).$$

De plus, on a $I_{\text{TB},n}(z) = P_0(z) + \sum_{j=1}^4 P_j(z) \text{Li}_j(1/z)$ avec des polynômes $P_0, \dots, P_4 \in \mathbb{Q}[z]$ vérifiant $P_j(z) = (-1)^{j+1} z^4 P_j(1/z)$ pour tout $j \in \{1, \dots, 4\}$, $P_1(1) = 0$ et $d_n^{4-j} P_j(z) \in \mathbb{Z}[z]$ pour tout $j \in \{0, \dots, 4\}$ (ceci sera généralisé au paragraphe 2.3). En particulier, on en déduit

$$(17) \quad I_{\text{TB},n}(1) = 2(u_{\text{TB},n} \zeta(3) - v_{\text{TB},n}) \text{ avec } 2d_n u_{\text{TB},n} \in \mathbb{Z} \text{ et } 2d_n^4 v_{\text{TB},n} \in \mathbb{Z}.$$

Mais ceci ne suffit pas à démontrer l'irrationalité de $\zeta(3)$, car $(\sqrt{2} - 1)^4 e^4 > 1$.

Une identité de Bailey ([Zu4], Proposition 2 ; [Sl], formule (4.7.1.3)) donne $I_{\text{TB},n}(1) = I_{c,n}(1)$. Une telle identité ne peut pas avoir lieu pour tout z , car $\text{Li}_4(1/z)$ apparaît dans la décomposition en polylogarithmes de $I_{\text{TB},n}(z)$ mais pas dans celle de $I_{c,n}(z)$. Par ailleurs Zudilin a démontré une identité générale ([Zu3], Théorème 5) qui écrit une série hypergéométrique très bien équilibrée sous la forme d'une intégrale généralisant celles introduites par Beukers [Be1], Vasilenko [V] et Vasilyev ([Va1], [Va2]). Dans notre cas particulier, cette identité est $I_{\text{TB},n}(1) = I_{\tau,n}(1)$. Enfin, en utilisant les algorithmes décrits dans [PWZ] on peut démontrer que $I_{\text{TB},n}(1)$ ([Ri2], §5.1 ; [Zu5]), ainsi que $u_{\text{TB},n}$ et $v_{\text{TB},n}$ [Kr], vérifient la relation de récurrence (1). On en déduit $u_{\text{TB},n} = u_{E,n}$ et $v_{\text{TB},n} = v_{E,n}$, d'où $u_{\text{TB},n} \in \mathbb{Z}$ et $2d_n^3 v_{\text{TB},n} \in \mathbb{Z}$ (ce qui est plus précis que (17)).

1.10. Preuve utilisant des formes modulaires

Dans ce paragraphe, on esquisse une preuve due à Beukers [Be6] de l'irrationalité de $\zeta(3)$. Les outils mis en œuvre sont exposés dans [Se] (Chapitre VII) et [Za1].

Pour τ dans le demi-plan de Poincaré \blacktriangledown , posons $q = e^{2i\pi\tau}$ et considérons les séries d'Eisenstein $E_2(\tau) = 1 - 24 \sum_{n \geq 1} \sigma_1(n) q^n$ et $E_4(\tau) = 1 + 240 \sum_{n \geq 1} \sigma_3(n) q^n$. On pose :

$$\begin{aligned} E(\tau) &= \frac{1}{24} (-5E_2(\tau) + 2E_2(2\tau) - 3E_2(3\tau) + 30E_2(6\tau)) \\ \text{et } F(\tau) &= \frac{1}{40} (E_4(\tau) - 28E_4(2\tau) + 63E_4(3\tau) - 36E_4(6\tau)). \end{aligned}$$

Alors $E(\tau)$, respectivement $F(\tau)$, est une forme modulaire de poids 2, resp. 4, pour $\Gamma_0(6)$. Si $F(\tau) = \sum_{n \geq 1} f_n q^n$ désigne le développement de Fourier de F à l'infini (où elle s'annule), on pose $f(\tau) = \sum_{n \geq 1} \frac{f_n}{n^3} q^n$. On a alors $(\frac{d}{d\tau})^3 f(\tau) = (2i\pi)^3 F(\tau)$.

Considérons la fonction modulaire pour $\Gamma_0(6)$ donnée par :

$$t(\tau) = \left(\frac{\Delta(6\tau)\Delta(\tau)}{\Delta(2\tau)\Delta(3\tau)} \right)^{1/2} = q \prod_{\substack{n \geq 1 \\ \text{pgcd}(n,6) = 1}} (1 - q^n)^{12},$$

avec $\Delta(\tau) = q \prod_{n \geq 1} (1 - q^n)^{24}$. Elle n'a ni zéro ni pôle dans ∇ . Au voisinage de $q = 0$, $t(\tau) = q - 12q^2 + 66q^3 - \dots$ s'écrit comme une série entière en q , à coefficients entiers, avec un rayon de convergence égal à 1. Elle admet une réciproque locale, notée $q(t) \in \mathbf{Z}[[t]]$. Par composition, on peut donc définir des suites $(u_{M,n})$ et $(v_{M,n})$ par :

$$E(q(t)) = \sum_{n \geq 0} u_{M,n} t^n \in \mathbf{Z}[[t]]$$

$$\text{et } E(q(t))f(q(t)) = \sum_{n \geq 0} v_{M,n} t^n \in \mathbf{Q}[[t]] \text{ avec } v_{M,0} = 0 \text{ et } d_n^3 v_{M,n} \in \mathbf{Z} \text{ pour tout } n \geq 1.$$

Notons, pour $k \in \mathbf{Z}$, w_k l'opérateur d'Atkin-Lehner défini par $(w_k g)(\tau) = 6^{-k/2} \tau^{-k} g(\frac{-1}{6\tau})$. Alors $w_2(E) = -E$ et $w_4(F) = -F$. De cette seconde égalité (et d'un lemme de Hecke : voir [We], §5) découle la relation $w_{-2}(h) = -h$, en posant $h(\tau) = L(F, 3) - f(\tau)$, où $L(F, s)$ est la fonction L de F . Il vient alors $w_0(Eh) = Eh$, c'est-à-dire que la fonction $E(\tau)h(\tau)$ est invariante par la substitution $\tau \mapsto \frac{-1}{6\tau}$.

Considérons maintenant les rayons de convergence. La fonction $t(\tau)$ est ramifiée seulement au-dessus des points $(\sqrt{2} - 1)^4$, $(\sqrt{2} + 1)^4$ et ∞ . Au-dessus de $(\sqrt{2} - 1)^4$, le seul point de ramification (modulo $\Gamma_0(6)$) est $\tau = i/\sqrt{6}$; il est d'indice deux, et les deux branches en ce point sont échangées par l'involution $\tau \mapsto \frac{-1}{6\tau}$. Comme $E(\tau)h(\tau)$ est invariante par cette involution, on peut définir Eh comme une fonction de t au voisinage de $t = (\sqrt{2} - 1)^4$, et en fait sur tout le disque $|t| < (\sqrt{2} + 1)^4$. Cela signifie que la série $\sum_{n \geq 0} (L(F, 3)u_{M,n} - v_{M,n})t^n$ a un rayon de convergence supérieur ou égal à $(\sqrt{2} + 1)^4$, c'est-à-dire qu'on a :

$$\limsup_{n \rightarrow \infty} \frac{\log |L(F, 3)u_{M,n} - v_{M,n}|}{n} \leq \log((\sqrt{2} - 1)^4).$$

Ceci conclut la démonstration de l'irrationalité de $L(F, 3)$. Or on peut calculer explicitement $L(F, s)$. En effet, quand $\text{Re}(s) > 4$ on a, pour tout entier $j \geq 1$:

$$L(E_4(j\tau), s) = 1 + 240 \sum_{n \geq 1} \frac{\sigma_3(n)}{(jn)^s} = 1 + 240 \sum_{d,e \geq 1} \frac{d^3}{(jde)^s} = 1 + 240 \zeta(s) \zeta(s-3) j^{-s}.$$

On en déduit immédiatement $L(F, s) = -2\zeta(s)\zeta(s-3)$, d'où $L(F, 3) = \zeta(3)$.

Comme $E(\tau)$ est une forme modulaire de poids 2 et $t(\tau)$ une fonction modulaire, la fonction $E(q(t))$ de la variable t est solution [Za2] (voir aussi [Be4], p. 58) d'une équation

différentielle linéaire $\triangleright y = 0$, d'ordre trois. On peut la déterminer explicitement :

$$\triangleright = (t^4 - 34t^3 + t^2) \frac{d^3}{dt^3} + (6t^3 - 153t^2 + 3t) \frac{d^2}{dt^2} + (7t^2 - 112t + 1) \frac{d}{dt} + (t - 5).$$

Cette équation différentielle vérifiée par la série génératrice des $u_{M,n}$ montre qu'ils satisfont à la relation de récurrence (1) : on a donc $u_{M,n} = u_{R,n}$ (voir aussi [Be7]). En posant $V(t) = E(q(t))f(q(t))$ on montre [Za2] que $\triangleright V = 5$, d'où $v_{M,n} = v_{R,n}$.

Une base de solutions de l'équation différentielle $\triangleright y = 0$ est donnée par $E(q(t))$, $\tau(t)E(q(t))$ et $\tau^2(t)E(q(t))$ (voir aussi [BP], Corollaire 2). La seule solution qui soit régulière en 0 est $E(q(t))$ (à proportionnalité près). De plus, la construction de \triangleright montre [Za2] que c'est un carré symétrique, ce qui peut se vérifier directement (voir [Dw1]).

REMARQUE 1.5. — Le point de vue adopté dans ce paragraphe est lié “individuellement” à $\zeta(3)$ (qui est vu comme valeur spéciale d'une fonction L), par opposition aux méthodes utilisées dans les paragraphes 1.3 à 1.9, où $\zeta(3)$ apparaissait comme la valeur en 1 d'un polylogarithme.

Cette preuve de l'irrationalité de $\zeta(3)$ s'exprime naturellement en termes des séries génératrices $U(t) = \sum_{n \geq 0} u_n t^n$ et $V(t) = \sum_{n \geq 0} v_n t^n$ des approximations rationnelles de $\zeta(3)$ (voir [Po2], [Be6] et [Ch], §5 pour d'autres preuves dans le même esprit). L'aspect arithmétique consiste à démontrer que les coefficients de $U(t)$ sont entiers, et que d_n^3 est un dénominateur commun aux n premiers coefficients de $V(t)$: c'est une majoration p -adique de ces coefficients, pour toute place finie p . L'aspect analytique est une minoration, par $(1 + \sqrt{2})^4$, du rayon de convergence (archimédien) de la série entière $\zeta(3)U(t) - V(t)$. En particulier, $U(t)$ et $V(t)$ sont des G -fonctions de Siegel. La série $U(t)$ est une solution de l'équation différentielle $\triangleright y = 0$; la conjecture de Bombieri-Dwork prédit ([Dw1], [Dw2] ; voir aussi [An] et [Dw3]) que \triangleright provient de la géométrie.

Or, pour $t \in \mathbf{P}^1(\mathbf{C}) \setminus \{0, 1, (\sqrt{2} \pm 1)^4, \infty\}$, Beukers et Peters construisent [BP] une surface K3 X_t birationnellement équivalente à la surface projective S_t d'équation affine $1 - (1 - xy)z - txyz(1 - x)(1 - y)(1 - z) = 0$. Ils montrent que si ω_t est l'unique 2-forme holomorphe sur X_t (à proportionnalité près), et si τ_t est un certain 2-cycle (constant pour la connexion de Gauss-Manin), alors $U(t)$ est l'intégrale de ω_t sur τ_t . En particulier $\triangleright y = 0$ est l'équation de Picard-Fuchs de cette famille de surfaces : elle provient bien de la géométrie.

1.11. Congruences

De nombreux auteurs ont étudié des propriétés de congruence sur les nombres d'Apéry u_n . Par exemple, Chowla, Cowles et Cowles [CCC] ont conjecturé $u_p \equiv 5 \pmod{p^3}$ pour tout $p \geq 5$ premier. Cette conjecture a été démontrée par plusieurs auteurs (voir par exemple [Ges], [Su], ...). De nombreuses autres congruences ont été prouvées, pour les nombres d'Apéry et certaines de leurs généralisations.

Notons $\sum_{n \geq 1} \gamma_n q^n = q \prod_{n \geq 1} (1 - q^{2n})^4 (1 - q^{4n})^4$ l'unique forme parabolique normalisée de poids 4 pour $\Gamma_0(8)$. Pour $r \geq 1$, $m \geq 1$ impair et p premier impair, on a la congruence suivante (qui ressemble à celles d'Atkin - Swinnerton-Dyer, voir [Haz] §VI.33) :

$$(18) \quad u_{\frac{1}{2}(mp^r-1)} - \gamma_p u_{\frac{1}{2}(mp^{r-1}-1)} + p^3 u_{\frac{1}{2}(mp^{r-2}-1)} \equiv 0 \pmod{p^r}$$

avec la convention $u_t = 0$ si $t \notin \mathbf{Z}$. Beukers la démontre [Be7] en utilisant la construction modulaire du paragraphe 1.10. On en déduit $u_{\frac{p-1}{2}} \equiv \gamma_p \pmod{p}$, congruence dont Beukers a conjecturé [Be7] qu'elle est vraie modulo p^2 . Ceci a été prouvé par Ishikawa [Is] si p ne divise pas $u_{\frac{p-1}{2}}$, puis par Ahlgren et Ono [AO] dans le cas général. Ahlgren et Ono utilisent des séries hypergéométriques sur \mathbf{F}_p et la modularité de la variété d'équation $x + \frac{1}{x} + y + \frac{1}{y} + z + \frac{1}{z} + w + \frac{1}{w} = 0$ (dont la famille de surfaces K3 considérée par Beukers-Peters est un quotient : voir [PS], Théorème 4).

Pour $r, m \geq 1$ et $p \geq 5$ premier, Beukers a démontré [Be5], de manière élémentaire, qu'on a $u_{mp^r-1} \equiv u_{mp^{r-1}-1} \pmod{p^{3r}}$. La même congruence, mais seulement modulo p^r , s'interprète en disant que $\int_0^T U(t) dt$ est (vue comme série formelle en T) le logarithme d'une loi de groupe formel sur \mathbf{Z} qui est isomorphe à \mathbf{G}_m sur \mathbf{Z} ([Be5] ; voir aussi l'appendice de [SB] ou [Haz], §VI.33).

2. IRRATIONALITÉ D'UNE INFINITÉ DE $\zeta(2k+1)$

2.1. Énoncé des résultats

Dans cette partie, on démontre les résultats suivants, dont le premier implique le théorème 0.3 :

THÉORÈME 2.1 ([Ri1], [BR]). — *Pour $\ell \geq 3$ impair, notons δ_ℓ la dimension du \mathbf{Q} -espace vectoriel engendré par $1, \zeta(3), \zeta(5), \dots, \zeta(\ell)$. Pour tout $\varepsilon > 0$ il existe un entier ℓ_0 tel que pour tout $\ell \geq \ell_0$ impair on ait :*

$$\delta_\ell \geq \frac{1 - \varepsilon}{1 + \log(2)} \log(\ell).$$

REMARQUE 2.2. — Si dans le théorème 2.1 on remplace $\frac{1-\varepsilon}{1+\log(2)}$ par $\frac{1}{3}$ alors [BR] on peut prendre $\ell_0 = 3$.

THÉORÈME 2.3 ([BR]). — *Il existe un entier impair ℓ , avec $\ell \leq 169$, tel que $1, \zeta(3)$ et $\zeta(\ell)$ soient linéairement indépendants sur \mathbf{Q} .*

Ce théorème a été amélioré par Zudilin [Zu2], qui remplace 169 par 145, grâce à un raffinement du lemme 2.12 ci-dessous.

Les deux ingrédients essentiels de la démonstration du théorème 2.1 sont l'absence de $\zeta(2), \zeta(4), \dots, \zeta(\ell-1)$ d'une part, et la minoration en $\log(\ell)$ de la dimension d'autre part. Seule cette deuxième idée est utile pour démontrer le théorème suivant.

THÉORÈME 2.4 ([Ri2]). — Soient $z \in \mathbf{Q}$, $|z| > 1$, et $\varepsilon > 0$. Il existe un entier ℓ_0 (qui dépend de z et ε) tel que, pour tout $\ell \geq \ell_0$, la dimension du \mathbf{Q} -espace vectoriel engendré par $1, \text{Li}_1(1/z), \text{Li}_2(1/z), \dots, \text{Li}_\ell(1/z)$ soit minorée par $\frac{1-\varepsilon}{1+\log(2)} \log(\ell)$.

En conséquence, pour tout nombre rationnel z de valeur absolue supérieure à 1 il existe une infinité d'entiers j tels que $\text{Li}_j(1/z)$ soit irrationnel. Par ailleurs, quand z est un entier négatif tel que $|z| > (4\ell)^{\ell(\ell-1)}$, Nikishin a démontré [Ni] que les nombres $1, \text{Li}_1(1/z), \text{Li}_2(1/z), \dots, \text{Li}_\ell(1/z)$ sont linéairement indépendants sur \mathbf{Q} ; sa méthode a inspiré en partie la construction exposée au paragraphe suivant. Hata a raffiné ([Hat1], [Hat2]) le résultat de Nikishin : par exemple $1, \text{Li}_1(1/z)$ et $\text{Li}_2(1/z)$ sont linéairement indépendants sur \mathbf{Q} pour $z \leq -5$ ou $z \geq 7$.

2.2. Structure de la preuve

Soient a et r deux entiers, avec $a \geq 3$ et $1 \leq r < \frac{a}{2}$. Soit $n \geq 1$. Définissons \mathbf{R}_n et \mathbf{S}_n (qui dépendent aussi de a et r) par :

$$\begin{aligned} \mathbf{R}_n(k) &= 2n!^{a-2r} \left(k + \frac{n}{2}\right) \frac{(k - rn)_{rn} (k + n + 1)_{rn}}{(k)_{n+1}^a} \\ &= 2n!^{a-2r} \left(k + \frac{n}{2}\right) \frac{(k-1)(k-2)\dots(k-rn)(k+n+1)(k+n+2)\dots(k+(r+1)n)}{k^a (k+1)^a \dots (k+n)^a} \end{aligned}$$

et

$$(19) \quad \mathbf{S}_n(z) = \sum_{k \geq 1} \mathbf{R}_n(k) z^{-k}.$$

Cette série converge absolument pour tout nombre complexe z tel que $|z| \geq 1$, car $\mathbf{R}_n(k) = O(k^{-2})$ quand k tend vers l'infini.

Les propriétés de cette série étudiées au paragraphe 2.3 permettent de démontrer les théorèmes 2.1 (en prenant $z = 1$ et a pair), 2.3 (avec $z = 1$, $a = 169$, $r = 10$ et n impair ; on utilise le théorème d'Apéry) et 2.4 (avec $z \in \mathbf{Q}$, $z > 1$; pour $z < -1$ il suffirait de modifier le lemme 2.9). Les trois preuves sont parallèles ; on détaille dans ce paragraphe la structure de celle du théorème 2.1.

On suppose a pair ; on construit des formes linéaires en $1, \zeta(3), \zeta(5), \dots, \zeta(a-1)$ grâce à la proposition suivante :

PROPOSITION 2.5. — Supposons a pair. Notons d_n le p.p.c.m des entiers de 1 à n . Alors il existe des nombres rationnels $\kappa_0, \kappa_3, \kappa_5, \dots, \kappa_{a-1}$ tels que :

- (1) On a $\mathbf{S}_n(1) = \kappa_0 + \kappa_3 \zeta(3) + \kappa_5 \zeta(5) + \kappa_7 \zeta(7) + \dots + \kappa_{a-1} \zeta(a-1)$.
- (2) Pour tout $j \in \{0, 3, 5, \dots, a-1\}$ on a $\limsup_{n \rightarrow +\infty} |\kappa_j|^{1/n} \leq 2^{a-2r} (2r+1)^{2r+1}$.
- (3) Pour tout $j \in \{0, 3, 5, \dots, a-1\}$, le nombre rationnel $d_n^a \kappa_j$ est un entier.
- (4) Il existe un réel $\psi_{r,a} > 0$ tel que $\lim_{n \rightarrow +\infty} |\mathbf{S}_n(1)|^{1/n} = \psi_{r,a} \leq \frac{2^{r+1}}{r^{a-2r}}$.

En fait on conjecture que l'amélioration suivante est possible :

CONJECTURE 2.6 ([Ri2]). — *Dans l'assertion (3) de la proposition 2.5, on peut remplacer d_n^a par d_n^{a-1} .*

REMARQUE 2.7. — En prenant $a = 4$ (et $r = 1$), on obtient les formes linéaires en 1 et $\zeta(3)$ du paragraphe 1.9, donc la conjecture 2.6 est vraie quand $a = 4$. Elle est démontrée aussi quand $a = 6$ et $r = 1$ (voir la fin du paragraphe 2.4). On ne connaît pas de conséquence directe de cette conjecture, mais une version forte de celle-ci pourrait éventuellement permettre de démontrer que parmi $\zeta(5)$, $\zeta(7)$ et $\zeta(9)$, l'un au moins est irrationnel (voir la remarque 3.4). En tout cas, il serait intéressant d'obtenir une preuve de la conjecture 2.6 grâce à une interprétation (par exemple géométrique, comme au paragraphe 1.10) de $\kappa_0, \dots, \kappa_{a-1}$.

La proposition 2.5 fournit des formes linéaires en 1, $\zeta(3)$, \dots , $\zeta(a-1)$ (si a est pair). Si cette suite de formes linéaires tend vers 0, sans être nulle à partir d'un certain rang, alors l'un au moins des nombres $\zeta(3)$, \dots , $\zeta(a-1)$ est irrationnel. Cette remarque sera utilisée pour démontrer le théorème 0.4. Ici on veut obtenir les théorèmes 2.1 à 2.4, donc on a besoin d'un critère d'indépendance linéaire, qui donne une minoration plus fine de la dimension du \mathbb{Q} -espace vectoriel engendré par 1, $\zeta(3)$, \dots , $\zeta(a-1)$. On va utiliser à cet effet le théorème 2.8 ci-dessous.

La meilleure minoration qu'on puisse espérer est donnée par le principe des tiroirs, de la manière suivante. Soient α et β des réels, avec $0 < \alpha < 1$ et $\beta > 1$. Soient $\theta_1, \dots, \theta_s$ des réels qui engendrent un \mathbb{Q} -espace vectoriel de dimension au moins $1 - \frac{\log(\alpha)}{\log(\beta)}$. Alors il existe une suite (ℓ_n) de formes linéaires en $\theta_1, \dots, \theta_s$ dont les coefficients entiers $p_{j,n}$ vérifient $\limsup_{n \rightarrow +\infty} |p_{j,n}|^{1/n} \leq \beta$ pour tout j et telle que $\limsup_{n \rightarrow +\infty} |\ell_n(\theta_1, \dots, \theta_s)|^{1/n} \leq \alpha$. Essentiellement, plus la dimension du \mathbb{Q} -espace vectoriel engendré est grande, plus les formes linéaires qu'on peut construire sont petites. On cherche une réciproque à cette assertion. Une contrainte supplémentaire est nécessaire : si $\frac{\theta_2}{\theta_1}$ est un nombre de Liouville, on peut construire des formes linéaires extrêmement petites même si la dimension du \mathbb{Q} -espace vectoriel engendré est seulement 2. Ce contre-exemple ne tient plus si on demande que les formes linéaires en $\theta_1, \dots, \theta_s$ ne soient pas trop petites. On a alors la réciproque suivante (pour une preuve, voir [Ne1] ou [Col], §II.1) :

THÉORÈME 2.8 ([Ne1]). — *Soient $\theta_1, \dots, \theta_s$ des réels. Pour tout $n \geq 1$, soit $\ell_n = p_{1,n}X_1 + \dots + p_{s,n}X_s$ une forme linéaire à coefficients entiers. Soient α et β des réels, avec $0 < \alpha < 1$ et $\beta > 1$.*

Supposons qu'on ait $\limsup_{n \rightarrow +\infty} |p_{j,n}|^{1/n} \leq \beta$ pour tout j compris entre 1 et s , et

$$\lim_{n \rightarrow +\infty} |\ell_n(\theta_1, \dots, \theta_s)|^{1/n} = \alpha.$$

Alors le \mathbb{Q} -espace vectoriel engendré par $\theta_1, \dots, \theta_s$ est de dimension au moins $1 - \frac{\log(\alpha)}{\log(\beta)}$.

Pour déduire le théorème 2.1 de la proposition 2.5 et de ce critère d'indépendance linéaire, il suffit de considérer $d_n^a \mathbf{S}_n(1)$, qui est une forme linéaire à coefficients entiers en $1, \zeta(3), \zeta(5), \dots, \zeta(a-1)$. On choisit a suffisamment grand, et r égal à la partie entière de $\frac{a}{(\log(a))^2}$. Alors r^r est négligeable devant c^a (pour toute constante c), et on peut prendre β essentiellement égal à $(2e)^a = e^{a(1+\log(2))}$ et α essentiellement majoré par r^{-a} , qui est de l'ordre de $e^{-a \log(a)}$. Cela démontre le théorème 2.1.

2.3. Quelques détails sur la preuve

Soit z un nombre complexe de module supérieur ou égal à 1. La série $\mathbf{S}_n(z)$ peut s'écrire comme une série hypergéométrique très bien équilibrée, de la manière suivante :

$$\mathbf{S}_n(z) = z^{-rn-1} n!^{a-2r} \frac{(rn)!((r+1)n+2)_{rn+1}}{(rn+1)_{n+1}^a} \times$$

$${}_aF_{a+2} \left(\begin{matrix} (2r+1)n+2, & (r+\frac{1}{2})n+2, & rn+1, & \dots, & rn+1 \\ (r+\frac{1}{2})n+1, & (r+1)n+2, & \dots, & (r+1)n+2 \end{matrix} \middle| z^{-1} \right).$$

Cette identité provient de simplifications dans les symboles de Pochhammer.

2.3.1. *Représentation intégrale et estimation analytique.* On a la représentation intégrale suivante, pour $|z| \geq 1$:

$$\mathbf{S}_n(z) = \frac{((2r+1)n+2)!}{n!^{2r+1}} z^{(r+1)n+1} \int_{[0,1]^{a+1}} \left(\frac{\prod_{j=1}^{a+1} t_j^r (1-t_j)}{(z-t_1 t_2 \dots t_{a+1})^{2r+1}} \right)^n \frac{z+t_1 \dots t_{a+1}}{(z-t_1 \dots t_{a+1})^3} dt_1 \dots dt_{a+1}.$$

Cette formule (voir par exemple [RZ], Lemme 1) se déduit de l'écriture de $\mathbf{S}_n(z)$ comme série hypergéométrique : pour $|z| > 1$ on applique les relations (4.1.2) et (1.5.21) de [Sl], puis on prolonge à $|z| = 1$ par continuité (voir la preuve du lemme 2 de [BR]). On peut aussi obtenir une preuve directe en développant en série le dénominateur de l'intégrande ([Col], [Hab]).

En calculant le maximum sur $[0, 1]^{a+1}$ de la fonction dont on intègre la puissance n -ième, on déduit de cette représentation intégrale l'estimation analytique suivante :

LEMME 2.9. — On suppose $z \in \mathbf{R}$, $z \geq 1$. Le polynôme

$$Q_{r,a,z}(s) = r s^{a+2} - (r+1) s^{a+1} + (r+1) z s - r z$$

admet une racine unique $s_0 \in [0, 1]$, et elle vérifie $s_0 > \frac{r}{r+1}$. De plus, si

$$\phi_{r,a,z} = z^{-r} ((r+1)s_0 - r)^r (r+1 - r s_0)^{r+1} (1 - s_0)^{a-2r},$$

alors

$$\lim_{n \rightarrow \infty} |\mathbf{S}_n(z)|^{1/n} = \phi_{r,a,z} \leq \frac{2^{r+1}}{z^r r^{a-2r}}.$$

Pour démontrer ce lemme, il suffit d'adapter les preuves du lemme 2.2 de [Ri2] et du lemme 3 de [BR]. On pourrait aussi donner une démonstration élémentaire de ce comportement asymptotique, sans utiliser la représentation intégrale (comme la deuxième preuve du lemme 3 de [BR]). Enfin, une troisième possibilité serait d'écrire $\mathbf{S}_n(z)$ comme intégrale complexe et d'appliquer la méthode du col ; mais cette méthode est très difficile à mettre en œuvre quand r , a et z sont des paramètres.

REMARQUE 2.10. — Pour démontrer les théorèmes 2.1 et 2.4, il suffit de connaître l'existence de la limite de $|\mathbf{S}_n(z)|^{1/n}$, et sa majoration par $\frac{2^{r+1}}{z^r r^{a-2r}}$. La valeur exacte de $\phi_{r,a,z}$ n'est utile que pour obtenir des estimations numériques précises (par exemple pour le théorème 2.3).

2.3.2. *Décomposition en polylogarithmes.* Pour démontrer que $\mathbf{S}_n(z)$ est une combinaison linéaire (à coefficients rationnels) de $1, \text{Li}_1(1/z), \dots, \text{Li}_a(1/z)$ quand $|z| > 1$, il suffit de décomposer la fraction rationnelle \mathbf{R}_n en éléments simples, sous la forme suivante :

$$(20) \quad \mathbf{R}_n(k) = \sum_{i=0}^n \sum_{j=1}^a \frac{c_{i,j}}{(k+i)^j}$$

où les coefficients $c_{i,j}$ sont des rationnels, donnés par

$$(21) \quad c_{i,j} = \frac{1}{(a-j)!} \left(\frac{d}{dX} \right)^{a-j} (\mathbf{R}_n(X)(X+i)^a)|_{X=-i}.$$

On a pour $|z| > 1$:

$$\begin{aligned} \mathbf{S}_n(z) &= \sum_{i=0}^n \sum_{j=1}^a c_{i,j} \sum_{k \geq 1} \frac{z^{-k}}{(k+i)^j} \\ &= \sum_{i=0}^n \sum_{j=1}^a c_{i,j} z^i \text{Li}_j(1/z) - \sum_{i=0}^n \sum_{j=1}^a c_{i,j} \sum_{q=1}^i \frac{z^{i-q}}{q^j}, \end{aligned}$$

d'où

$$(22) \quad \mathbf{S}_n(z) = P_0(z) + \sum_{j=1}^a P_j(z) \text{Li}_j(1/z)$$

en posant

$$(23) \quad P_0(z) = - \sum_{\ell=0}^{n-1} \left(\sum_{i=\ell+1}^n \sum_{j=1}^a \frac{c_{i,j}}{(i-\ell)^j} \right) z^\ell$$

et

$$(24) \quad P_j(z) = \sum_{i=0}^n c_{i,j} z^i \text{ pour } j \in \{1, \dots, a\}.$$

Bien sûr, les P_j et les $c_{i,j}$ dépendent aussi de n , a et r .

2.3.3. *Propriété de symétrie.* La fonction \mathbf{R}_n vérifie la propriété de symétrie suivante :

$$\mathbf{R}_n(-k - n) = (-1)^{a(n+1)+1} \mathbf{R}_n(k).$$

Cette symétrie est rendue possible par la présence des deux facteurs de Pochhammer au numérateur de $\mathbf{R}_n(k)$: quand k est changé en $-k - n$, ils sont permutés (on applique la formule $(-\alpha)_p = (-1)^p (\alpha - p + 1)_p$).

L'unicité du développement en éléments simples montre que $c_{i,j} = (-1)^{j+a(n+1)+1} c_{n-i,j}$ pour tous $i \in \{0, \dots, n\}$ et $j \in \{1, \dots, a\}$, ce qui donne pour tout $j \in \{1, \dots, a\}$:

$$(25) \quad P_j(z) = (-1)^{j+a(n+1)+1} z^n P_j(1/z).$$

En particulier, si $j + a(n + 1)$ est pair alors $P_j(1) = 0$. De plus on a $P_1(1) = 0$, car $P_1(1) = \sum_{i=0}^n c_{i,1}$ est l'opposé du résidu à l'infini de \mathbf{R}_n (on peut aussi faire tendre z vers 1 dans (22) et constater que le seul terme qui puisse tendre vers l'infini est $P_1(z)\text{Li}_1(1/z)$). Quand a est pair, on obtient donc :

$$\mathbf{S}_n(1) = P_0(1) + P_3(1)\zeta(3) + P_5(1)\zeta(5) + \dots + P_{a-1}(1)\zeta(a-1).$$

Quand a est impair et n pair, on obtient de même une forme linéaire en $1, \zeta(2), \zeta(4), \dots, \zeta(a-1)$ dont on peut se servir pour montrer qu'une infinité de puissances de π sont linéairement indépendantes sur \mathbf{Q} , i.e. que π est transcendant. On peut aussi en déduire une mesure de transcendance de π , à la manière de Reyssat [Re2].

Enfin, quand a et n sont impairs, on obtient une forme linéaire en $1, \zeta(3), \zeta(5), \dots, \zeta(a)$; c'est ce qu'on utilise pour démontrer le théorème 2.3.

2.3.4. *ficients de la forme linéaire.*

LEMME 2.11. — Pour tout $j \in \{0, \dots, a\}$ on a :

$$\limsup_{n \rightarrow +\infty} |P_j(z)|^{1/n} \leq 2^{a-2r} (2r+1)^{2r+1} |z|.$$

PREUVE : On peut suivre la démonstration du lemme 4 de [BR] en écrivant la formule de Cauchy sur le cercle C de centre $-i$ et de rayon $1/2$:

$$c_{i,j} = \frac{1}{2i\pi} \int_C \mathbf{R}_n(t) (t+i)^{j-1} dt.$$

On majore ensuite le module de l'intégrande, et le lemme en découle. Une autre preuve, qui conduit à une majoration légèrement moins précise, est donnée dans [Col] et [Hab].

2.3.5. *Estimation arithmétique.* Les polynômes P_0, \dots, P_a sont à coefficients rationnels ; on a besoin d'un dénominateur commun pour leurs coefficients.

LEMME 2.12. — Pour tout $j \in \{0, \dots, a\}$, le polynôme $d_n^{a-j} P_j(z)$ est à coefficients entiers.

REMARQUE 2.13. — On peut ([Zu2], §4) raffiner ce lemme, ce qui permet de remplacer 169 par 145 dans l'énoncé du théorème 2.3. Cependant, des exemples montrent qu'on ne peut pas espérer remplacer d_n^{a-j} par d_n^{a-1-j} . La conjecture 2.6 signifie que pour $z = 1$ on a des compensations particulières qui font chuter le dénominateur.

PREUVE : Posons $F_s(X) = \frac{(X-sn)_n}{(X)_{n+1}}$ et $G_s(X) = \frac{(X+sn+1)_n}{(X)_{n+1}}$ pour tout $s \in \{1, \dots, r\}$, ainsi que $H(X) = \frac{n!}{(X)_{n+1}}$ et $I(X) = 2X + n$. Alors on a $F_s(X) = \sum_{p=0}^n \frac{f_{p,s}}{X+p}$ avec $f_{p,s} = (-1)^{n-p} \binom{n}{p} \binom{p+sn}{n} \in \mathbf{Z}$, et de même (avec des notations évidentes) $g_{p,s} \in \mathbf{Z}$ et $h_p \in \mathbf{Z}$ pour tous p, s . On obtient alors le développement en éléments simples de $\mathbf{R}_n(X) = (\prod_{s=1}^r F_s(X)) \cdot (\prod_{s=1}^r G_s(X)) \cdot H(X)^{a-2r} \cdot I(X)$ en faisant le produit des développements des facteurs. On utilise les formules $\frac{2X+n}{X+p} = 2 + \frac{n-2p}{X+p}$ et $\frac{1}{(X+p)(X+p')} = \frac{1}{(p'-p)(X+p)} + \frac{1}{(p-p')(X+p')}$ pour $p \neq p'$; les dénominateurs n'apparaissent que par application de la seconde. Ce calcul montre que $d_n^{a-j} c_{i,j}$ est entier pour tous i, j , ce qui achève la preuve (suivant [Col] et [Hab]) du lemme.

2.4. Quelques remarques

Soit Q_n un polynôme à coefficients rationnels, de degré inférieur ou égal à $a(n+1) - 1$. On peut toujours considérer $\mathbf{R}_n(k) = \frac{Q_n(k)}{\binom{k}{n+1}^a}$ et $\mathbf{S}_n(z) = \sum_{k \geq 1} \mathbf{R}_n(k) z^{-k}$, qui converge quand $|z| > 1$. Une difficulté majeure consiste à bien choisir le polynôme Q_n .

Quel que soit ce choix, on peut décomposer \mathbf{R}_n en éléments simples, définir P_0, \dots, P_a et obtenir une décomposition de $\mathbf{S}_n(z)$ en polylogarithmes : toutes les formules du paragraphe 2.3.2 restent valables. Pour obtenir une forme linéaire en valeurs de ζ , il faut¹ faire tendre z vers 1. Tous les termes de la décomposition en polylogarithmes ont une limite finie, sauf peut-être $P_1(z)\text{Li}_1(1/z)$. C'est pourquoi on suppose $P_1(1) = 0$, ce qui signifie que \mathbf{R}_n n'a pas de résidu à l'infini, i.e. $\deg(Q_n) \leq a(n+1) - 2$; alors la série qui définit $\mathbf{S}_n(z)$ converge absolument dès que $|z| \geq 1$.

En outre on souhaite² obtenir une forme linéaire en les $\zeta(2k+1)$ seulement, c'est-à-dire avoir $P_j(1) = 0$ pour tout $j \geq 2$ pair. Pour assurer cela il est suffisant d'avoir une propriété de symétrie du polynôme Q_n , en l'occurrence $Q_n(-k-n) = (-1)^{a(n+1)+1} Q_n(k)$. C'est cette remarque qui constitue le cœur des progrès récents ([Ri1], [BR]). On ne sait pas du tout la généraliser, par exemple pour construire des formes linéaires en $\zeta(s)$ dans lesquelles les s appartenant à une certaine progression arithmétique n'apparaissent pas.

La forme linéaire $\mathbf{S}_n(1)$ ne sera intéressante que si elle tend suffisamment vite vers 0 quand n tend vers l'infini. Intuitivement, ce sera le cas si les premiers termes de la série qui définit $\mathbf{S}_n(1)$ sont nuls. C'est pourquoi on cherche un polynôme $Q_n(k)$ qui s'annule aux premiers entiers, en l'occurrence entre 1 et rn ; ceci signifie que $Q_n(k)$ est multiple

¹Voir cependant la remarque 2.14.

²Sauf pour démontrer le théorème 2.4 ; pour ce dernier, le polynôme $Q_n(k) = (k-rn)_{rn}$ convient aussi. C'est celui qui est utilisé dans le Chapitre 2 de [Ri2].

de $(k - rn)_{rn}$. Il s'agit en fait d'un problème de type Padé : on demande aux polynômes P_0, \dots, P_a d'être tels que

$$\mathbf{S}_n(z) = P_0(z) + \sum_{j=1}^a P_j(z) \text{Li}_j(1/z) = O(z^{-rn-1}) \text{ quand } z \rightarrow \infty.$$

Parmi tous les polynômes symétriques $Q_n(k)$ multiples de $(k - rn)_{rn}$ (donc nécessairement aussi multiples de $(k + n + 1)_{rn}$), on a intérêt à en prendre un de degré minimal, pour que $\mathbf{S}_n(1)$ soit aussi petit que possible. Si $a(n+1)$ est impair, le polynôme $(k - rn)_{rn}(k + n + 1)_{rn}$ a la bonne parité, et on peut considérer $Q_n(k) = n!^{a-2r}(k - rn)_{rn}(k + n + 1)_{rn}$: on obtient la série hypergéométrique bien équilibrée de [Ri1] et [BR]. Si $a(n+1)$ est pair, pour obtenir le bon signe dans la propriété de symétrie de Q_n on est amené à introduire un facteur $k + \frac{n}{2}$, ce qui donne la série très bien équilibrée du paragraphe 2.2. Dans les deux cas, $\mathbf{S}_n(z)$ est la solution unique d'un problème de Padé (voir [Hu2] et [FR]).

Plus a est grand (en prenant, pour chaque a , la valeur optimale de r), plus la forme linéaire à coefficients entiers $d_n^a \mathbf{S}_n(1)$ est petite (et la présence, ou l'absence, du facteur $k + \frac{n}{2}$ a une influence négligeable sur ce comportement). Donc si on cherche des formes linéaires en $1, \zeta(3), \zeta(5), \dots, \zeta(2\ell + 1)$, celles obtenues avec la série très bien équilibrée pour $a = 2\ell + 2$ seront meilleures que celles obtenues avec la série bien équilibrée pour $a = 2\ell + 1$ et n pair. Ceci n'a aucune influence quand ℓ tend vers l'infini, mais peut s'avérer crucial si ℓ est fixé (comme dans le théorème 0.4). En outre, si la conjecture 2.6 (qui n'a aucun équivalent pour des séries seulement bien équilibrées) est vraie alors il suffit de multiplier $\mathbf{S}_n(1)$ par d_n^{a-1} , ce qui donne une forme linéaire encore plus petite. Pour $a = 4$, on retrouve ainsi les formes linéaires d'Apéry en 1 et $\zeta(3)$ (ce qui n'est pas le cas avec la série bien équilibrée quand $a = 3$).

REMARQUE 2.14. — Pour démontrer le théorème 2.1 on pourrait évaluer les formes linéaires en polylogarithmes en $z = -1$ plutôt qu'en $z = 1$. Ceci induit peu de changements. Le plus notable est que $\log(2) = -\text{Li}_1(-1)$ remplace le divergent $\text{Li}_1(1)$; pour $\ell \geq 2$ on a $\text{Li}_\ell(-1) = -(1 - 2^{1-\ell})\zeta(\ell)$. Pour $a = 3$ et $z = -1$ les formes linéaires construites au paragraphe 2.3 sont [Kr] celles utilisées par Apéry ([Ap1], [Po1]) pour prouver que $\zeta(2)$ est irrationnel. En particulier d_n^2 suffit comme dénominateur des coefficients de cette forme linéaire. Plus généralement, la conjecture 2.6 devrait être valable aussi quand a est impair et $z = -1$.

Considérons l'opérateur différentiel hypergéométrique suivant, où $\delta = z \frac{d}{dz}$:

$$\mathbf{L} = \delta^{a+1} \left(\delta - \frac{n}{2} - 1 \right) \left(\delta - (r+1)n - 1 \right) - z \left(\delta - n \right)^{a+1} \left(\delta - \frac{n}{2} + 1 \right) \left(\delta + rn + 1 \right).$$

L'écriture de $\mathbf{S}_n(z)$ comme série hypergéométrique très bien équilibrée montre que $\mathbf{S}_n(z)$ est une solution de l'équation différentielle $\mathbf{L}y = 0$. Par monodromie on voit, grâce à (22), que pour tout $b \in \{1, \dots, a\}$ la fonction $\sum_{j=b}^a (-1)^{j-1} P_j(z) \frac{\log^{j-b}(z)}{(j-b)!}$ est aussi une solution

de $\mathbf{L}y = 0$. En particulier pour $b = a$ on obtient le polynôme P_a , qu'on peut écrire comme polynôme hypergéométrique très bien équilibré (avec un petit abus de langage : ici les paramètres inférieurs $-\frac{n}{2}$ et $-(r+1)n$ sont négatifs, mais la série ${}_{a+3}F_{a+2}$ est quand même bien définie) :

$$P_a(z) = (-1)^{rn} n!(rn)!((r+1)n)!n!^{-2r-1} \times {}_{a+3}F_{a+2} \left(\begin{matrix} -n, & -\frac{n}{2} + 1, & rn + 1, & -n, & \dots, & -n \\ & -\frac{n}{2}, & -(r+1)n, & 1, & \dots, & 1 \end{matrix} \middle| z \right).$$

L'aspect bien équilibré de ce polynôme hypergéométrique lui confère (voir [And] ou [AAR], §3.5) la propriété de réciprocité (25). En effet, si $y(z)$ est une solution de l'équation différentielle $\mathbf{L}y = 0$ alors $z^n y(1/z)$ est aussi une solution de cette même équation. Quant aux autres polynômes P_{a-1}, \dots, P_1 , ils s'obtiennent par la méthode de Frobenius (voir [Inc]) et vérifient, eux aussi, (25). Toutes ces considérations valent aussi pour la série bien équilibrée de [Ri1] et [BR], et permettent [Hu2] d'écrire celle-ci comme solution unique d'un problème de Padé.

Un autre intérêt des définitions utilisées dans ce texte est que $\mathbf{S}_n(1)$ possède (pour a pair) plusieurs représentations intégrales assez simples. Tout d'abord, on a ([Zu3], Théorème 5) l'intégrale suivante, qui généralise $I_{\kappa,n}(1)$ et les intégrales introduites par Vasilenko [V] et Vasilyev ([Va1], [Va2]) :

$$(26) \quad \mathbf{S}_n(1) = \frac{(rn)!^2}{n!^{2r}} \int_{[0,1]^{a-1}} \frac{\prod_{j=1}^{a-1} x_j^{rn} (1-x_j)^n}{(Q_{a-1}(x_1, \dots, x_{a-1}))^{rn+1}} dx_1 \dots dx_{a-1},$$

en posant $Q_{a-1}(x_1, \dots, x_{a-1}) = 1 - x_1(1 - x_2(\dots(1 - x_{a-1})\dots))$. Vasilyev a démontré [Va2] que si $a = 6$ et $r = 1$ alors cette intégrale s'écrit $\kappa'_0 + \kappa'_3 \zeta(3) + \kappa'_5 \zeta(5)$ avec $d_n^5 \kappa'_0$, $d_n^5 \kappa'_3$ et $d_n^5 \kappa'_5$ entiers. Ceci prouve la conjecture 2.6 dans ce cas. Il n'est pas évident que κ'_0 , κ'_3 et κ'_5 soient les $P_0(1)$, $P_3(1)$ et $P_5(1)$ du paragraphe 2.3, mais cela découle de l'indépendance linéaire conjecturale de 1, $\zeta(3)$ et $\zeta(5)$.

D'autre part, en appliquant à (26) un théorème de Zlobin [Zl] ou le changement de variables qui figure dans [Fi1] (§2) on obtient l'intégrale suivante, qui ressemble à celles utilisées par Sorokin ([So2], [So3]) :

$$\mathbf{S}_n(1) = \frac{(rn)!^2}{n!^{2r}} \int_{[0,1]^{a-1}} \frac{\prod_{j=1}^{a-1} x_j^{rn} (1-x_j)^n dx_j}{(1-x_1 x_2)^{n+1} (1-x_1 x_2 x_3 x_4)^{n+1} \dots (1-x_1 \dots x_{a-2})^{n+1} (1-x_1 \dots x_{a-1})^{n+1}}.$$

Il serait intéressant d'arriver à démontrer le théorème 2.1 en utilisant seulement des intégrales multiples comme celle-ci (ou celle de (26)). Le problème est qu'a priori on s'attend à ce qu'une telle intégrale $(a-1)$ -uple soit une forme linéaire, à coefficients rationnels, en les polyzêtas de poids au plus $(a-1)$ (voir [Wa] et [Zl], Théorème 3). Or le théorème 5 de [Zu3] montre que ces intégrales sont égales à $\mathbf{S}_n(1)$, donc seuls 1 et les valeurs de ζ aux entiers impairs apparaissent.

3. RÉSULTATS QUANTITATIFS

3.1. Exposant d'irrationalité de $\zeta(3)$

On appelle *exposant d'irrationalité* d'un nombre réel irrationnel α , et on note $\mu(\alpha)$, la borne inférieure de l'ensemble des réels ν pour lesquels il n'existe qu'un nombre fini de nombres rationnels p/q tels que $|\alpha - \frac{p}{q}| < \frac{1}{q^\nu}$. La théorie des fractions continues ([HW], §11.1), ou le principe des tiroirs de Dirichlet ([HW], §11.3), montre qu'un exposant d'irrationalité est toujours supérieur ou égal à 2. Si α est algébrique, Liouville a démontré ([Li] ; voir aussi [HW], §11.7) que $\mu(\alpha)$ est inférieur ou égal au degré de α . Ce résultat a été amélioré par Roth en 1955 : on a $\mu(\alpha) = 2$ pour tout nombre algébrique irrationnel α (voir [FN], Chapitre 1, §7). On a aussi $\mu(\alpha) = 2$ pour presque tout réel α , au sens de la mesure de Lebesgue ([HW], §11.11). À l'opposé, un nombre de Liouville est un nombre dont l'exposant d'irrationalité est infini : il est extrêmement bien approché par des nombres rationnels (un exemple de tel nombre est $\sum_{k \geq 1} \frac{1}{10^{k!}}$).

Les formes linéaires d'Apéry montrent que l'exposant d'irrationalité de $\zeta(3)$ est majoré par 13,4179 (voir [FN], Chapitre 2, §5.6) ; en particulier $\zeta(3)$ n'est pas un nombre de Liouville. Ce résultat a été amélioré notamment par Hata [Hat3] puis Rhin-Viola, qui ont démontré la meilleure majoration de $\mu(\zeta(3))$ connue à ce jour :

THÉORÈME 3.1 ([RV]). — *L'exposant d'irrationalité de $\zeta(3)$ est majoré par 5,5139, c'est-à-dire qu'il n'existe qu'un nombre fini de nombres rationnels p/q tels que*

$$|\zeta(3) - \frac{p}{q}| < \frac{1}{q^{5,5139}}.$$

Pour obtenir ce résultat, Rhin et Viola considèrent les intégrales suivantes :

$$(27) \quad J_n = \int_0^1 \int_0^1 \int_0^1 \frac{u^{hn}(1-u)^{ln}v^{kn}(1-v)^{sn}w^{jn}(1-w)^{qn}}{(1-w(1-uv))^{(q+h-r)n+1}} du dv dw,$$

où h, \dots, s sont des paramètres dont on fixe les valeurs de la manière suivante : $h = 16$, $j = 17$, $k = 19$, $l = 15$, $q = 11$, $r = 9$, $s = 13$. Si on prenait tous ces paramètres égaux à un même entier, on obtiendrait les intégrales du paragraphe 1.3, donc la suite des formes linéaires d'Apéry (ou, plus précisément, une suite extraite), conduisant à la même mesure d'irrationalité. L'intérêt réside donc dans le fait de ne pas prendre tous les paramètres égaux ; l'asymptotique obtenue pour $J_n^{1/n}$ est un peu moins bonne, mais on gagne beaucoup sur les dénominateurs par lesquels il faut multiplier J_n pour obtenir une forme linéaire en 1 et $\zeta(3)$ à coefficients entiers. Ce gain provient de l'action sur des intégrales de la forme (27) d'un groupe isomorphe au produit semi-direct $H \circ \underset{\cong}{\cong} 5$, où H est l'hyperplan d'équation $\varepsilon_1 + \dots + \varepsilon_5 = 0$ dans $(\mathbf{Z}/2\mathbf{Z})^5$. D'autres interprétations de cette action de groupe se trouvent dans [Zu4] et [Fi2].

REMARQUE 3.2. — Les majorations de $\mu(\zeta(3))$ mentionnées ci-dessus sont effectives : on peut donner une majoration explicite de la hauteur $\max(|p|, |q|)$ des approximations

rationnelles p/q “exceptionnellement bonnes”. Ceci contraste avec le théorème de Roth, dans lequel on sait seulement majorer le nombre d’exceptions p/q , mais pas leur hauteur.

3.2. Irrationalité d’un nombre parmi $\zeta(5), \dots, \zeta(21)$

Soit a un entier pair, avec $a \geq 6$. Dans ce paragraphe, on construit (en suivant [Ri3]) des formes linéaires à coefficients rationnels en $1, \zeta(5), \zeta(7), \dots, \zeta(a+1)$. Si, après multiplication par un dénominateur commun des coefficients, elles tendent vers zéro sans être nulles à partir d’un certain rang, alors l’un au moins des nombres $\zeta(5), \zeta(7), \dots, \zeta(a+1)$ est irrationnel ; c’est ce qui va se produire avec $a = 20$. On pose :

$$\overline{\mathbf{R}}_n(k) = n!^{a-6} \left(k + \frac{n}{2}\right) \frac{(k-n)_n^3 (k+n+1)_n^3}{(k)_{n+1}^a}$$

et

$$\overline{\mathbf{S}}_n(z) = \frac{1}{2} \sum_{k=1}^{\infty} \overline{\mathbf{R}}_n'(k) z^{-k}.$$

On développe $\overline{\mathbf{R}}_n$ en éléments simples, ce qui définit des coefficients $\overline{c}_{i,j}$ (les formules (20) et (21) restant valables). On définit $\overline{P}_1, \dots, \overline{P}_a$ à partir des $\overline{c}_{i,j}$ par la relation (24) ; seul \overline{P}_0 est défini par une formule légèrement différente :

$$\overline{P}_0(z) = - \sum_{\ell=0}^{n-1} \left(\sum_{i=\ell+1}^n \sum_{j=1}^a \frac{j(j+1)\overline{c}_{i,j}}{2(i-\ell)^{j+2}} \right) z^{\ell}.$$

On obtient la décomposition suivante exactement comme au paragraphe 2.3.2, mais un décalage se produit car on dérive $\overline{\mathbf{R}}_n$ (voir le paragraphe 1.4) :

$$\overline{\mathbf{S}}_n(z) = \overline{P}_0(z) + \sum_{j=1}^a \frac{j(j+1)}{2} \overline{P}_j(z) \operatorname{Li}_{j+2}(1/z).$$

Les arguments du paragraphe 2.3.3 restent valables, et montrent (car a est pair) que $\overline{\mathbf{S}}_n(1)$ est une forme linéaire à coefficients rationnels en $1, \zeta(5), \zeta(7), \dots, \zeta(a+1)$. De plus un dénominateur commun pour ces coefficients est $2d_n^{a+2}$; on conjecture ([Ri2], §5.1) que $2d_n^{a+1}$ convient aussi. La majoration de ces coefficients (qui est effectuée au paragraphe 2.3.4) est inutile ici : elle servait à appliquer le critère de Nesterenko, dont on n’a pas besoin puisqu’on applique seulement la remarque évidente qu’une forme linéaire, à coefficients entiers, en des rationnels fixés ne peut pas être arbitrairement petite sans être nulle.

Le point délicat de la preuve est l’estimation asymptotique de $\overline{\mathbf{S}}_n(1)$. En effet, on ne connaît pas d’écriture de $\overline{\mathbf{S}}_n(1)$ comme intégrale multiple réelle. On utilise donc la méthode du col. Posons

$$K_n(u) = \frac{-1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \overline{\mathbf{R}}_n(s) \left(\frac{\pi}{\sin(\pi s)} \right)^3 e^{us} ds,$$

où c est un réel avec $0 < c < n + 1$, et u un nombre complexe tel que $\operatorname{Re}(u) \leq 0$ et $|\operatorname{Im}(u)| < 3\pi$. Cette intégrale est à rapprocher de celle notée $I_{c,n}(z)$ au paragraphe 1.5. On peut appliquer le théorème des résidus, pour faire apparaître les pôles de l'intégrande qui sont situés aux entiers $n + 1, n + 2, \dots$. Au voisinage d'un tel entier k , on a $(\frac{\pi}{\sin(\pi s)})^3 = \frac{(-1)^k}{(s-k)^3} + \frac{(-1)^k \pi^2}{2(s-k)} + O(s-k)$. On obtient donc (voir [He] et [Zu2] pour des résultats analogues) :

$$K_n(u) = \frac{\pi^2 + u^2}{2} \sum_{k=n+1}^{\infty} \overline{\mathbf{R}}_n(k)(-e^u)^k + u \sum_{k=n+1}^{\infty} \overline{\mathbf{R}}'_n(k)(-e^u)^k + \frac{1}{2} \sum_{k=n+1}^{\infty} \overline{\mathbf{R}}''_n(k)(-e^u)^k.$$

En choisissant $u = i\pi$, le premier terme disparaît, et on obtient $\overline{\mathbf{S}}_n(1) = \operatorname{Re}(K_n(i\pi))$.

La méthode du col donne ([Ri3], Lemme 5) deux nombres complexes non nuls c_0 et α , qu'on peut calculer, tels que $K_n(i\pi) \sim c_0 n^{-8} e^{\alpha n}$ quand n tend vers l'infini. Comme la partie imaginaire de α n'est pas un multiple entier de π , il existe une suite strictement croissante $\varphi(n)$ d'entiers tels que l'argument de $c_0 e^{\alpha \varphi(n)}$, vu modulo 2π , ait une limite autre que $\pm\pi/2$. On a alors :

$$\lim_{n \rightarrow \infty} |\overline{\mathbf{S}}_{\varphi(n)}(1)|^{1/\varphi(n)} = e^{\operatorname{Re}(\alpha)}.$$

Le choix $a = 20$ donne $\operatorname{Re}(\alpha) = -22,02\dots$ d'où $\operatorname{Re}(\alpha) + a + 2 < 0$. Donc la forme linéaire $d_{\varphi(n)}^{22} \overline{\mathbf{S}}_{\varphi(n)}(1)$ en $1, \zeta(5), \zeta(7), \dots, \zeta(21)$, à coefficients entiers, tend vers 0 quand n tend vers l'infini et est non nulle pour n assez grand. Cela montre que l'un au moins parmi $\zeta(5), \zeta(7), \dots, \zeta(21)$ est irrationnel.

REMARQUE 3.3. — Si on savait démontrer la conjecture mentionnée ci-dessus (i.e. que $2d_n^{a+1} \overline{\mathbf{P}}_j(1)$ est un entier pour tout j), on pourrait ([Ri2], §5.1) appliquer la même méthode avec $a = 18$, et démontrer ainsi que l'un au moins des nombres $\zeta(5), \zeta(7), \dots, \zeta(19)$, est irrationnel.

3.3. Irrationalité d'un nombre parmi $\zeta(5), \zeta(7), \zeta(9)$ et $\zeta(11)$

La structure de la preuve est la même que dans le paragraphe précédent. La différence principale vient de dénominateurs nettement plus petits, grâce à une étude fine de leurs valuations p -adiques et à l'utilisation d'une fraction rationnelle modifiée :

$$\tilde{\mathbf{R}}_n(k) = \frac{\prod_{u=1}^{10} ((13 + 2u)n)!}{(27n)!^{16}} (37n + 2k) \frac{(k - 27n)_{27n}^3 (k + 37n + 1)_{27n}^3}{\prod_{u=1}^{10} (k + (12 - u)n)_{(13+2u)n+1}}.$$

Pour $|z| \geq 1$ on pose $\tilde{\mathbf{S}}_n(z) = \frac{1}{2} \sum_{k=1}^{\infty} \tilde{\mathbf{R}}_n''(k) z^{-k}$. La décomposition en éléments simples $\tilde{\mathbf{R}}_n(k) = \sum_{j=1}^{10} \sum_{i=(j+1)n}^{(36-j)n} \frac{\tilde{c}_{i,j}}{(k+i)^j}$ définit les $\tilde{c}_{i,j}$ à partir desquels on construit les polynômes $\tilde{P}_j(z) = \sum_{i=(j+1)n}^{(36-j)n} \tilde{c}_{i,j} z^i$ pour $j \in \{1, 2, \dots, 10\}$ et

$$\tilde{P}_0(z) = - \sum_{\ell=0}^{35n-1} \left(\sum_{j=1}^{10} \sum_{i=\max((j+1)n, \ell+1)}^{(36-j)n} \frac{j(j+1)\tilde{c}_{i,j}}{2(i-\ell)^{j+2}} \right) z^{\ell}.$$

On a alors $\tilde{\mathbf{S}}_n(z) = \tilde{P}_0(z) + \sum_{j=1}^{10} \frac{j(j+1)}{2} \tilde{P}_j(z) \text{Li}_{j+2}(1/z)$.

Le problème est de majorer de façon très précise le dénominateur des rationnels $\tilde{c}_{i,j}$. En suivant la méthode utilisée pour démontrer le lemme 2.12, on obtiendrait $d_{33n}^{10-j} \tilde{c}_{i,j} \in \mathbf{Z}$ pour tous i et j . Une étude fine de la valuation p -adique des coefficients binomiaux permet d'obtenir un dénominateur nettement plus petit : on trouve un entier Φ_n "assez grand" tel que $d_{33n}^{10-j} \Phi_n^{-1} \tilde{c}_{i,j} \in \mathbf{Z}$. On en déduit directement que $2d_{35n}^3 d_{34n}^8 d_{33n}^8 \Phi_n^{-1} \tilde{P}_j(z)$ est à coefficients entiers pour tout $j \in \{0, 1, \dots, 10\}$.

La symétrie $\tilde{\mathbf{R}}_n(-37n - k) = -\tilde{\mathbf{R}}_n(k)$ donne $z^{37n} \tilde{P}_j(1/z) = (-1)^{j+1} \tilde{P}_j(z)$, d'où $\tilde{P}_j(1) = 0$ pour $j = 2, 4, \dots, 10$. En outre on a $\tilde{P}_1(1) = 0$ car $\tilde{\mathbf{R}}_n(k) = O(k^{-2})$ quand k tend vers l'infini. Donc $\tilde{\mathbf{S}}_n(1)$ est une forme linéaire en $1, \zeta(5), \zeta(7), \zeta(9)$ et $\zeta(11)$. Pour l'estimer, et démontrer qu'elle est non nulle pour une infinité de n , on transforme $\tilde{\mathbf{S}}_n(1)$ en une intégrale complexe, à laquelle on applique la méthode du col (voir [Zu2], §2). On obtient les comportements asymptotiques suivants quand n tend vers l'infini : $\limsup |\tilde{\mathbf{S}}_n(1)|^{1/n} \leq e^{-227,58\dots}$, $\limsup |\Phi_n^{-1}|^{1/n} \leq e^{-176,75\dots}$ et $(d_{35n}^3 d_{34n}^8 d_{33n}^8)^{1/n} \rightarrow e^{403}$. Comme $403 < 227,58 + 176,75$ on obtient la conclusion cherchée.

REMARQUE 3.4. — Zudilin conjecture ([Zu4], §9) que des compensations ont lieu quand $z = 1$, ce qui permettrait de trouver un dénominateur plus petit pour les $P_j(1)$. Peut-être pourrait-on alors démontrer que parmi $\zeta(5), \zeta(7)$ et $\zeta(9)$ l'un au moins est irrationnel.

REMARQUE 3.5. — En utilisant des méthodes similaires, on peut démontrer [Zu2] que pour tout $\ell \geq 1$ impair l'un au moins des nombres $\zeta(\ell + 2), \zeta(\ell + 4), \dots, \zeta(8\ell - 1)$, est irrationnel.

BIBLIOGRAPHIE

- [AO] S. AHLGREN et K. ONO – *A Gaussian hypergeometric series evaluation and Apéry number congruences*, J. Reine Angew. Math. 518 (2000), 187-212.
- [AG] G. ALMKVIST et A. GRANVILLE – *Borwein and Bradley's Apéry-like formulae for $\zeta(4n + 3)$* , Experiment. Math. 8.2 (1999), 197-203.
- [An] Y. ANDRÉ – *G-functions and geometry*, Aspects of Math. E13, Vieweg, 1989.
- [AnJ] R. ANDRÉ-JEANNIN – *Irrationalité de la somme des inverses de certaines suites récurrentes*, C. R. Acad. Sci. Paris, Ser. I 308 (1989), 539-541.
- [And] G.E. ANDREWS – *The well-poised thread: an organized chronicle of some amazing summations and their implications*, Ramanujan J. 1.1 (1997), 7-23.
- [AAR] G.E. ANDREWS, R. ASKEY et R. ROY – *Special Functions*, The Encyclopedia of Mathematics and its Applications 71 (G.-C. Rota ed.), Cambridge University Press, Cambridge, 1999.

- [Ap1] R. APÉRY – *Irrationalité de $\zeta(2)$ et $\zeta(3)$* , in : Journées Arithmétiques (Luminy, 1978), Astérisque 61 (1979), 11-13.
- [Ap2] R. APÉRY – *Interpolation de fractions continues et irrationalité de certaines constantes*, in: Comité des Travaux Historiques et Scientifiques (CTHS), Bulletin de la Section des Sciences III (Mathématiques), Bibliothèque Nationale, Paris, 1981, 37-53.
- [AW] R. ASKEY et J.A. WILSON – *A recursive relation generalizing those of Apéry*, J. Austral. Math. Soc. 36 (1984), 267-278.
- [As] W. VAN ASSCHE – *Approximation theory and analytic number theory*, in: Special Functions and Differential Equations (Madras, 1997), Allied Publishers, New Delhi, 1998, 336-355.
- [BR] K.M. BALL et T. RIVOAL – *Irrationalité d'une infinité de valeurs de la fonction zêta aux entiers impairs*, Invent. Math. 146.1 (2001), 193-207.
- [BO] C. BATUT et M. OLIVIER – *Sur l'accélération de la convergence de certaines fractions continues*, Sémin. de Théorie des Nombres de Bordeaux 1979-1980, exp. no. 23 (25 p.).
- [Be1] F. BEUKERS – *A note on the irrationality of $\zeta(2)$ and $\zeta(3)$* , Bull. London Math. Soc. 11.3 (1979), 268-272.
- [Be2] F. BEUKERS – *Padé-approximations in number theory*, in: Padé approximation and its applications (Amsterdam, 1980), Lecture Notes in Math. 888, Springer, 1981, 90-99.
- [Be3] F. BEUKERS – *The values of polylogarithms*, in: Topics in classical number theory (Budapest, 1981), Colloq. Math. Soc. János Bolyai 34, 1984, 219-228.
- [Be4] F. BEUKERS – *Irrationality of π^2 , periods of an elliptic curve and $\Gamma_1(5)$* , in: Approximations diophantiennes et nombres transcendants (Luminy, 1982), D. Bertrand et M. Waldschmidt eds., Progress in Math. 31, Birkhäuser, 1983, 47-66.
- [Be5] F. BEUKERS – *Some congruences for the Apéry numbers*, J. Number Th. 21 (1985), 141-155.
- [Be6] F. BEUKERS – *Irrationality proofs using modular forms*, in: Journées Arithmétiques (Besançon, 1985), Astérisque 147-148 (1987), 271-283.
- [Be7] F. BEUKERS – *Another Congruence for the Apéry Numbers*, J. Number Th. 25 (1987), 201-210.
- [BP] F. BEUKERS et C.A.M. PETERS – *A family of K3 surfaces and $\zeta(3)$* , J. Reine Angew. Math. 351 (1984), 42-54.
- [BB] J. BORWEIN et D. BRADLEY – *Empirically determined Apéry-like formulae for $\zeta(4n + 3)$* , Experiment. Math. 6 (1997), 181-194.
- [BE] P. BORWEIN et T. ERDÉLYI – *Polynomials and Polynomial inequalities*, Graduate Texts in Math. 161, Springer, 1995.

- [BV] P. BUNDSCHUH et K. VÄÄNÄNEN – *Arithmetical investigations of a certain infinite product*, Compositio Math. 91 (1994), 175-199.
- [Ca1] P. CARTIER – *Démonstration automatique d'identités et fonctions hypergéométriques (d'après Zeilberger)*, Sémin. Bourbaki 1991-92, exp. no. 746, Astérisque 206 (1992), 41-91.
- [Ca2] P. CARTIER – *Fonctions polylogarithmes, nombres polyzêtas et groupes pro-unipotents*, Sémin. Bourbaki 2000-01, exp. no. 885, à paraître dans Astérisque.
- [CCC] S. CHOWLA, J. COWLES et M. COWLES – *Congruence properties of Apéry numbers*, J. Number Th. 12 (1980), 188-190.
- [Ch] G.V. CHUDNOVSKY – *Transcendental numbers*, in: Number theory, Proc. Southern Illinois Conf. (Carbondale, 1979), Lecture Notes in Math. 751, Springer, 45-69.
- [Coh1] H. COHEN – *Démonstration de l'irrationalité de $\zeta(3)$ (d'après Apéry)*, Sémin. de Théorie des Nombres de Grenoble, octobre 1978 (9 p.).
- [Coh2] H. COHEN – *Généralisation d'une construction de R. Apéry*, Bull. Soc. Math. France 109 (1981), 269-281.
- [Col] P. COLMEZ – *Arithmétique de la fonction zêta*, Journées X-UPS 2002, à paraître.
- [Di] J. DIEUDONNÉ – *Calcul infinitésimal*, Collection Méthodes, Hermann, 1968.
- [Dw1] B. DWORK – *On Apéry's differential operator*, Groupe d'étude d'analyse ultramétrique 1979-1981, exp. no. 25 (6 p.).
- [Dw2] B. DWORK – *Arithmetic theory of differential equations*, Symposia Math. 24 (INDAM, Rome, 1979), Academic Press, 1981, 225-243.
- [Dw3] B. DWORK, G. GEROTTO et F.J. SULLIVAN – *An introduction to G-functions*, Annals of Math. Studies 133, Princeton Univ. Press, 1994.
- [FN] N.I. FEL'DMAN et YU.V. NESTERENKO – *Number Theory IV, Transcendental Numbers*, A.N. Parshin et I.R. Shafarevich eds., Encyclopaedia of Mathematical Sciences 44, Springer, 1998.
- [Fi1] S. FISCHLER – *Formes linéaires en polyzêtas et intégrales multiples*, C. R. Acad. Sci. Paris, Ser. I 335.1 (2002), 1-4.
- [Fi2] S. FISCHLER – *Groupes de Rhin-Viola et intégrales multiples*, Actes des Rencontres Arithmétiques de Caen (juin 2001), soumis au J. Théor. Nombres Bordeaux.
- [FR] S. FISCHLER et T. RIVOAL – *Approximants de Padé et séries hypergéométriques équilibrées*, Rapport de recherche 2002-27, LMNO, Université de Caen ; soumis.
- [Gel] A.O. GEL'FOND – *Calcul des différences finies*, Dunod, 1963.
- [Ges] I. GESSEL – *Some congruences for Apéry numbers*, J. Number Th. 14 (1982), 362-368.
- [Gu1] L.A. GUTNIK – *The irrationality of certain quantities involving $\zeta(3)$* , Uspekhi Mat. Nauk [Russian Math. Surveys] 34.3 (1979), 190 [200].

- [Gu2] L.A. GUTNIK – *On the irrationality of some quantities containing $\zeta(3)$* , Acta Arith. 42.3 (1983), 255-264 (en russe) ; traduction dans Amer. Math. Soc. Transl. 140 (1988), 45-55.
- [Hab] L. HABSIEGER – *Introduction to diophantine approximation*, en préparation.
- [HW] G.H. HARDY et E.M. WRIGHT – *An introduction to the theory of numbers*, 3ème édition, Oxford Univ. Press, 1954.
- [Hat1] M. HATA – *On the linear independence of the values of polylogarithmic functions*, J. Math. Pures Appl. 69.2 (1990), 133-173.
- [Hat2] M. HATA – *Rational approximations to the dilogarithm*, Trans. Amer. Math. Soc. 336.1 (1993), 363-387.
- [Hat3] M. HATA – *A new irrationality measure for $\zeta(3)$* , Acta Arith. 92.1 (2000), 47-57.
- [Haz] M. HAZEWINKEL – *Formal groups and applications*, Pure and Applied Mathematics 78, Academic Press, 1978.
- [He] T.G. HESSAMI PILEHROOD – *Linear independence of vectors with polylogarithmic coordinates*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. no. 6 [Moscow Univ. Math. Bull. 54.6] (1999), 54-56 [40-42].
- [Hu1] M. HUTTNER – *Équations différentielles fuchsiennes. Approximations du dilogarithme, de $\zeta(2)$ et de $\zeta(3)$* , Pub. IRMA Lille 43 (1997).
- [Hu2] M. HUTTNER – *Constructible sets of linear differential equations and effective rational approximations of G -functions*, Pub. IRMA Lille 59 (2002).
- [Inc] E.L. INCE – *Ordinary differential equations*, Dover Publ., 1926.
- [Ing] A.E. INGHAM – *The distribution of prime numbers*, Cambridge Univ. Press, 1932.
- [Is] T. ISHIKAWA – *On Beukers' conjecture*, Kobe J. Math. 6 (1989), 49-52.
- [Ko] M. KOECHER – *Letter*, Math. Intelligencer 2 (1980), 62-64.
- [Kr] C. KRATTENTHALER – Communication personnelle, 28 Octobre 2002.
- [La] S. LANG – *Algebra*, 3ème édition, Addison-Wesley, 1993.
- [Le] D. LESHCHINER – *Some new identities for $\zeta(k)$* , J. Number Th. 13 (1981), 355-362.
- [Li] J. LIOUVILLE – *Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques*, J. Math. Pures Appl. 16 (1851), 133-142.
- [Lu] Y.L. LUKE – *The special functions and their approximations*, Volume I, Mathematics in Science and Engineering 53, Academic Press, 1969.
- [Me] M. MENDÈS-FRANCE – *Roger Apéry et l'irrationnel*, La Recherche 97 (1979), 170-172.
- [Ne1] YU.V. NESTERENKO – *On the linear independence of numbers*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. no. 1 [Moscow Univ. Math. Bull. 40.1] (1985), 46-49 [69-74].

- [Ne2] YU.V. NESTERENKO – *A few remarks on $\zeta(3)$* , Mat. Zametki [Math. Notes] 59.6 (1996), 865-880 [625-636].
- [Ne3] YU.V. NESTERENKO – *Integral identities and constructions of approximations to zeta-values*, Actes des Rencontres Arithmétiques de Caen (juin 2001), soumis au J. Théor. Nombres Bordeaux.
- [Ni] E.M. NIKISHIN – *On the irrationality of the values of the functions $F(x,s)$* , Mat. Sbornik 109.3 [Math. USSR-Sb. 37.3] (1979), 410-417 [381-388].
- [NS] E.M. NIKISHIN et V.N. SOROKIN – *Rational approximations and orthogonality*, Translations of Math. Monographs 92, Amer. Math. Soc., 1991.
- [NZM] I. NIVEN, H.S. ZUCKERMAN et H.L. MONTGOMERY – *An introduction to the theory of numbers*, 5ème édition, J. Wiley, 2000.
- [Oe] J. OESTERLÉ – *Polylogarithmes*, Sémin. Bourbaki 1992-93, exp. no. 762, Astérisque 216 (1993), 49-67.
- [PS] C. PETERS et J. STIENSTRA – *A pencil of $K3$ -surfaces related to Apéry's recurrence for $\zeta(3)$ and Fermi surfaces for potential zero*, in: Arithmetics of complex manifolds (Erlangen, 1988), W.P. Barth et H. Lange eds., Lecture Notes in Math. 1399, Springer, 110-127.
- [PWZ] M. PETKOVŠEK, H.S. WILF et D. ZEILBERGER – *$A=B$* , A.K. Peters, 1996.
- [Po1] A. VAN DER POORTEN – *A proof that Euler missed... Apéry's proof of the irrationality of $\zeta(3)$* , Math. Intelligencer 1.4 (1978/79), 195-203.
- [Po2] A. VAN DER POORTEN – *Some wonderful formulae... footnotes to Apéry's proof of the irrationality of $\zeta(3)$* , Sémin. Delange-Pisot-Poitou, 20e année, 1978-79, no. 29 (7p.).
- [Po3] A. VAN DER POORTEN – *Some wonderful formulas ... an introduction to polylogarithms*, in: Proceedings of the Queen's Number Theory Conference (Kingston, 1979), Queen's Papers in Pure and Applied Mathematics 54, 1980, 269-286.
- [Pr1] M. PREVOST – *A new proof of the irrationality of $\zeta(2)$ and $\zeta(3)$ using Padé approximants*, J. Comp. Appl. Math. 67 (1996), 219-235.
- [Pr2] M. PREVOST – *On the irrationality of $\sum \frac{t^n}{A\alpha^n + B\beta^n}$* , J. Number Th. 73 (1998), 139-161.
- [Re1] E. REYSSAT – *Irrationalité de $\zeta(3)$ selon Apéry*, Sémin. Delange-Pisot-Poitou, 20e année, 1978-79, no. 6 (6 p.).
- [Re2] E. REYSSAT – *Mesures de transcendance pour les logarithmes de nombres rationnels*, in: Approximations diophantiennes et nombres transcendants (Luminy, 1982), D. Bertrand et M. Waldschmidt eds., Progress in Math. 31, Birkhäuser, 1983, 47-66.
- [RV] G. RHIN et C. VIOLA – *The group structure for $\zeta(3)$* , Acta Arith. 97.3 (2001), 269-293.

- [Ri1] T. RIVOAL – *La fonction zêta de Riemann prend une infinité de valeurs irrationnelles aux entiers impairs*, C. R. Acad. Sci. Paris, Sér. I 331.4 (2000), 267-270.
- [Ri2] T. RIVOAL – *Propriétés diophantiennes des valeurs de la fonction zêta de Riemann aux entiers impairs*, thèse de doctorat, Univ. de Caen (2001). Disponible sur <http://theses-EN-ligne.in2p3.fr>.
- [Ri3] T. RIVOAL – *Irrationalité d'au moins un des neuf nombres $\zeta(5)$, $\zeta(7)$, ..., $\zeta(21)$* , Acta Arith. 103.2 (2002), 157-167.
- [Ri4] T. RIVOAL – *Séries hypergéométriques et irrationalité des valeurs de la fonction zêta de Riemann*, Actes des Journées Arithmétiques de Lille (juillet 2001), à paraître au J. Théor. Nombres Bordeaux.
- [RZ] T. RIVOAL et W. ZUDILIN – *Diophantine properties of numbers related to Catalan's constant*, Prépublication 315 de l'Institut de Mathématiques de Jussieu (janvier 2002), Univ. Paris 6, soumis.
- [Se] J.P. SERRE – *Cours d'arithmétique*, Presses Univ. de France, 1970.
- [Sl] L.J. SLATER – *Generalized hypergeometric functions*, Cambridge Univ. Press, 1966.
- [So1] V.N. SOROKIN – *Hermite-Padé approximations for Nikishin systems and the irrationality of $\zeta(3)$* , Uspekhi Mat. Nauk [Russian Math. Surveys] 49.2 (1994), 167-168 [176-177].
- [So2] V.N. SOROKIN – *A transcendence measure for π^2* , Mat. Sbornik [Sb. Math.] 187.12 (1996), 87-120 [1819-1852].
- [So3] V.N. SOROKIN – *Apéry's theorem*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. no. 3 [Moscow Univ. Math. Bull. 53.3] (1998), 48-53 [48-52].
- [SB] J. STIENSTRA et F. BEUKERS – *On the Picard-Fuchs equation and the formal Brauer group of certain elliptic K3-surfaces*, Math. Ann. 271 (1985), 269-304.
- [Su] B. SURY – *On a conjecture of Chowla et al.*, J. Number Th. 72 (1998), 137-139.
- [V] O.N. VASILENKO – *Certain formulae for values of the Riemann zeta function at integral points*, in: Number theory and its applications, Proceedings of the science-theoretical conference (Tashkent, 1990), p. 27 (en russe).
- [Va1] D.V. VASILYEV – *Some formulas for Riemann zeta-function at integer points*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. no. 1 [Moscow Univ. Math. Bull. 51.1] (1996), 81-84 [41-43].
- [Va2] D.V. VASILYEV – *On small linear forms for the values of the Riemann zeta-function at odd integers* (en russe), Doklady NAN Belarusi (Reports of the Belarus National Academy of Sciences) 45.5 (2001), 36-40.
- [Wa] M. WALDSCHMIDT – *Valeurs zêta multiples : une introduction*, J. Théor. Nombres Bordeaux 12.2 (2000), 581-595.
- [We] A. WEIL – *Remarks on Hecke's lemma and its use*, in: Oeuvres scientifiques - Collected Papers III, Springer, 1979, 405-412.

- [WW] E.T. WHITTAKER et G.N. WATSON – *A course of modern analysis*, 4ème édition, Cambridge Univ. Press, 1927.
- [Za1] D. ZAGIER – *Introduction to modular forms*, in: From number theory to physics (Les Houches, 1989), M. Waldschmidt, P. Moussa, J.M. Luck et C. Itzykson eds., Springer, 1992, 238-291.
- [Za2] D. ZAGIER – Cours au Collège de France, mai 2001.
- [Ze1] D. ZEILBERGER – *Closed form (pun intended !)*, in: A tribute to Emil Grosswald: Number theory and related analysis, Contemporary Math. 143 (M. Knopp et M. Sheingorn eds.), Amer. Math. Soc., 1993, 579-607.
- [Ze2] D. ZEILBERGER – *Computerized deconstruction*, à paraître dans Adv. Applied Math.
- [Zl] S.A. ZLOBIN – *Integrals expressible as linear forms in generalized polylogarithms*, Mat. Zametki [Math. Notes] 71.5 (2002), 782-787 [711-716].
- [Zu1] W. ZUDILIN – *One of the numbers $\zeta(5)$, $\zeta(7)$, $\zeta(9)$, $\zeta(11)$ is irrational*, Uspekhi Mat. Nauk [Russian Math. Surveys] 56.4 (2001), 149-150 [774-776].
- [Zu2] W. ZUDILIN – *Irrationality of values of the Riemann zeta function*, Izvestiya RAN Ser. Mat. [Izv. Math.] 66.3 (2002), 49-102 [489-542].
- [Zu3] W. ZUDILIN – *Well-poised hypergeometric series for diophantine problems of zeta values*, Actes des Rencontres Arithmétiques de Caen (juin 2001), soumis au J. Théor. Nombres Bordeaux.
- [Zu4] W. ZUDILIN – *Arithmetic of linear forms involving odd zeta values*, preprint, math.NT/0206176.
- [Zu5] W. ZUDILIN – *An elementary proof of Apéry's theorem*, preprint, math.NT/0202159.

Stéphane FISCHLER

École Normale Supérieure

Département de Mathématiques

et Applications

UMR 8553 du CNRS

45, rue d'Ulm

F-75230 Paris Cedex 05

E-mail : fischler@dma.ens.fr

**EXEMPLES D'INSTABILITÉS
POUR DES ÉQUATIONS D'ONDES NON LINÉAIRES**
[d'après G. Lebeau]

par **Guy MÉTIVIER**

1. INSTABILITÉ DES ÉQUATIONS D'ONDES SURCRITIQUES

Dans l'étude des équations hyperboliques non linéaires, l'équation d'onde

$$(1) \quad (\partial_t^2 - \Delta_x)u + |u|^{p-1}u = 0, \quad u|_{t=0} = u_0, \quad \partial_t u|_{t=0} = u_1,$$

est un modèle de base pour l'analyse mathématique. On s'intéresse ici au cas de la dimension trois ($x \in \mathbb{R}^3$) et, pour simplifier, on se limite au cas où p est un entier impair ($|u|^{p-1}u = u^p$). En multipliant l'équation par $\partial_t u$ et en intégrant par parties, on voit que, formellement, on a conservation de l'énergie $E(u(t)) := \mathcal{E}(u(t), \partial_t u(t))$ avec :

$$(2) \quad \mathcal{E}(u_0, u_1) := \int_{\mathbb{R}^3} \left(\frac{1}{2}|u_1|^2 + \frac{1}{2}|\nabla_x u_0|^2 + \frac{u_0^{p+1}}{p+1} \right) dx.$$

L'existence et l'unicité de solutions fortes du problème de Cauchy est bien connue dans le cas sous-critique $p \leq 3$ et dans le cas critique $p = 5$ (voir par exemple [14], [17], [6], [4], [16], [12], [1]). En particulier, le semi-groupe solution $S_t : (u(0), \partial_t u(0)) \mapsto (u(t), \partial_t u(t))$ est uniformément Lipschitzien sur les bornés de $\dot{H}^1 \times L^2$.

Dans le cas surcritique ($p \geq 7$), on dispose simplement d'un théorème d'existence et d'unicité locales des solutions régulières et, pour des données de Cauchy dans l'espace d'énergie, d'un théorème d'existence globale sans unicité de solutions faibles bornées dans l'espace d'énergie, i.e. vérifiant $E(u(t)) \leq \mathcal{E}(u_0, u_1)$. Un corollaire du travail de G. Lebeau [11] est que le problème de Cauchy (1) surcritique ($p \geq 7$) est mal posé au sens de Hadamard. Compte tenu du théorème local d'existence et d'unicité, l'obstacle vient de l'évolution des singularités. G. Lebeau considère le cas le plus simple d'une singularité ponctuelle (à l'origine) et de solutions radiales, c'est-à-dire invariantes par rotation. Un corollaire du résultat de G. Lebeau est le suivant :

THÉORÈME 1.1. — *Si p est un entier impair supérieur ou égal à 7, il existe des familles de données initiales radiales $\underline{u}^\delta = (u_0^\delta, u_1^\delta)$ et $\underline{v}^\delta = (v_0^\delta, v_1^\delta)$, C^∞ en dehors de l'origine,*

d'énergie bornée par 1, dont la différence est asymptotiquement nulle dans l'espace des fonctions C^∞ plates à l'origine :

$$\lim_{\delta \rightarrow 0} \left\| |x|^{-k} (\underline{u}^\delta - \underline{v}^\delta) \right\|_{H^s \times H^{s-1}} = 0, \quad \forall k, s$$

telles que toutes les solutions faibles radiales u^δ, v^δ de (1) (il en existe) vérifient

$$(3) \quad \liminf_{\delta \rightarrow 0} \left\| u^\delta(t_\delta) - v^\delta(t_\delta) \right\|_{L^{p+1}(\mathbb{R}^3)} > 0,$$

où $t_\delta \rightarrow 0$ quand $\delta \rightarrow 0$.

Ce résultat d'instabilité montre que le problème de Cauchy (1) est mal posé au sens de Hadamard. En supposant que les opérateurs solutions S_t existent, pour tout $T > 0$ arbitrairement petit, la famille $\{S_t\}_{t \in [0, T]}$ n'est pas équi-uniformément continue sur les boules de l'espace d'énergie.

Pour les solutions radiales, en dimension trois, on pose $u(t, x) = rg(t, r)$, avec $r = |x|$. L'équation s'écrit

$$(4) \quad (\partial_t^2 - \partial_r^2)g + \frac{g^p}{r^{p-1}} = 0, \quad g|_{t=0} = g_0, \quad \partial_t g|_{t=0} = g_1.$$

Pour cette équation, tous les p sont sous-critiques (en dehors de $r = 0$) et par vitesse finie de propagation, pour des données initiales lisses dans $\{r > 0\}$, on a existence, unicité et régularité de solutions faibles (radiales) dans $\{t < r\}$ (sous le cône d'onde).

G. Lebeau construit ses solutions u^δ à l'aide de solutions de (4) dont les données initiales g_0 et g_1 sont à support compact, de classe C^∞ en dehors de l'origine, et vérifient

$$(5) \quad g_0(r) \sim r^\gamma (c_0 + c_1 r^\beta + \dots), \quad g_1(r) \sim r^{\gamma-1-\beta} (d_0 + d_1 r^\beta + \dots).$$

Les paramètres β et γ sont tels que :

$$(6) \quad \frac{p-2}{p+1} < \gamma < \frac{p-3}{p-1}, \quad \beta = \frac{p-3}{2} - \gamma \frac{p-1}{2} > 0.$$

Ce choix assure en particulier que l'énergie initiale

$$\mathcal{E}(g_0, g_1) := \int_0^\infty \left(\frac{1}{2} |g_1|^2 + \frac{1}{2} |\partial_r g_0|^2 + \frac{g_0^{p+1}}{p+1} \right) dr$$

est finie.

Étant donné les deux suites de coefficients (c_0, c_1, \dots) et (d_0, d_1, \dots) avec $(c_0, d_0) \neq (0, 0)$, l'idée de G. Lebeau est de construire et contrôler sur un cône $t \leq cr$ (avec $c < 1$ petit) deux solutions g et g' associées à des données initiales (g_0, g_1) et (g'_0, g'_1) vérifiant toutes deux (5), mais qui diffèrent significativement sur une courbe $t = t(r) \ll r$. Plus précisément, on montre que

$$(7) \quad \int_{J_\delta} |(g - g')(t_\delta, r)|^{p+1} \frac{dr}{r^{p-1}} \geq c |J_\delta| \delta^{\gamma(p+1)-(p-1)},$$

où J_δ est un certain intervalle centré en $\delta > 0$ et de longueur $\approx \delta^{2+\beta}/t_\delta \ll \delta$ et t_δ une suite tendant vers 0. On renvoie à [11] pour un énoncé précis.

Pour ce faire, on utilise un changement d'échelle. On introduit les nouvelles variables

$$(8) \quad t = \otimes s, \quad r = \otimes x \quad g(t, r) = \otimes^\gamma f(s, x).$$

Avec $h = \otimes^\beta$, l'équation devient

$$h^2(\partial_s^2 - \partial_x^2)f + \frac{f^{p+1}}{x^{p-1}} = 0 \quad f|_{s=0} = f_0, \quad \partial_s f|_{s=0} = f_1$$

et (5) devient

$$f_0(x) \sim x^\gamma \sum_{k \geq 0} h^k c_k x^{k\beta}, \quad f_1(x) \sim x^{\gamma-1-\beta} \sum_{k \geq 0} h^k d_k x^{k\beta}$$

On se ramène alors à étudier des solutions de (11) dans un voisinage de $s = 0$, $x = 1$. Revenant à la notation habituelle t pour la variable de temps, on s'est donc ramené à étudier au voisinage de $t = 0$, $x = x_0 \neq 0$, un problème de la forme

$$(11) \quad h^2(\partial_t^2 - \partial_x^2)u + \frac{\partial F}{\partial u}(x, u) = 0 \quad u|_{t=0} = u_0, \quad \partial_t u|_{t=0} = u_1$$

$$(12) \quad u_0(x) \sim \sum_{k \geq 0} h^k a_k(x), \quad u_1(x) \sim \sum_{k \geq 0} h^k b_k(x),$$

avec $F(x, u) = \frac{u^{p+1}}{(p+1)x^{p-1}}$. L'objectif est de montrer l'instabilité du problème (11).

THÉORÈME 1.2. — *Il existe une fonction $t(h)$ qui tend vers zéro avec h et des familles de données initiales $\underline{u}^h = (\underline{u}_0^h, \underline{u}_1^h)$ et $\underline{v}^h = (v_0^h, v_1^h)$ bornées dans C^∞ sur l'intervalle $\{|x - x_0| \leq r\}$, telles que les solutions u^δ, v^δ de (11) sont définies sur $\Omega = \{|x - x_0| + t \leq r, 0 \leq t \leq t(h)\}$ et vérifient*

$$(13) \quad \underline{u}^h - \underline{v}^h = O(h^\infty) \quad \text{et} \quad \|(u^h - v^h)|_{t=t(h)}\|_{L^2} \geq c.$$

2. UN PEU D'OPTIQUE GÉOMÉTRIQUE FORTEMENT NON LINÉAIRE

Les résultats classiques sur le problème de Cauchy montrent que (11) possède une unique solution régulière sur un intervalle de longueur $\approx h$. L'objectif serait de construire des solutions sur les temps d'ordre $O(1)$. L'optique géométrique propose de chercher des solutions de la forme

$$(14) \quad u_h(t, x) = U\left(t, x, \frac{\varphi(t, x)}{h}\right), \quad U(t, x, \theta) \sim \sum_{k \geq 0} h^k U_k(t, c, \theta),$$

où les $U_k(t, x, \theta)$ sont des fonctions régulières, 2π -périodiques en θ , et où la phase φ décrit les oscillations rapides dues au facteur $h^2 \partial_t^2$ de l'équation. Comme il n'y a pas d'oscillations initiales on impose $\varphi(0, x) \equiv 0$.

Une littérature très abondante est consacrée aux calculs de développements de ce type pour des équations non linéaires et à leur justification (voir par exemple [18] [13], [9]

et leur bibliographie). Le régime standard de l'optique géométrique dite faiblement non linéaire (cf. [8]) correspond à des solutions de la forme $u(t, x) = hU(t, x, \varphi/h)$ ou à des non linéarités de la forme $V(x)u + hF_1(x, u)$. Cette situation est maintenant bien comprise (cf. [9]). Dans le cas d'amplitudes supérieures ($u = h^\kappa U(t, x, \varphi/h)$, $\kappa < 1$) la situation est plus délicate. L'existence de solutions formelles ne peut avoir lieu que sous certaines hypothèses de structure pour les équations (appelées conditions de transparence dans [10], et qui, pour les équations quasi-linéaires correspondent à des conditions de dégénérescence linéaire de la valeur propre, cf. [3]). Le point important est que les conditions d'existence de solutions BKW ne garantissent pas leur stabilité. En particulier, l'existence de solutions exactes ayant le développement formel trouvé n'est pas garantie (cf. [10], [3]). C'est exactement le phénomène observé par G. Lebeau pour les équations d'ondes (11). D'une part, il précise la construction de solutions formelles esquissée dans [18] et d'autre part il montre leur instabilité.

En reportant (14) dans (11), on obtient la suite d'équations

$$(15) \quad \sigma^2 \partial_\theta^2 U_0 + F'_u(x, U_0) = 0,$$

$$(16) \quad \sigma^2 \partial_\theta^2 U_1 + F''_{u,u}(x, U_0)U_1 + T \partial_\theta U_0 = 0,$$

$$(17) \quad \sigma^2 \partial_\theta^2 U_k + F''_{u,u}(x, U_0)U_k + T \partial_\theta U_{k-1} + \triangleright U_{k-2} + R_k(x, U_0, \dots, U_{k-1}) = 0.$$

avec $\triangleright = \partial_t^2 - \partial_x^2$,

$$(18) \quad \sigma^2 = (\partial_t \varphi)^2 - (\partial_x \varphi)^2, \quad T := 2\varphi'_t \partial_t - 2\varphi'_x \partial_x + \triangleright \varphi,$$

et R_k est une fonction lisse de ses arguments. On obtient de même les conditions initiales

$$(19) \quad U_k(0, x, 0) = a_k(x),$$

$$(20) \quad \varphi'_t(0, x)(\partial_\theta U_k)(0, x, 0) + (\partial_t U_{k-1})(0, x, 0) = b_k(x).$$

(on a utilisé que $\varphi(0, x) = 0$).

THÉORÈME 2.1. — *La suite d'équations (15)-(20) possède une unique solution formelle (U, φ) avec $U = \sum h^k U_k$.*

Pour la preuve, on renvoie à [11]. On esquisse ici le début de la construction, qui est la détermination de φ et U_0 . La grande différence avec l'optique faiblement non linéaire, est que l'équation «eikonale» pour φ est couplée à l'équation de transport pour U_0 (cf. aussi [15] dans le cas quasi-linéaire).

1) Comme σ est indépendant de θ , l'équation (15) est une équation différentielle ordinaire en θ dépendant des paramètres σ et x . Pour σ et x fixés, les solutions dépendent de deux paramètres. Fixer la période élimine un paramètre. D'autre part, l'équation est invariante par translation en θ . Les solutions 2π -périodiques de (15) sont donc

$$(21) \quad U_0(t, x, \theta) = K(\sigma, x, \theta + \Theta(t, x))$$

avec Θ une fonction déphasage arbitraire et $K(\sigma, x, \theta) = x\sigma^{\frac{2}{p-1}}G(\theta)$, où G est l'unique solution 2π -périodique de $G'' + G^p = 0$ vérifiant $G'(0) = 0$, $G(0) > 0$.

Par ailleurs, les données initiales (19) (20) déterminent $\partial_t \varphi|_{t=0} = \sigma|_{t=0}$ et $\Theta|_{t=0}$.

REMARQUE 2.2. — Dans le cas linéaire (ou faiblement non linéaire) (i.e. $F'_u(x, U_0) = V(x)U_0$, avec $V > 0$), toutes les solutions de (15) (i.e. une famille à deux paramètres) ont la même période qui vaut 2π si φ vérifie l'équation eikonale $\sigma^2 = V$. Dans le cas fortement non linéaire, le paramétrage des solutions est radicalement différent : pour tout σ , il y a une famille à un paramètre de solutions de période fixée.

2) L'équation (16) est une équation différentielle linéaire en θ pour U_1 de la forme $\mathcal{L}U_1 = -T\partial_\theta U_0$ où \mathcal{L} est le linéarisé de (15) en U_0 . L'invariance par translation de (15) implique que \mathcal{L} a un noyau. Plus précisément \mathcal{L} est autoadjoint et a un noyau de dimension un engendré par $\partial_\theta K(\sigma, x, \cdot + \Theta)$. L'équation (16) admet donc une solution 2π -périodique si et seulement si le membre de droite

$$-T\partial_\theta U_0 = -(T + Z(\Theta)\partial_\theta)\partial_\theta K(\sigma, x, \cdot + \Theta)$$

est orthogonal au noyau. On a noté $Z = 2\varphi'_t \partial_t - 2\varphi'_x \partial_x$. Cela conduit à l'équation :

$$\int_0^{2\pi} (T\partial_\theta K + Z(\Theta)\partial_\theta^2 K)(\partial_\theta K) d\theta = 0$$

ou encore, en notant

$$J(t, x) := \frac{1}{2} \int_0^{2\pi} (\partial_\theta K)^2 d\theta = \frac{1}{2} x^2 \sigma^{\frac{4}{p-1}} \int_0^{2\pi} (\partial_\theta G(\theta))^2 d\theta,$$

$$(22) \quad \partial_t \varphi \partial_t J - \partial_x \varphi \partial_x J + \triangleright \varphi J = 0.$$

Avec (18), on voit que φ est déterminée par les équations

$$(23) \quad \begin{cases} \partial_t q = \partial_x(\rho J) \\ \partial_t \rho = \partial_x(q/J) \end{cases}, \quad \rho = \partial_x \varphi, \quad q = \partial_t \varphi J.$$

Comme on connaît $\varphi|_{t=0} = 0$ et $\partial_t \varphi|_{t=0}$ par l'étape 1, on voit que φ est maintenant bien déterminée.

Notant \mathcal{L}_0 l'opérateur \mathcal{L} associé à $\Theta = 0$ et \mathcal{L}_0^{-1} son inverse partiel sur $\ker \mathcal{L}_0^\perp$, on a donc $U_1(t, x, \theta) = V_1(t, x, \theta + \Theta(t, x))$ avec

$$(24) \quad V_1 = \lambda_1 \partial_\theta K - \mathcal{L}_0^{-1}(T\partial_\theta K) - Z(\Theta)\mathcal{L}_0^{-1}(\partial_\theta^2 K)$$

et λ_1 une fonction arbitraire de (t, x) .

Les conditions initiales (19) (20) avec $k = 1$ déterminent $\lambda_1|_{t=0}$ et $\partial_t \Theta|_{t=0}$.

3) L'équation (17) pour $k = 2$ est de la forme $\mathcal{L}U_2 = F_2(t, x, \theta + \Theta)$ où F_2 s'exprime à l'aide de K et V_1 et est quadratique en V_1 . La condition de résolubilité

$$\int_0^{2\pi} F_2(t, x, \theta) \partial_\theta K(t, x, \theta) d\theta = 0$$

s'écrit a priori sous la forme

$$A\lambda_1^2 + B\lambda_1 + C = 0.$$

PROPOSITION 2.3. — *On a $A \equiv 0$, $B \equiv 0$ et l'équation $C = 0$ s'explique comme une équation hyperbolique du second ordre pour Θ .*

La première identité n'utilise que l'invariance par translation de (15). La seconde exprime une compatibilité ou condition de transparence très forte des équations. Elle utilise que U_1 vérifie l'équation (16). Il est à peu près évident que C est un opérateur du second ordre en Θ . L'hyperbolicité se vérifie directement.

Connaissant les données initiales de Θ par les étapes 1) et 2), on en déduit Θ , et U_0 est maintenant complètement déterminé.

L'équation (17) pour $k = 3$ détermine λ_1 donc U_1 , puis les équations suivantes permettent de trouver tous les termes U_k par récurrence.

3. LE MÉCANISME DE L'INSTABILITÉ

Par le procédé de sommation de Borel, le Théorème 2.1 permet de construire au voisinage de $(0, x_0)$ des solutions approchées de (11), c'est-à-dire telles que

$$(25) \quad h^2(\partial_t^2 - \partial_x^2)u_{app} + \frac{\partial F}{\partial u}(x, u_{app}) = O(h^\infty), \quad u_{app}|_{t=0} = u_0, \quad \partial_t u_{app}|_{s=0} = u_1$$

dont les données initiales vérifient (12). Le problème est maintenant de construire des solutions exactes de (11), voisines de u_{app} . Les résultats standard de l'optique géométrique, ici les méthodes banales de perturbation, nous disent que cela est possible *pour des temps* $t = O(h)$. Le problème est donc d'étudier la stabilité des solutions approchées u_{app} pour des temps plus longs. L'optimum serait d'atteindre des temps $O(1)$. En fait, on va voir que pour des temps d'ordre $h|\ln h|$ des instabilités apparaissent.

Dans (11), effectuons le changement d'échelle $t = hs$, pour obtenir l'équation

$$(26) \quad \partial_s^2 u + F'_u(x, u) = h^2 \partial_x^2 u.$$

On obtient des solutions asymptotiques

$$(27) \quad u \sim \sum_{k \geq 0} h^{2k} u_{2k}$$

en résolvant la cascade d'équations :

$$(28) \quad \partial_s^2 u_0 + F'_u(x, u_0) = 0,$$

$$(29) \quad \partial_s^2 u_1 + F''_{u,u}(x, u_0)u_1 = \partial_x^2 u_0,$$

etc.

(Le lecteur est invité à réfléchir aux liens existant entre ces solutions et les solutions données par le Théorème 2.1.) Les solutions de (28) sont globales et périodiques (ce qui justifie l'apparition d'oscillations en t/h ou φ/h pour les solutions de (11)). Par contre les solutions de (29) et des équations suivantes ont une croissance polynomiale en s , ce qui indique que le développement (27) n'a aucune chance d'être valable pour des $s = O(h^{-1})$ (d'où la nécessité d'introduire un ansatz sophistiqué comme en (14)). En regardant de près, on peut faire converger (27) pour des temps $s = O(h^{-\alpha})$ pour un $\alpha > 0$ dépendant de p . Mais les premiers phénomènes apparaissent en temps $s = O(|\ln h|)$.

L'idée est la suivante. Les équations ci-dessus sont gouvernées par l'équation différentielle en s (28) et ses linéarisées. La variable x n'intervient que comme paramètre. Le terme $h^2 \partial_x^2$ est considéré comme une perturbation. Tout cela n'a de sens que pour des données très régulières en x et h , i.e. des perturbations uniformément basse fréquence. Au contraire, les données initiales (12) permettent des perturbations d'amplitude $O(h^\infty)$ à haute fréquence $O(h^{-1})$. Si ces fréquences sont amplifiées exponentiellement, alors la perturbation est de l'ordre

$$O(h^\infty)e^{\mu s}$$

et si $\mu > 0$, la perturbation devient $O(1)$ en temps $s \gg |\ln h|$.

G. Lebeau étudie d'abord la stabilité linéaire des solutions approchées u_{app} construites par le Théorème 2.1. Le linéarisé de (11) autour de u_{app} est :

$$(30) \quad h^2(\partial_t^2 - \partial_x^2)\dot{u} - \partial_u^2 F(x, u_{app})\dot{u} = \dot{f}, \quad \dot{u}|_{t=0} = \dot{u}_0, \quad \partial_t \dot{u}|_{t=0} = \dot{u}_1.$$

Compte tenu de (21), le potentiel $\partial_u^2 F(x, u_{app})$ est une perturbation de $p\sigma^2 G^{p-1}(\varphi/h + \Theta)$. Après un changement de variables et de fonctions (prenant notamment φ comme nouvelle variable de temps), on se ramène dans [11] à une perturbation de

$$(31) \quad h^2(\partial_t^2 - \alpha \partial_x^2) + pG^{p-1}(t/h + \tilde{\Theta}).$$

On devrait arriver au même constat plus simplement. Puisqu'on va s'intéresser à des temps très courts, $t \ll h^{1-\varepsilon}$ pour tout $\varepsilon > 0$, on n'a pas besoin de toute la finesse de la description des solutions asymptotiques du paragraphe 2. Les solutions approchées du type (27) existant sur des temps $|t| \leq h^{1-\varepsilon_0}$ pour un certain $\varepsilon_0 > 0$ devraient suffire. On considérerait alors le linéarisé de (26) qui est une perturbation de

$$\partial_s^2 - h^2 \partial_x^2 + \partial_u^2 F(x, u_0).$$

Explicitant la solution u_0 de (28), on obtient à nouveau un opérateur de la forme (31).

Pour donner une idée du mécanisme, considérons dans la variable $s = t/h$ l'opérateur (31) avec α et $\tilde{\Theta}$ constants :

$$(32) \quad \mathcal{M} := \partial_s^2 - h^2 \partial_x^2 + pG^{p-1}(s + s_0).$$

On étudie le comportement en temps s grand des solutions de $\mathcal{M}u = 0$. Après transformation de Fourier, on se ramène aux équations

$$(33) \quad \mathcal{M}_\lambda := \partial_s^2 + pG^{p-1}(s + s_0) + \lambda, \quad \lambda = h^2 \xi^2.$$

On note $E_\lambda(s)$ la matrice 2×2 donnant l'évolution des solutions de $\mathcal{M}_\lambda u = 0$:

$$E_\lambda(s) \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} u(s) \\ u'(s) \end{pmatrix}$$

où u est la solution de $\mathcal{M}_\lambda u = 0$, $u(0) = a_0$, $u'(0) = a_1$. Le potentiel G^{p-1} est 2π -périodique (et même π -périodique puisque $p-1$ est pair et que $G(s+\pi) = -G(s)$). On a donc

$$(34) \quad E_\lambda(s) = E_\lambda(s')(E_\lambda(2\pi))^k, \quad s = s' + 2k\pi.$$

Le comportement en grand temps est donc donné par le comportement des itérés Δ_λ^k , où $\Delta_\lambda := E_\lambda(2\pi)$. La Proposition 4.1 de [11] donne les informations cruciales sur le spectre des Δ_λ . Retenons ici :

PROPOSITION 3.1. — *Il existe $\mu_0 > 0$ et $\lambda_0 > 0$ tels que $e^{2\pi\mu_0}$ est valeur propre de Δ_{λ_0} et pour tout $\lambda \geq 0$ les valeurs propres de Δ_λ sont de partie réelle au plus égale à $e^{2\pi\mu_0}$.*

Notons $E(s)$ l'évolution de $(u, \partial_s u)$ par \mathcal{M} et $\Delta = E(2\pi)$. Par le théorème de Plancherel, on a :

PROPOSITION 3.2. — *Il existe une constante C et une norme sur $\mathcal{H} = H^1(\mathbb{R}) \oplus L^2(\mathbb{R})$ équivalente à la norme usuelle telle que*

$$\|\Delta\|_{\mathcal{L}(\mathcal{H})} = e^{2\pi\mu_0} \quad \text{et pour tout } s \geq 0 : \quad \|E(s)\|_{\mathcal{L}(\mathcal{H})} \leq C e^{s\mu_0}.$$

De plus, ces majorations sont essentiellement optimales pour des fonctions bien polarisées et à spectre très localisé autour de $\sqrt{\lambda_0}/h$. Soit $\underline{V} = {}^t(v_0, v_1)$ un vecteur propre de Δ_{λ_0} associé à la valeur propre (simple) $e^{2\pi\mu_0} > 1$. Soit $V(s) = E_{\lambda_0}(s)\underline{V}$. Alors

$$|V(s)| \geq c e^{s\mu_0}.$$

Soit ξ_0 tel que $\xi_0^2 = \lambda_0$. Considérons des données initiales

$$(35) \quad {}^t(\dot{u}_0(x), \dot{u}_1(x)) = \chi(x) e^{i\xi_0 x/h} \underline{V}$$

où $\chi \in C_0^\infty(\mathbb{R})$ est une fonction de localisation égale à un sur l'intervalle I . Soit $\dot{U}(s) = {}^t(\dot{u}, \partial_s \dot{u}) = E(s) {}^t(\dot{u}_0, \dot{u}_1)$. Par vitesse finie de propagation, on a

$$(36) \quad \dot{U}(s) = e^{ix\xi_0/h} V(s)$$

pour x dans un intervalle $I' \subset\subset I$ et $s \in [0, T/h]$, $T > 0$ ne dépendant que de I et I' . En particulier, il existe $T > 0$ et $c > 0$ tels que, pour $s \leq T/h$, on a :

$$(37) \quad \|\dot{U}(s)\|_{\mathcal{H}} \geq ce^{s\mu_0}$$

On obtient bien sûr des solutions réelles en prenant la partie réelle des solutions que l'on vient de construire.

Une partie technique du travail de G. Lebeau consiste à montrer que l'étude esquissée ci-dessus, qui est assez claire dans le cas où les coefficients α et $\tilde{\Theta}$ de (31) sont constants, s'étend au cas de coefficients variables. L'idée est que le calcul en x est de type semi-classique et donc que le figeage des coefficients donne une bonne approximation.

4. INSTABILITÉS NON LINÉAIRES

On trouve dans la littérature un certain nombre de résultats montrant que l'instabilité linéaire induit une instabilité non linéaire. Voir par exemple [7] pour un résultat assez général abstrait sous des hypothèses spectrales pour le générateur de l'évolution linéaire. On décrit ici le principe de l'analyse de [11] sans rentrer dans les détails des preuves. Des variantes de ce principe ont déjà été utilisées dans [5] et [10].

On part d'une solution approchée u_{app} de (11) et on cherche les solutions exactes sous la forme $u_{ex} = u_{app} + u_{cor}$. On obtient pour $u = (u_{cor}, \partial_s u_{cor})$, dans les variables appropriées, une équation de la forme

$$(38) \quad \partial_s u - A_h u = N_h(u) + f_h$$

où A_h est un opérateur linéaire, N_h est la somme d'un terme linéaire petit en h et de termes au moins quadratiques, et l'erreur f_h est $O(h^\infty)$.

On considère différents jeux de données initiales pour u_{cor} . Typiquement, on choisit

$$(39) \quad u|_{s=0} = a_\kappa := \kappa \alpha(h) \underline{a}, \quad \alpha(h) = O(h^\infty)$$

avec $\kappa = 0$ ou 1 et \underline{a} un terme source de l'instabilité linéaire. Les deux données initiales vérifient donc $a_0 - a_1 = O(h^\infty)$ et on espère construire des solutions u_κ dont la différence est amplifiée exponentiellement :

$$\|u_0(s) - u_1(s)\| \geq c\alpha(h)e^{s\mu_0}, \quad c > 0.$$

Si les solutions vivent assez longtemps pour que $\alpha(h)e^{s\mu_0} \geq c' > 0$, la différence $u_0 - u_1$ est effectivement d'ordre 1 et on a bien mis en évidence l'instabilité non linéaire annoncée.

On a vu au paragraphe 3 que la donnée instable \underline{a} contient des oscillations en $e^{\pm ix\xi_0/h}$. Ces oscillations vont se propager aux solutions, et par non linéarité on doit considérer toutes leurs harmoniques. Cela conduit à introduire la variable rapide $X = x\xi_0/h$ et à chercher les solutions de (38) sous la forme

$$(40) \quad u(s, x) = \mathbf{u}(s, x, x\xi_0/h)$$

avec $\mathbf{u}(s, x, X)$ périodique en X .

REMARQUE 4.1. — La solution approchée est elle-même construite comme une fonction périodique de $s = t/h$ (ou plutôt chez Lebeau $s = \varphi/h$). Si bien que dans les variables originales, on est confronté à *des oscillations à deux phases* φ et $x\xi_0$. Si les développements à une phase φ ont une certaine consistance, il n'y a aucune raison a priori pour que les développements à deux phases soient stables. On retrouve là la problématique de [10].

Considérant f_h comme une fonction indépendante de X , il *suffit* de résoudre l'équation

$$(41) \quad \partial_s \mathbf{u} - \mathbf{A}_h \mathbf{u} = \mathbf{N}_h(\mathbf{u}) + f_h, \quad \mathbf{u}|_{s=0} = \mathbf{a},$$

où \mathbf{A}_h et \mathbf{N}_h sont les extensions évidentes de A_h et N_h . On note $\mathbf{E}_h(s, s')$ le groupe d'évolution de $\partial_s - \mathbf{A}_h$, autrement dit $(\partial_s - \mathbf{A}_h)\mathbf{E}_h(s, s') = 0$ avec $\mathbf{E}_h(s, s) = \text{Id}$. Supposons ici que l'on a (cf. Proposition 5.1 et Lemme 5.4 de [11]).

HYPOTHÈSES. — Il existe un espace \mathbf{H} , des constantes $C_0, h_0 > 0$ et $\mu_0 > 0$ telles que pour $0 < h \leq h_0$ on a :

i) pour $0 \leq s' \leq s$ et $\mathbf{f} \in \mathbf{H}$:

$$(42) \quad \|\mathbf{E}_h(s, s')\mathbf{f}\|_{\mathbf{H}} \leq C_0 e^{(s-s')\mu_0} \|\mathbf{f}\|_{\mathbf{H}},$$

ii) pour \mathbf{f} et \mathbf{g} dans la boule unité de \mathbf{H} on a

$$(43) \quad \|\mathbf{N}_h(\mathbf{f})\|_{\mathbf{H}} \leq C_0 (h\|\mathbf{f}\|_{\mathbf{H}} + \|\mathbf{f}\|_{\mathbf{H}}^2),$$

$$(44) \quad \|\mathbf{N}_h(\mathbf{f}) - \mathbf{N}_h(\mathbf{g})\|_{\mathbf{H}} \leq C_0 \|\mathbf{f} - \mathbf{g}\|_{\mathbf{H}} (h + \|\mathbf{f}\|_{\mathbf{H}} + \|\mathbf{g}\|_{\mathbf{H}}),$$

iii) il existe une donnée initiale \mathbf{a} dans \mathbf{H} telle que

$$(45) \quad \|\mathbf{a}\| = 1 \quad \text{et} \quad \|\mathbf{E}_h(s, 0)\mathbf{a}\|_{\mathbf{H}} \geq \frac{1}{C_0} e^{s\mu_0}.$$

On considère le terme source f_h de (41) et on suppose que

$$(46) \quad \|f_h\|_{\mathbf{H}} = O(h^\infty).$$

On peut alors choisir une fonction $s(h)$ telle que

$$(47) \quad |\ln h| = o(s(h)) \quad \text{et} \quad s(h) = o(h^{-\varepsilon}) \quad \forall \varepsilon > 0,$$

$$(48) \quad e^{s(h)\mu_0} \|f_h\|_{\mathbf{H}} = O(h^\infty).$$

REMARQUE. Les estimations (42) à (45) ne sont utiles que pour $s \leq s(h)$.

On résout (41) en cherchant \mathbf{u} tel que

$$\mathbf{u} = \Phi(\mathbf{u}) := \mathbf{E}_h(s, 0)\mathbf{a} + \int_0^s \mathbf{E}_h(s, s')(f_h(s') + \mathbf{N}_h(\mathbf{u}(s'))) ds'.$$

Les estimations ci-dessus permettent d'appliquer le Théorème du point fixe pour des données initiales \mathbf{a} assez petites.

PROPOSITION 4.2. — Il existe $C_1, C_2, h_1 > 0$ et $\varepsilon_1 > 0$ tels que si $0 < h \leq h_1$, $0 < \varepsilon \leq \varepsilon_1$ et

$$(49) \quad \|\mathbf{a}\|_{\mathbf{H}} \leq \varepsilon e^{-s(h)\mu_0}$$

alors le problème de Cauchy (41) admet une unique solution $\mathbf{u} \in C^0([0, s(h)]; \mathbf{H})$ telle que

$$(50) \quad \|\mathbf{u}(s)\|_{\mathbf{H}} \leq C_1(\varepsilon + h)e^{(s-s(h))\mu_0},$$

$$(51) \quad \|\mathbf{u}(s) - \mathbf{E}_h(s, 0)\mathbf{a}\|_{\mathbf{H}} \leq C_2(h + \sqrt{h}\varepsilon + \varepsilon^2)e^{(s-s(h))\mu_0}.$$

PREUVE. On montre que si \mathbf{u} satisfait (50), alors il en est de même de $\Phi(\mathbf{u})$. La norme du premier terme de $\Phi(\mathbf{u})$ est en

$$C_0 e^{s\mu_0} \|\mathbf{a}\|_{\mathbf{H}} \leq C_0 \varepsilon e^{(s-s(h))\mu_0} \leq \frac{1}{4} C_1 \varepsilon e^{(s-s(h))\mu_0}$$

si $C_1 \geq 4C_0$. Avec (48), le second terme est en

$$\left(\int_0^s C_0 e^{(s-s')\mu_0} ds' \right) C h^2 e^{-s(h)\mu_0} \leq \frac{1}{\mu_0} C_0 C h^2 e^{(s-s(h))\mu_0} \leq \frac{1}{4} C_1 h e^{(s-s(h))\mu_0}$$

si h est assez petit. On note que, pour $s \leq s(h)$, (50) implique que $\|\mathbf{u}(s)\|_{\mathbf{H}} \leq 1$, pourvu que ε et h soient assez petits. Alors la norme du troisième terme de $\Phi(\mathbf{u})$ est en

$$\int_0^s C_0^2 e^{(s-s')\mu_0} h C_1 (\varepsilon + h) e^{(s'-s(h))\mu_0} ds' + \int_0^s C_0^2 e^{(s-s')\mu_0} C_1^2 (\varepsilon + h)^2 e^{2(s'-s(h))\mu_0} ds'.$$

La première intégrale se majore pour h petit par

$$C_0^2 h s(h) C_1 (\varepsilon + h) e^{(s-s(h))\mu_0} \leq \frac{1}{4} C_1 (\varepsilon + h) e^{(s-s(h))\mu_0}$$

puisque $hs(h)$ tend vers zéro par (47). La seconde intégrale est majorée par

$$\frac{1}{\mu_0} C_0^2 C_1^2 (\varepsilon + h)^2 e^{2(s-s(h))\mu_0} \leq \frac{1}{4} C_1 (\varepsilon + h) e^{(s-s(h))\mu_0}$$

pour ε et h assez petits.

On montre de même que l'application Φ est contractante, d'où l'existence de la solution \mathbf{u} . En reprenant les calculs ci-dessus, notamment en utilisant que $hs(h) = O(\sqrt{h})$, on obtient (51). \triangleright

On considère alors deux données de Cauchy,

$$\mathbf{a}_0 = 0 \quad \text{et} \quad \mathbf{a}_1 = \varepsilon e^{-s(h)\mu_0} \underline{\mathbf{a}},$$

qui donnent naissance à deux solutions \mathbf{u}_0 et \mathbf{u}_1 . La Proposition 4.2 et (45) impliquent que

$$\|u_0(s) - u_1(s)\|_{\mathbf{H}} \geq \frac{\varepsilon}{C_0} e^{(s-s(h))\mu_0} - 2C_2(h + \sqrt{h}\varepsilon + \varepsilon^2) e^{(s-s(h))\mu_0} \geq \frac{\varepsilon}{2C_0} e^{(s-s(h))\mu_0}$$

PROPOSITION 4.3. — *Il existe $\varepsilon > 0$ et $c > 0$ tels que, pour h assez petit, les problèmes de Cauchy (41) avec les données initiales \mathbf{a}_0 et \mathbf{a}_1 ont des solutions \mathbf{u}_0 et \mathbf{u}_1 dans $C^0([0, s(h)]; \mathbf{H})$. En outre, on a lorsque h tend vers zéro :*

$$\|\mathbf{a}_0 - \mathbf{a}_1\|_{\mathbf{H}} = O(h^\infty) \quad \text{et} \quad \|(\mathbf{u}_0 - \mathbf{u}_1)|_{s=s(h)}\|_{\mathbf{H}} \geq c.$$

BIBLIOGRAPHIE

- [1] H. BAHOURI, P. GÉRARD – *High frequency approximation of solutions to critical nonlinear wave equations*, Amer. J. Math. 121 (1999), 131-175.
- [2] H. BAHOURI, J. SHATAH – *Decay estimates for the critical semi-linear wave equation*, Ann. I.H.P., Anal. non linéaire 15 (1998), 783-789.
- [3] C. CHEVERRY, O. GUÈS, G. MÉTIVIER – *Oscillations fortes sur un champ linéairement dégénéré*, preprint
- [4] J. GINIBRE, A. SOFFER, G. VELO – *The global Cauchy problem for the critical nonlinear wave equation*, J. Funct. Anal. 110 (1992), 96–130.
- [5] E. GRENIER – *On the nonlinear instability of Euler and Prandtl equations*, Comm. Pure Appl. Math. 53 (2000), no. 9, 1067–1091.
- [6] M. GRILLAKIS – *Regularity and asymptotic behaviour of the wave equation with a critical nonlinearity*, Ann. of Math. 132 (1990), 485–509.
- [7] S. FRIEDLANDLER, W. STRAUSS, M. VISHIK – *Nonlinear instability in an ideal fluid*, Ann. Inst. Henri Poincaré, Vol. 14, no. 2 (1997), 187-209.
- [8] J. HUNTER, A. MAJDA, R. ROSALES – *Resonantly interacting weakly nonlinear hyperbolic waves II: several space variables*, Stud. Appl. Math. 75 (1986), 187-226.
- [9] J.-L. JOLY, G. MÉTIVIER, J. RAUCH – *Recent results in non-linear geometric optics*. Hyperbolic problems : theory, numerics, applications, Vol. II (Zürich, 1998), 723–736, Internat. Ser. Numer. Math. 130, Birkhäuser, Basel, 1999.
- [10] J.-L. JOLY, G. MÉTIVIER, J. RAUCH – *Transparent Nonlinear Geometric Optics and Maxwell-Bloch Equations*, J. Diff. Equ. 166 (2000), 175-250.
- [11] G. LEBEAU – *Nonlinear optic and supercritical wave equation*, preprint et *Optique non linéaire et ondes surcritiques*, Séminaire EDP École Polytechnique, 1999-2000.
- [12] H. LINDBLAD, C. SOGGE – *On existence and scattering with minimal regularity for semilinear wave equations*, J. Funct. Anal. 130 (1995), 357–426.
- [13] V. MASLOV, G. OMEL'YANOV – *Geometric Asymptotics for Nonlinear PDE*, Trans. Mathematical Monographs, vol 202 (2000)
- [14] J. RAUCH – *The u^5 Klein-Gordon equation. II. Anomalous singularities for semi-linear wave equations*, in Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. I (Paris, 1978/1979), 335–364.

- [15] D. SERRE – *Oscillations non linéaires des systèmes hyperboliques : méthodes et résultats qualitatifs*, Ann. Inst. H. Poincaré, Anal. Non Linéaire 8 (1991), 351–417.
- [16] J. SHATAH, M. STRUWE – *Well posedness in the energy estimate for semi-linear wave equations with critical growth*, I.M.R.N. (1994), 303-309.
- [17] M. STRUWE – *Globally regular solutions to the u^5 Klein-Gordon equation*, Ann. Scuola Norm. Sup. Pisa 15 (1988), 495–513
- [18] G.B. WHITHAM – *Nonlinear Waves*, Wiley-Intersciences series of texts, John Wiley and Sons, 1993.

Guy MÉTIVIER

Université de Bordeaux 1

M.A.B.

Cours de la Libération

F-33405 Talence Cedex

E-mail : metivier@univ-bordeaux1.fr

LA CONJECTURE DES SOUFFLETS

[d'après I. Sabitov]

par Jean-Marc SCHLENKER

1. LES POLYÈDRES FLEXIBLES

1.1. Les polyèdres dans \mathbf{R}^3

On considère ici des polyèdres non nécessairement convexes dans l'espace euclidien de dimension 3, qu'on notera simplement \mathbf{R}^3 . Un polyèdre sera donné sous la forme d'un complexe simplicial fini P_0 homéomorphe à une surface compacte orientable, muni d'une application $\phi : P_0 \rightarrow \mathbf{R}^3$ affine sur les triangles.

On dira que P est *plongé* lorsque l'application ϕ est injective, et qu'il est seulement *immergé* lorsque ϕ est injective au voisinage de chaque point.

Comme on a supposé P_0 orientable, un polyèdre délimite un domaine fermé borné de \mathbf{R}^3 , qui n'est pas nécessairement connexe lorsque P_0 n'est pas plongé. On peut considérer son volume ; il sera toujours question ici du volume *algébrique*, c'est-à-dire que chaque composante connexe du complémentaire du polyèdre sera comptée avec un signe et éventuellement une multiplicité. Bien sûr, pour les polyèdres plongés, ces distinctions n'interviennent que pour le signe du volume.

1.2. Polyèdres flexibles et infinitésimalement flexibles

Aspects de la rigidité. Étant donné un polyèdre P dans \mathbf{R}^3 , on peut se demander s'il est rigide, c'est-à-dire si on peut le déformer sans changer la forme de ses faces (aux déplacements de \mathbf{R}^2 près). Cette question a en fait plusieurs versions principales distinctes :

- la **rigidité infinitésimale** : est-il possible de trouver un déplacement infinitésimal (i.e. au premier ordre) des sommets de P dans \mathbf{R}^3 , qui ne change pas au premier ordre la longueur de ses arêtes — et donc, puisque les faces sont des triangles, qui ne change pas la forme de ses faces ? Bien entendu il faut se restreindre aux déformations qui ne sont pas triviales (i.e. qui ne proviennent pas d'un

déplacement de P , ou en termes plus pédants d'un champ de Killing de \mathbf{R}^3 restreint aux sommets de P) ;

- la **flexibilité** : peut-on trouver une famille continue à un paramètre $(P_t)_{t \in [0,1]}$, avec $P_0 = P$, telle que pour tout $t \in [0, 1]$, P_t ait la même combinatoire que P et des arêtes de même longueur ? Là encore on doit considérer des « flexions » non triviales, c'est-à-dire qui ne consistent pas simplement en un déplacement de P ;
- la **rigidité globale** : si P' est un polyèdre de \mathbf{R}^3 qui a la même combinatoire que P et des arêtes de même longueur, P' est-il l'image de P par un déplacement ?

Ces propriétés sont en général distinctes, même si des relations existent entre elles. Elles existent d'ailleurs non seulement pour les polyèdres mais aussi pour les surfaces régulières. Par exemple, on ne sait toujours pas si les surfaces compactes régulières de \mathbf{R}^3 peuvent être flexibles (dans la classe des surfaces régulières).

Réalisabilité. Une question liée à la rigidité des polyèdres est la réalisabilité d'une métrique. Étant donné un polyèdre P_0 , vu comme un objet combinatoire — un complexe simplicial fini homéomorphe à une surface compacte orientable — et étant donné, pour chaque arête de P_0 , un nombre réel positif, on se demande s'il existe un polyèdre dans \mathbf{R}^3 , de même combinatoire que P_0 , et dont les longueurs des arêtes sont les nombres qu'on s'est donné.

Pour les polyèdres convexes on dispose grâce à Aleksandrov [Ale58] d'une réponse partielle à cette question, et plus précisément d'une réponse complète à une question parallèle (voir plus bas). Dans le cas général, on sait très peu de choses, voir quand même [BZ95].

Rigidité d'ordre supérieur. Il existe des notions de rigidité d'ordre supérieur, similaires à la rigidité infinitésimale. La définition même de ces notions est intéressante ; on renvoie le lecteur à [Con93, CS94], et à [Sab92] pour les questions analogues concernant les surfaces régulières.

Rigidité d'autres structures. Les questions de rigidité ne se posent pas seulement pour les polyèdres, mais aussi pour d'autres structures qui modélisent plus précisément les problèmes appliqués où la rigidité intervient (mécanique, architecture, etc). Voir par exemple [Con93, TW00, CW94].

1.3. Polyèdres convexes

Quand on se restreint aux polyèdres convexes, la situation se simplifie considérablement. Le premier résultat dans ce domaine résout la question de la rigidité globale des polyèdres convexes dans \mathbf{R}^3 . Les polyèdres convexes qu'on considère ici sont des bords de corps convexes polyédraux dans \mathbf{R}^3 .

THÉORÈME 1.1 (Legendre, Cauchy [Cau13]). — *Les polyèdres euclidiens convexes sont uniquement déterminés par leur combinatoire et la forme de leurs faces.*

Ce théorème est en général attribué à Cauchy. Pourtant, l'une au moins des idées principales de la preuve est due à Legendre [LegII], qui a démontré le théorème pour certaines classes de polyèdres. La contribution de Legendre est mal connue car elle se trouve seulement dans la première édition de ses *Éléments de géométrie*, mais n'a pas été reprise dans les éditions ultérieures. Elle a été «redécouverte» par Lebesgue, puis à nouveau récemment par Sabitov. Par ailleurs la preuve de Cauchy contenait deux petites erreurs corrigées respectivement par Steinitz [Ste16] et Lebesgue [Leb09].

Legendre ne prétend d'ailleurs pas à la paternité de l'énoncé ; il remarque en effet, suivant Robert Simson, que l'énoncé du théorème 1.1 se trouve de manière au moins implicite dans les *Éléments* d'Euclide (livre XI, définitions 9 et 10), mais sans démonstration.

Ce théorème a aussi une version infinitésimale, obtenue par M. Dehn. Elle peut être obtenue en utilisant les arguments de la preuve de Cauchy (mais la preuve de Dehn était différente).

THÉORÈME 1.2 (Dehn [Deh16]). — *Les polyèdres euclidiens convexes sont infinitésimalement rigides.*

La preuve peut dans une certaine mesure être étendue à certains polyèdres non convexes, mais qui possèdent certaines propriétés des polyèdres convexes (voir [Sto68, RR00]).

De la rigidité à la réalisation des métriques. Le résultat de rigidité de Cauchy est un élément fondamental dans un beau résultat d'Aleksandrov [Ale58], qui peut être considéré comme son extension naturelle. Pour l'énoncer, il faut d'abord parler de métriques à singularités coniques sur les surfaces.

Considérons un cône dans l'espace euclidien de dimension 3. Sa métrique induite est plate sauf en son sommet, où elle est singulière. En ce point, elle peut être obtenue en quotientant le revêtement universel de $\mathbf{R}^2 \setminus \{0\}$ par une rotation d'angle θ autour de 0. L'angle θ est alors appelé angle total autour du point singulier, et $2\pi - \theta$ est la courbure singulière en ce point.

Une métrique plate à singularités coniques sur une surface est une métrique qui est plate, c'est-à-dire localement isométrique au plan euclidien, sauf en un nombre fini de points qui ont un voisinage isométrique à un voisinage d'un sommet d'un cône, comme on vient de les décrire.

THÉORÈME 1.3 (Aleksandrov [Ale58]). — *Soit P un polyèdre euclidien convexe. La métrique induite sur son bord est une métrique plate sur S^2 à singularités coniques, et toutes les singularités sont à courbure singulière positive. Réciproquement, toute métrique de ce type est induite sur le bord d'un unique polyèdre euclidien convexe.*

Dans cet énoncé, l'unicité doit s'entendre aux isométries globales de \mathbf{R}^3 près. On est donc passé d'un résultat de *caractérisation* des polyèdres par leur combinatoire et la forme de leurs faces, à une *relation bijective* entre les polyèdres et certaines métriques sur la sphère.

Le résultat d'Aleksandrov est démontré d'une manière implicite. Pour chaque métrique à singularité conique à courbure positive sur la sphère, il énonce l'existence d'un polyèdre dont c'est la métrique induite — mais sans donner aucune indication sur la combinatoire de ce polyèdre.

Une question ouverte. Il existe quand même quelques questions ouvertes concernant la rigidité infinitésimale des polyèdres euclidiens convexes. En particulier :

CONJECTURE 1.4 (Stoker [Sto68]). — *Soit P un polyèdre convexe. Si une déformation infinitésimale de P ne change pas, au premier ordre, ses angles dièdres, alors elle ne change pas non plus les angles intérieurs de ses faces.*

On peut aussi énoncer une version non infinitésimale de cette question : si deux polyèdres convexes ont la même combinatoire, et si leurs angles dièdres sont les mêmes, les angles intérieurs de leurs faces sont-ils identiques ? On trouvera aussi ci-dessous la version hyperbolique de cette conjecture, dont l'énoncé est plus simple encore.

1.4. Polyèdres flexibles

Les octaèdres flexibles. Bricard [Bri97] a caractérisé tous les octaèdres euclidiens flexibles. Il a montré qu'il n'existe pas de tel octaèdre plongé ou même immergé, mais qu'il existe trois familles d'octaèdres — ayant des intersections entre leurs faces au voisinage de certains de leurs sommets — qui sont flexibles.

Les polyèdres « plongés » flexibles. Au cours des années 1970, la question de la rigidité des polyèdres est revenue sur le devant de la scène. Le principal résultat a été la construction par R. Connelly [Con77, Kui79] d'un exemple de polyèdre flexible plongé dans \mathbf{R}^3 .

Peu de temps après la découverte de Connelly, d'autres exemples plus simples ont été mis à jour ; le plus simple exemple connu à ce jour a été construit par K. Steffen, et il n'a que neuf sommets (il est décrit dans [Con79]). Le lecteur intéressé pourra trouver sur le web une multitude de pages décrivant ces exemples, parfois avec de belles images.

Ces exemples restent tout de même assez exceptionnels ; les polyèdres sont « génériquement » rigides, ça a été montré par H. Gluck [Glu75] quand le genre est 0, et dans [Sab02] pour un genre quelconque.

2. LE THÉORÈME DES SOUFFLETS

C'est le résultat suivant.

THÉORÈME 2.1 (Sabitov [Sab96]). — *Soit P un polyèdre flexible dans \mathbf{R}^3 . Lors d'une flexion de P , son volume ne change pas.*

Notons que cet énoncé ne s'applique pas aux déformations infinitésimales d'un polyèdre. Par exemple, si un polyèdre a un « faux » sommet s , dont un voisinage est contenu dans

un plan p , il est facile de voir que les déformations infinitésimales obtenues en déplaçant s orthogonalement à p ne changent pas les longueurs des arêtes, tout en changeant le volume.

2.1. Historique ?

L'origine de cette conjecture n'est pas parfaitement établie. D'après une rumeur insistante, sa première vérification expérimentale aurait eu lieu, dans les années 1970, à l'I.H.É.S. grâce à un modèle de polyèdre flexible qui s'y trouve toujours, et qui est pourvu d'un trou dans l'une de ses faces. D. Sullivan y aurait soufflé la fumée de sa pipe avant d'actionner le polyèdre ; ne voyant pas de fumée sortir, il aurait conclu que le volume reste constant au cours de la déformation. La conjecture est en tous cas énoncée en détail dans l'exposé [Con80] de R. Connelly au Congrès International des Mathématiciens d'Helsinki en 1978, avec une référence à D. Sullivan.

2.2. Propriétés algébriques du volume

La formule de Héron. Soit T un triangle euclidien, et soient a, b et c les longueurs de ses côtés. Une formule qui porte parfois le nom du mathématicien grec Héron exprime l'aire de T en fonction des longueurs des côtés : si $p = (a + b + c)/2$ est le demi-périmètre de T , alors son aire vérifie l'équation :

$$A^2 - p(p - a)(p - b)(p - c) = 0 .$$

Le lecteur intéressé devrait en trouver la preuve sans trop d'efforts.

Il existe une formule analogue qui donne le volume d'un simplexe de dimension trois en fonction de la longueur de ses arêtes. Elle était apparemment connue de Tartaglia, et probablement avant lui, et a été redécouverte en particulier par L. Euler. On ne l'énoncera pas ici car la formule de Cayley-Menger (voir plus bas) fournit une autre formule de ce type, conceptuellement plus simple.

Le théorème 2.2 ci-dessous peut être vu comme une extension de cette formule. Il assure que le volume de chaque polyèdre de \mathbf{R}^3 est racine d'une équation polynomiale dont les coefficients dépendent seulement (de manière polynomiale) des longueurs de ses arêtes.

Le volume des polyèdres comme racine. Le théorème des Soufflets s'explique par le résultat plus fondamental suivant.

THÉORÈME 2.2 (Sabitov). — *Soit P_0 un polyèdre (combinatoire) ayant e arêtes. Il existe des polynômes c_0, c_1, \dots, c_n en e variables à coefficients entiers tels que, si P est un polyèdre de \mathbf{R}^3 combinatoirement équivalent à P_0 et dont les longueurs des arêtes sont l_1, \dots, l_e , alors $12\text{vol}(P)$ est racine de :*

$$X^{2n} + c_1(l_1^2, \dots, l_e^2)X^{2n-2} + c_2(l_1^2, \dots, l_e^2)X^{2n-4} + \dots + c_n(l_1^2, \dots, l_e^2) = 0 .$$

Ce théorème implique directement le théorème 2.1. Sa preuve, donnée dans [Sab96], est constructive. Il admet un autre énoncé, légèrement plus faible mais dont la preuve est plus agréable, bien qu'elle ne soit pas constructive.

Rappelons que, si L est un corps et $R \subset L$ est un sous-anneau, un élément $x \in L$ est entier sur R s'il existe un polynôme unitaire à coefficients dans R dont x est racine.

THÉORÈME 2.3 (Connelly, Sabitov, Walz [CSW97]). — *Soit P un polyèdre dans \mathbf{R}^3 . Alors $12\text{vol}(P)$ est entier sur l'anneau engendré par les carrés des longueurs des arêtes de P .*

Nous nous contenterons ici d'indiquer la preuve du théorème 2.3 ; le lecteur intéressé par la preuve du théorème 2.2 est invité à consulter [Sab98c] ou [Sab98b] pour une preuve légèrement différente.

Notons qu'il n'est pas nécessaire pour le théorème 2.2 de supposer que le polyèdre considéré est plongé ; en fait la preuve est de nature algébrique, si bien que les dégénérescences sont autorisées.

3. LA PREUVE

Il existe maintenant plusieurs preuves de la conjecture des Soufflets, à travers les théorèmes 2.2 ou 2.3. Nous présentons ici une preuve du théorème 2.3 due à Connelly, Sabitov et Walz [CSW97], qui a l'avantage d'être conceptuellement simple et de nécessiter peu de calculs. Par contre elle n'est pas constructive concernant le polyèdre unitaire dont le volume est racine.

Le principe de la preuve est de montrer, par un argument récursif sur la « complexité » des polyèdres, que, pour chaque polyèdre (combinatoire) P , chaque place sur \mathbf{R} qui est finie sur le corps engendré par les carrés des longueurs des arêtes de P est aussi finie sur $12\text{vol}(P)$.

3.1. Préliminaires algébriques

On utilise la notion usuelle suivante de place sur un corps ; voir par exemple [Lan72].

DÉFINITION 3.1. — *Soient L et F des corps. Une **place** sur L à valeurs dans F est une application $\phi : L \rightarrow F \cup \{\infty\}$ telle que, pour tous $x, y \in L$:*

- $\phi(x + y) = \phi(x) + \phi(y)$.
- $\phi(xy) = \phi(x)\phi(y)$.
- $\phi(1) = 1$.

Cette définition sous-entend que ∞ se comporte par rapport à l'addition et à la multiplication comme on peut s'y attendre : si $a \in F$, alors $a \pm \infty = \infty$, $a/\infty = 0$, $a\infty = \infty$ si

$a \neq 0$, et $1/0 = \infty$. Les expressions ∞/∞ , 0∞ et $\infty \pm \infty$ ne sont pas définies, et les conditions de la définition précédente ne s'appliquent que lorsque les termes correspondants sont définis.

Le lemme suivant est classique, voir [Lan72].

LEMME 3.2. — *Soit L un corps contenant un anneau R . Un élément $x \in L$ est entier sur R si et seulement si toute place sur L qui est finie sur R est finie en x .*

On en déduit par exemple immédiatement le résultat important suivant : si $x, y \in L$ sont des éléments entiers sur R , alors $x + y$ et $x - y$ sont entiers sur L .

3.2. Le déterminant de Cayley-Menger

Soient p_0, p_1, \dots, p_n des points de \mathbf{R}^N , et, pour $1 \leq i, j \leq n$, soit $d_{ij} := \|p_i - p_j\|$. Le déterminant de Cayley-Menger des p_i est défini par :

$$CM(p_0, p_1, \dots, p_n) := \begin{vmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{01}^2 & d_{02}^2 & \dots & d_{0n}^2 \\ 1 & d_{01}^2 & 0 & d_{12}^2 & \dots & d_{1n}^2 \\ 1 & d_{02}^2 & d_{12}^2 & 0 & \dots & d_{2n}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & d_{0n}^2 & d_{1n}^2 & d_{2n}^2 & \dots & 0 \end{vmatrix}$$

À un coefficient près, ce déterminant donne le carré du volume du simplexe de dimension n dont les sommets sont les p_i . La relation précise est indiquée par le lemme élémentaire mais fondamental suivant. On pourra consulter [Ber77] pour la preuve.

LEMME 3.3. — *Soit $N \in \mathbf{N} \setminus \{0\}$, et soient $p_0, p_1, \dots, p_n \in (\mathbf{R}^N)^n$. Alors :*

$$CM(p_0, p_1, \dots, p_n) = (-1)^{n+1} 2^n (n!)^2 \text{vol}_n^2(\Sigma(p_0, p_1, \dots, p_n)) ,$$

où $\Sigma(p_0, p_1, \dots, p_n)$ est le simplexe dont les sommets sont les p_i , et vol_n est le volume orienté n -dimensionnel.

Conséquences. On utilisera deux conséquences distinctes de ce lemme. La première, obtenue en prenant $n = 3$, indique que le volume d'un simplexe de \mathbf{R}^3 est une fonction polynomiale des carrés des longueurs de ses arêtes.

COROLLAIRE 3.4. — *Soient $p_0, p_1, p_2, p_3 \in \mathbf{R}^3$. Alors :*

$$CM(p_0, p_1, p_2, p_3) = 2(12\text{vol}(\Sigma(p_0, p_1, p_2, p_3)))^2 .$$

COROLLAIRE 3.5. — *Soient $p_0, p_1, p_2, p_3 \in \mathbf{R}^3$. Alors $12\text{vol}(\Sigma(p_0, p_1, p_2, p_3))$ est entier sur l'anneau engendré par les carrés des longueurs des distances entre les p_i .*

Preuve. — La symétrie du déterminant de Cayley-Menger permet de montrer sans grande difficulté que, lorsqu'on le développe, chacun des termes apparaît deux fois. Ainsi, $(12\text{vol}(\Sigma(p_0, p_1, p_2, p_3)))^2$ est dans l'anneau engendré par les carrés des longueurs entre les p_i , et le résultat suit. \triangleright

Ceci implique en particulier directement que le théorème 2.2 est vrai pour les simplexes, et plus généralement pour tous les polyèdres qu'on peut décomposer en simplexes sans ajouter d'arête ; par exemple pour les polyèdres ci-dessous.

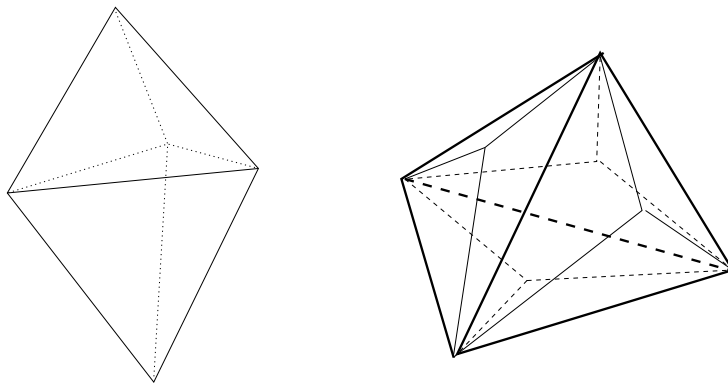


FIGURE 1. Polyèdres décomposables en simplexes

La seconde conséquence est que, lorsqu'on considère cinq points dans \mathbf{R}^3 , les carrés des longueurs des segments qui les joignent satisfont à une équation polynomiale.

COROLLAIRE 3.6. — Soient $p_0, p_1, p_2, p_3, p_4 \in \mathbf{R}^3$. Alors :

$$CM(p_0, p_1, p_2, p_3, p_4) = 0 .$$

3.3. Un lemme sur les pyramides

Le cœur de la preuve est constitué du lemme suivant, qui permet de montrer que si une place ϕ est finie en tous les carrés des longueurs des arêtes d'un polyèdre, on peut «souvent» extraire de ce polyèdre un simplexe Σ tel que ϕ est finie sur le volume au carré de Σ . Dans l'énoncé on identifie les éléments de $\{1, 2, \dots, n\}$ avec $\mathbf{Z}/n\mathbf{Z}$, ce qui revient à identifier de manière cyclique $n + 1$ à 1.

LEMME 3.7. — Soit $n \geq 4$ et soient q, p_1, p_2, \dots, p_n des points de \mathbf{R}^3 . Pour tous $i, j \in \mathbf{Z}/n\mathbf{Z}$, notons $d_i := d(q, p_i)$ et $d_{i,j} := d(p_i, p_j)$. Soit ϕ une place sur \mathbf{R} . Supposons que, pour tout $i \in \mathbf{Z}/n\mathbf{Z}$, $\phi(d_i^2)$ et $\phi(d_{i,i+1}^2)$ sont finis. Alors il existe $i \in \mathbf{Z}/n\mathbf{Z}$ tel que $\phi(d_{i,i+2}^2)$ est fini.

Exemples assez simples. Avant de continuer la preuve du théorème 2.3, on peut remarquer que le lemme précédent prouve déjà le résultat pour les pyramides.

Commençons par une pyramide Π dont la base est un quadrilatère $(abcd)$. Soit ϕ une place sur \mathbf{R} , qui est finie sur l'anneau engendré par les carrés des longueurs des côtés.

D'après le lemme 3.7, ϕ est finie aussi sur l'une des diagonales, c'est-à-dire soit $[ac]$, soit $[bd]$. Décomposons la pyramide en ajoutant cette arête, son volume apparaît alors comme la somme des volumes de deux simplexes S_1 et S_2 . D'après le corollaire 3.5, ϕ est finie sur $12\text{vol}(S_1)$ et sur $12\text{vol}(S_2)$, et donc aussi sur $\text{vol}(\Pi)$. Comme c'est vrai pour tout ϕ , le lemme 3.2 montre que le théorème 2.3 est vrai pour Π .

En fait cet argument s'applique essentiellement à n'importe quelle pyramide Π . Quand la base a plus de quatre côtés, il faut appliquer le lemme 3.7 pour montrer que, étant donné une place ϕ , on peut retirer un simplexe à Π en faisant apparaître une nouvelle arête a telle que ϕ est finie sur le carré de la longueur de a . Un argument de récurrence sur le nombre de côtés de la base montre ensuite que ϕ est finie sur $12\text{vol}(\Pi)$, et le lemme 3.2 permet à nouveau de conclure.

Preuve du lemme 3.7. On va procéder par l'absurde et faire l'hypothèse suivante :

$$\forall i \in \mathbf{Z}/n\mathbf{Z}, \phi(d_{i,i+2}^2) \text{ est infini} \quad (H)$$

On va utiliser (H) pour montrer, par récurrence sur i , que, pour tout $3 \leq i \leq n$, on a :

$$\phi(d_{1,i}^2) \text{ et } \phi(d_{1,i}^2/d_{1,i-1}^2) \text{ sont infinis} \quad (H_i)$$

(H_n) fournira une contradiction puisqu'on a supposé que $\phi(d_{1,n}^2)$ est fini.

Il est clair que (H) implique (H_3) ; on suppose donc (H_i) vrai, pour montrer (H_{i+1}). Il suffit, pour cela, de montrer que $\phi(d_{1,i+1}^2/(d_{1,i}d_{1-i,1+i})^2)$ est non nul ; d'après (H), il suit que $\phi(d_{1,i+1}^2/d_{1,i}^2)$ est infini, puis, avec (H_i), que $\phi(d_{1,i+1}^2)$ est infini. Pour obtenir ce résultat, on raisonne à nouveau par l'absurde et on suppose que $\phi(d_{1,i+1}^2/(d_{1,i}d_{1-i,1+i})^2) = 0$.

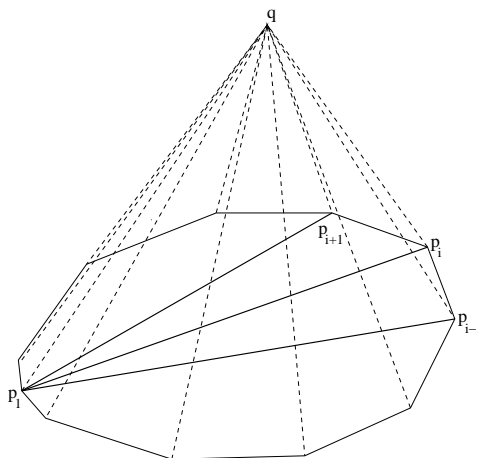


FIGURE 2. La récurrence

On applique d'abord le corollaire 3.6 aux points $p_1, p_{i-1}, p_i, p_{i+1}$ et q , et on obtient que :

$$\begin{vmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{1,i-1}^2 & d_{1,i}^2 & d_{1,i+1}^2 & d_1^2 \\ 1 & d_{1,i-1}^2 & 0 & d_{i-1,i}^2 & d_{i-1,i+1}^2 & d_{i-1}^2 \\ 1 & d_{1,i}^2 & d_{i-1,i}^2 & 0 & d_{i,i+1}^2 & d_i^2 \\ 1 & d_{1,i+1}^2 & d_{i-1,i+1}^2 & d_{i,i+1}^2 & 0 & d_{i+1}^2 \\ 1 & d_1^2 & d_{i-1}^2 & d_i^2 & d_{i+1}^2 & 0 \end{vmatrix} = 0 .$$

On divise la deuxième ligne et la deuxième colonne par $d_{1,i}^2$, et la cinquième ligne et la cinquième colonne par $d_{i-1,i+1}^2$. On obtient ainsi que :

$$\begin{vmatrix} 0 & \frac{1}{d_{1,i}^2} & 1 & 1 & \frac{1}{d_{i-1,i+1}^2} & 1 \\ \frac{1}{d_{1,i}^2} & 0 & \frac{d_{1,i-1}^2}{d_{1,i}^2} & 1 & \frac{d_{1,i+1}^2}{(d_{1,i}d_{i-1,i+1})^2} & \frac{d_1^2}{d_{1,i}^2} \\ 1 & \frac{d_{1,i-1}^2}{d_{1,i}^2} & 0 & d_{i-1,i}^2 & 1 & d_{i-1}^2 \\ 1 & 1 & d_{i-1,i}^2 & 0 & \frac{d_{i,i+1}^2}{d_{i-1,i+1}^2} & d_i^2 \\ \frac{1}{d_{i-1,i+1}^2} & \frac{d_{1,i+1}^2}{(d_{1,i}d_{i-1,i+1})^2} & 1 & \frac{d_{i,i+1}^2}{d_{i-1,i+1}^2} & 0 & \frac{d_{i+1}^2}{d_{i-1,i+1}^2} \\ 1 & \frac{d_1^2}{d_{1,i}^2} & d_{i-1}^2 & d_i^2 & \frac{d_{i+1}^2}{d_{i-1,i+1}^2} & 0 \end{vmatrix} = 0 .$$

On compose cette égalité par ϕ . D'après (H_i) , $\phi(d_{1,i-1}^2/d_{1,i}^2) = \phi(1/d_{1,i}^2) = 0$, et (H) indique que $\phi(1/d_{i-1,i+1}^2) = 0$. Ainsi :

$$\begin{vmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 1 & 0 & 0 & \phi(d_{i-1,i}^2) & \mathbf{1} & \phi(d_{i-1}^2) \\ 1 & \mathbf{1} & \phi(d_{i-1,i}^2) & 0 & 0 & \phi(d_i^2) \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 1 & 0 & \phi(d_{i-1}^2) & \phi(d_i^2) & 0 & 0 \end{vmatrix} = \phi(0) = 0 .$$

Tous les éléments de ce déterminant sont finis. Pour le développer, on voit que les deuxième ligne et deuxième colonne, ainsi que les cinquièmes ligne et colonne, ne contiennent qu'un seul élément non nul (en gras). Ainsi le déterminant est égal à 1, d'où la contradiction recherchée.

3.4. Découpages d'un polyèdre

Soit P un polyèdre dans \mathbf{R}^3 , et soit ϕ une place sur \mathbf{R} , qui prend des valeurs finies aux carrés des longueurs de P . Soit q un sommet de P par lequel ne passe aucun chemin fermé composé de trois arêtes de P qui ne bordent pas de face. Soient p_1, p_2, \dots, p_n les sommets adjacents à q . D'après le lemme 3.7, il existe alors $i_0 \in \mathbf{Z}/n\mathbf{Z}$ tel que $\phi(d(p_{i_0-1}, p_{i_0+1})^2)$ est fini. On appelle P' le polyèdre obtenu en «enlevant» à P le simplexe de sommets $q, p_{i_0-1}, p_{i_0}, p_{i_0+1}$: on supprime le sommet q et les arêtes et faces qui lui sont adjacentes, et on ajoute un triangle dont les sommets sont $p_{i_0-1}, p_{i_0}, p_{i_0+1}$.

On obtient ainsi en utilisant le corollaire 3.5 l'énoncé suivant.

COROLLAIRE 3.8. — *Si $12\phi(\text{vol}(P'))$ est fini, alors $12\phi(\text{vol}(P))$ est fini.*

Il faut noter pour la suite que, quand on passe de P à P' , la valence du sommet q diminue d'une unité.

Preuve du corollaire. — Il est clair que ϕ prend aussi des valeurs finies aux carrés des autres longueurs du simplexe Σ de sommets $q, p_{i_0-1}, p_{i_0}, p_{i_0+1}$. Donc, d'après le corollaire 3.5, ϕ prend une valeur finie en $12\text{vol}(\Sigma)$. Comme par hypothèse $\phi(12\text{vol}(P_i))$ est fini aussi, on obtient que $\phi(12\text{vol}(P))$ est fini. \triangleright

3.5. Un argument de récurrence

On associe à chaque polyèdre P , non nécessairement connexe, le quadruplet (g, n, c, v) , où g est la somme des genres des composantes connexes de P , c le nombre de ses composantes connexes, n le nombre de ses sommets, et v la valence minimale de ses sommets. Étant donné deux polyèdres P et P' , de quadruplets associés (g, c, n, v) et (g', c', n', v') respectivement, on dira que P est moins complexe que P' si l'une des conditions suivantes est satisfaite :

- $g < g'$;
- $g = g'$, et $c > c'$.
- $g = g'$, $c = c'$, et $n < n'$.
- $g = g'$, $n = n'$, $c = c'$, et $v < v'$.

Étant donné un polyèdre P , on lui associe une suite de polyèdres $P_0 = P, P_1, \dots, P_N$, avec $N \in \mathbf{Z}$, non nécessairement connexes, définis récursivement à partir de $P_0 = P$ en appliquant tant que c'est possible l'une des opérations suivantes.

- (1) S'il existe sur P_i un chemin fermé composé de trois arêtes, et qui ne borde pas de face, on coupe P_i suivant le triangle dont le bord est ce chemin. On obtient ainsi un polyèdre P_{i+1} , qui peut avoir soit un genre total inférieur à celui de P_i , soit une composante connexe de plus que P_i .
- (2) Sinon on choisit un sommet q de P_i de valence minimale, et on lui applique le corollaire 3.8 ; on appelle P_{i+1} le polyèdre P' obtenu.

Il est facile de vérifier que, dans tous les cas, P_{i+1} est moins complexe que P_i :

- dans le cas (1), parce que P_{i+1} a soit un genre total inférieur, soit un nombre de composantes connexes supérieur à ceux de P_i .
- dans le cas (2), parce que la valence minimale des sommets de P_{i+1} est inférieure d'une unité à celle des sommets de P_i .

Finalement, on vérifie que P_N est nécessairement une réunion disjointe de simplexes, sans quoi on pourrait trouver un polyèdre P_{N+1} moins complexe que P_N par l'une des opérations décrites ci-dessus.

Preuve du théorème 2.3. Soit P un polyèdre dans \mathbf{R}^3 , et soit ϕ une place qui est finie sur l'ensemble des carrés des longueurs des arêtes de P . Soit $P_0 = P, P_1, \dots, P_N$ la suite de polyèdres définie plus haut (qui dépend de ϕ). P_N est une réunion disjointe de simplexes et, par construction, ϕ est finie sur les carrés des longueurs de ses arêtes. Ainsi, $\phi(12\text{vol}(P_N))$ est fini d'après le corollaire 3.5.

La construction de $(P_i)_{0 \leq i \leq N}$ permet ensuite de montrer, par récurrence inverse sur i , que $\phi(12\text{vol}(P_i))$ est fini pour tout $i \in \{0, 1, \dots, N\}$, et ceci est en particulier vrai pour P . Le théorème suit donc par application du lemme 3.2.

4. PRÉCISIONS, VARIANTES

4.1. Polynômes minimaux

Dans [AS99], Astrelin et Sabitov montrent qu'il existe, pour chaque polyèdre (combinatoire) de genre $g = 0$, un unique polynôme minimal qui annule le volume.

THÉORÈME 4.1 (Astrelin, Sabitov). — *Soit P_0 un polyèdre (vu comme objet combinatoire) et soient a_1, \dots, a_N ses arêtes. Il existe un polynôme unitaire $Q \in \mathbf{Q}[l][V]$, à coefficients dans les polynômes à N indéterminées à coefficients rationnels, avec la propriété suivante. Soit P un polyèdre dans \mathbf{R}^3 ayant la combinatoire de P_0 , soient $(l_i)_{1 \leq i \leq N}$ les longueurs de ses arêtes et soit V son volume. Alors $Q(l_i^2)(12V) = 0$.*

De plus, il existe un unique polynôme ayant cette propriété et de degré minimal, et il divise tous les autres.

4.2. Autres applications

Les méthodes développées pour montrer le théorème 2.2 ont d'autres applications à la théorie métrique des polyèdres dans \mathbf{R}^3 . Par exemple, dans [Sab02], Sabitov montre que les longueurs de certaines « diagonales » d'un polyèdre — des segments joignant deux sommets, mais qui ne sont pas des arêtes — ont pour longueur des racines d'équations polynomiales dont les coefficients dépendent des longueurs des arêtes. Il en déduit une preuve du fait (connu depuis [Glu75] en genre 0) que les polyèdres dans \mathbf{R}^3 sont génériquement rigides. Le lecteur pourra trouver d'autres applications intéressantes dans [Sab98a, Sab01].

4.3. Généralisations directes

La nature essentiellement algébrique de la preuve du théorème 2.3, ainsi d'ailleurs que celle du théorème 2.2 qu'on pourra trouver dans [Sab98c], indiquent que le résultat n'est pas limité au cas où le corps de base est \mathbf{R} .

5. AU-DELÀ DE \mathbf{R}^3

On va considérer maintenant les généralisations possibles des résultats énoncés plus haut à d'autres contextes géométriquement intéressants.

5.1. L'espace de Minkowski

L'espace de Minkowski de dimension trois est simplement \mathbf{R}^3 , muni de la forme quadratique $dx^2 + dy^2 - dz^2$. C'est une variété lorentzienne plate complète, qu'on notera ici \mathbf{R}_1^3 . La structure affine provenant de sa connexion de Levi-Civita est identique à celle de \mathbf{R}^3 .

La rigidité infinitésimale des polyèdres convexes de \mathbf{R}_1^3 est une conséquence directe du résultat de Cauchy dans \mathbf{R}^3 ; en effet, il est immédiat que, si V est une déformation infinitésimale isométrique d'un polyèdre convexe $P \subset \mathbf{R}_1^3$, et si V a pour coordonnées (X, Y, Z) , alors le vecteur V' de coordonnées $(X, Y, -Z)$ est une déformation isométrique infinitésimale de P vu comme un polyèdre de \mathbf{R}^3 ; de plus, V' est trivial si et seulement si V l'est.

V. Alexandrov [Ale01] a montré que la preuve du théorème 2.2 peut être étendue à l'espace de Minkowski. La conjecture des Soufflets y est donc encore valable.

5.2. Polyèdres hyperboliques et sphériques

L'espace hyperbolique et la sphère. On peut aussi remplacer l'espace euclidien par un autre espace riemannien de dimension trois à courbure constante. Les deux candidats naturels sont la sphère de dimension trois munie de sa métrique habituelle, qu'il n'est pas nécessaire de présenter au lecteur, et l'espace hyperbolique de dimension trois, qu'on peut voir comme une sous-variété de l'espace de Minkowski de dimension quatre, muni de la métrique induite :

$$H^3 := \{x \in \mathbf{R}_1^4 \mid \langle x, x \rangle = 1 \wedge x_0 > 0\} .$$

Modèles projectifs. Il existe une application $\rho_H : H^3 \rightarrow \mathbf{R}^3$, dont l'image est la boule de rayon 1 dans \mathbf{R}^3 , et qui a la propriété remarquable d'envoyer les géodésiques de H^3 sur des segments géodésiques dans \mathbf{R}^3 . On en déduit le «modèle projectif», aussi appelé «modèle de Klein», de H^3 .

En conséquence, on peut associer à chaque polyèdre de H^3 un polyèdre de \mathbf{R}^3 , qui est simplement son image par ρ_H .

De même, il existe une application ρ_S entre un hémisphère (ouvert) de S^3 et \mathbf{R}^3 , qui envoie les demi-grands cercles de la source sur les géodésiques de \mathbf{R}^3 . On peut donc aussi associer à chaque polyèdre de S^3 contenu dans un hémisphère (ouvert) un polyèdre euclidien.

La rigidité infinitésimale est une propriété projective. Un résultat ancien mais frappant affirme que la rigidité infinitésimale est une propriété purement projective.

THÉORÈME 5.1 (Darboux [Dar93], Sauer [Sau35]). — *Soit P un polyèdre euclidien, et soit u une transformation projective de \mathbf{R}^3 . Alors P est infinitésimalement rigide si et seulement si $u(P)$ l'est.*

Cet énoncé est lié à un autre fait remarquable : un polyèdre hyperbolique (ou un polyèdre sphérique contenu dans un hémisphère) est infinitésimalement rigide si et seulement si le polyèdre euclidien qui lui est associé par le modèle projectif l'est.

THÉORÈME 5.2 (Pogorelov [Pog73]). — *Soit P un polyèdre hyperbolique (resp. sphérique). Alors P est infinitésimalement rigide si et seulement si son image dans le modèle projectif de H^3 (resp. de S^3) l'est.*

La relation avec le théorème précédent tient au fait que composer à droite les applications ρ_H ou ρ_S par une isométrie revient à composer à gauche par une transformation projective de \mathbf{R}^3 .

En d'autres termes, les images de P dans \mathbf{R}^3 pour deux choix différents du modèle projectif diffèrent par une transformation projective.

Polyèdres convexes. En utilisant le théorème précédent et le théorème de rigidité de Cauchy, on obtient immédiatement l'énoncé suivant.

COROLLAIRE 5.3. — *Les polyèdres convexes hyperboliques (resp. sphériques) sont infinitésimalement rigides.*

Dualité. Dans S^3 et H^3 , la rigidité des polyèdres convexes admet une propriété « duale » : les polyèdres convexes sont aussi déterminés par leur « métrique duale », une métrique à courbure constante 1 et à singularités coniques pour laquelle la somme des angles autour des points singuliers est supérieure à 2π pour H^3 , inférieure pour S^3 . La métrique duale d'un polyèdre est partiellement déterminée par ses angles dièdres.

Ces résultats s'expliquent par des phénomènes de dualité, d'une part à l'intérieur de la classe des polyèdres convexes de S^3 , d'autre part entre les polyèdres convexes de H^3 et les polyèdres convexes de type espace de l'espace de Sitter de dimension trois. Le phénomène hyperbolique a été mis à jour par Rivin et Hodgson [RH93], il est aussi relié aux travaux d'Andreev sur les angles dièdres des polyèdres hyperboliques convexes [And70, And71]. Dans \mathbf{R}^3 , ce phénomène de dualité prend une forme dégénérée, par exemple le phénomène dual de la rigidité infinitésimale des polyèdres convexes est la rigidité dans le problème de Minkowski.

Par contre, on a un analogue hyperbolique très simple de la conjecture de Stoker citée ci-dessus.

CONJECTURE 5.4. — *Soit $P \subset H^3$ un polyèdre convexe. Si une déformation infinitésimale de P ne change pas, au premier ordre, ses angles dièdres, alors elle est triviale.*

Notons d'ailleurs que l'analogie sphérique de cette conjecture est faux.

S'il existait un polyèdre P admettant une déformation infinitésimale qui ne change pas, au premier ordre, ses angles dièdres, on pourrait en prendre deux copies et les recoller en identifiant leurs bords. On obtiendrait ainsi une cône-variété hyperbolique M (voir les notes de Thurston [Thu97] pour une définition) admettant une déformation infinitésimale non triviale qui ne change pas l'angle autour des singularités.

De plus, la convexité de P signifierait que les angles autour des singularités de M serait partout inférieurs strictement à 2π , et on peut supposer qu'une cône-variété satisfaisant cette condition n'admet aucune déformation qui ne change pas les angles autour de ses singularités. C'est du moins un résultat récent de Hodgson et Kerckhoff pour les cônes-variétés dont le lieu singulier est un link, c'est-à-dire une réunion disjointe de courbes fermées.

Ces considérations de rigidité interviennent dans le théorème d'hyperbolisation des orbifolds de Thurston, voir [BP01] ou [CHK00].

Flexibilité et volume. On peut construire des exemples simples de polyèdres flexibles dans la sphère ; il suffit de considérer un polygone $p \subset S^2$, puis de prendre la suspension sphérique de p (ce qui revient à considérer p comme contenu dans un S^2 «équatorial» et à ajouter les deux pôles). Les flexions correspondent aux déformations isométriques de p dans S^2 . V. Alexandrov [Ale97] a remarqué que le volume intérieur de ces polyèdres varie au cours de ces flexions. Le théorème 2.1 est donc faux dans S^3 . Néanmoins ces exemples sont assez particuliers, et il n'est pas exclu qu'une version affaiblie du théorème soit valable, et/ou que l'analogie du théorème 2.1 soit vrai dans l'espace hyperbolique H^3 .

5.3. Dimensions supérieures

On ne connaît pas d'exemple de polyèdre plongé ou même immergé flexible dans \mathbf{R}^4 . Par contre, R. Connelly a donné un exemple très simple de polyèdre flexible. Le théorème 2.3 s'étend à ces polyèdres, cela a été montré par A. Walz.

5.4. Invariants de Dehn

Le troisième des vingt-trois problèmes posés par Hilbert au début du vingtième siècle concerne la possibilité, étant donné deux polyèdres P et P' de même volume dans \mathbf{R}^3 , de découper P et de recomposer les morceaux pour obtenir P' . Cette question a été résolue dès 1901 par Dehn [Deh01] : Dehn a généralisé un invariant introduit auparavant par Bricard [Bri96], appelé invariant de Dehn, préservé par l'opération de découpage et de recomposition. En fait deux polyèdres de \mathbf{R}^3 sont équivalents par découpage et recomposition si et seulement s'ils ont même volume et même invariant de Dehn, c'est un résultat de Sydler [Syd65, Jen68, Car86] qui a été montré en 1965.

R. Connelly demande si, lorsqu'on déforme un polyèdre flexible, son invariant de Dehn reste constant.

L'analogie du résultat de Dehn pour les polyèdres sphériques et hyperboliques reste inconnu. I. Sabitov suggère que cela pourrait être rapproché du fait que, pour ces polyèdres, le volume n'est pas fixé lors des déformations isométriques.

Remerciements. L'auteur tient à remercier Victor Alexandrov, Étienne Ghys, Idjad Sabitov et Rabah Souam pour de nombreuses remarques sur le texte.

BIBLIOGRAPHIE

- [Ale58] A.D. Alexandrov. *Konvexe polyeder*. Akademie-Verlag, Berlin, 1958.
- [Ale97] V. Alexandrov. An example of a flexible polyhedron with nonconstant volume in the spherical space. *Beiträge Algebra Geom.*, 38(1):11–18, 1997.
- [Ale01] V. Alexandrov. Flexible polyhedra in Minkowski 3-space. Preprint Jussieu, math.MG/0111003, 2001.
- [And70] E.M. Andreev. Convex polyhedra in Lobacevskii space. *Mat. Sb.(N.S.)*, 81(123):445–478, 1970.
- [And71] E.M. Andreev. On convex polyhedra of finite volume in Lobacevskii space. *Math. USSR Sbornik*, 12(3):225–259, 1971.
- [AS99] A.V. Astrelin and I.Kh. Sabitov. A canonical polynomial for the volume of a polyhedron. *Uspekhi Mat. Nauk*, 54(2(326)):165–166, 1999.
- [Ber77] M. Berger. *Géométrie. Vol. 3*. CEDIC, Paris, 1977. Convexes et polytopes, polyèdres réguliers, aires et volumes.
- [BP01] M. Boileau and J. Porti. Geometrization of 3-orbifolds of cyclic type. *Astérisque*, 272:208, 2001. Appendix A by Michael Heusener and Joan Porti.
- [Bri96] R. Bricard. Sur une question de géométrie relative aux polyèdres. *Nouv. Ann. Math.*, 15:331–334, 1896.
- [Bri97] R. Bricard. Mémoire sur la théorie de l'octaèdre articulé. *J. Math. Pur. Appl., Liouville*, 3:113–148, 1897.
- [BZ95] Yu. D. Burago and V. A. Zalgaller. Isometric piecewise-linear embeddings of two-dimensional manifolds with a polyhedral metric into \mathbf{R}^3 . *Algebra i Analiz*, 7(3):76–95, 1995.
- [Car86] P. Cartier. Décomposition des polyèdres: le point sur le troisième problème de Hilbert. *Astérisque*, 133-134:261–288, 1986. Sémin. Bourbaki (1984/85), Exposé 646.
- [Cau13] A.L. Cauchy. Sur les polygones et polyèdres, second mémoire. *Journal de l'École Polytechnique*, 19:87–98, 1813.
- [CHK00] D. Cooper, C.D. Hodgson, and S.P. Kerckhoff. *Three-dimensional orbifolds and cone-manifolds*, volume 5 of *MSJ Memoirs*. Mathematical Society of Japan, Tokyo, 2000. With a postface by Sadayoshi Kojima.

- [Con77] R. Connelly. A counterexample to the rigidity conjecture for polyhedra. *Inst. Haut. Étud. Sci., Publ. Math.*, 47:333–338, 1977.
- [Con80] R. Connelly. Conjectures and open questions in rigidity. In *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, pages 407–414, Helsinki, 1980. Acad. Sci. Fennica.
- [Con93] R. Connelly. Rigidity. In *Handbook of convex geometry, Vol. A, B*, pages 223–271. North-Holland, Amsterdam, 1993.
- [Con79] R. Connelly. A flexible sphere. *Math. Intelligencer*, 1(3):130–131, 1978/79.
- [CS94] R. Connelly and H. Servatius. Higher-order rigidity—what is the proper definition? *Discrete Comput. Geom.*, 11(2):193–200, 1994.
- [CSW97] R. Connelly, I. Sabitov, and A. Walz. The Bellows conjecture. *Beitr. Algebra Geom.*, 38:1–10, 1997. <http://www.emis.de/journals/BAG/>.
- [CW94] H. Crapo and W. Whiteley. Spaces of stresses, projections and parallel drawings for spherical polyhedra. *Beiträge Algebra Geom.*, 35(2):259–281, 1994.
- [Dar93] G. Darboux. *Leçons sur la théorie générale des surfaces. III, IV*. Éditions Jacques Gabay, Sceaux, 1993. Reprint of the 1894 original (III) and the 1896 original (IV), Cours de Géométrie de la Faculté des Sciences.
- [Deh01] M. Dehn. Über den Rauminhalt. *Math. Ann.*, 105:465–478, 1901.
- [Deh16] M. Dehn. Über den Starrheit konvexer Polyeder. *Math. Ann.*, 77:466–473, 1916.
- [Glu75] H. Gluck. Almost all simply connected closed surfaces are rigid. In *Geometric topology (Proc. Conf., Park City, Utah, 1974)*, pages 225–239. Lecture Notes in Math., Vol. 438. Springer, Berlin, 1975.
- [Jen68] B. Jensen. The algebra of polyhedra and the Dehn-Sydler theorem. *Math. Scand.*, 22:241–256, 1968.
- [Kui79] N.H. Kuiper. Sphères polyédriques flexibles dans E^3 , d’après Robert Connelly. In *Séminaire Bourbaki, 30e année (1977/78)*, volume 710 of *Lecture Notes in Math.*, pages 147–168, Exposé 514. Springer, Berlin, 1979.
- [Lan72] S. Lang. *Introduction to algebraic geometry*. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1972. Third printing, with corrections.
- [Leb09] H. Lebesgue. Démonstration complète du théorème de Cauchy sur l’égalité des polyèdres convexes. *Intermédiaire des Mathématiciens*, 16:113–120, 1909.
- [LegII] A.M. Legendre. *Éléments de géométrie*. Paris, 1793 (an II). Première édition, note XII, pages 321–334.
- [Pog73] A.V. Pogorelov. *Extrinsic Geometry of Convex Surfaces*. American Mathematical Society, 1973. Translations of Mathematical Monographs. Vol. 35.
- [RH93] I. Rivin and C. D. Hodgson. A characterization of compact convex polyhedra in hyperbolic 3-space. *Invent. Math.*, 111:77–111, 1993.
- [RR00] L. Rodríguez and H. Rosenberg. Rigidity of certain polyhedra in \mathbf{R}^3 . *Comment. Math. Helv.*, 75(3):478–503, 2000.

- [Sab92] I.Kh. Sabitov. Local theory of bendings of surfaces. In *Geometry, III*, volume 48 of *Encyclopaedia Math. Sci.*, pages 179–256. Springer, Berlin, 1992.
- [Sab96] I.Kh. Sabitov. The volume of a polyhedron as a function of its metric. *Fundam. Prikl. Mat.*, 2(4):1235–1246, 1996.
- [Sab98a] I.Kh. Sabitov. The generalized Heron-Tartaglia formula and some of its consequences. *Mat. Sb.*, 189(10):105–134, 1998.
- [Sab98b] I.Kh. Sabitov. A proof of the “bellows” conjecture for polyhedra of low topological genus. *Dokl. Akad. Nauk*, 358(6):743–746, 1998.
- [Sab98c] I.Kh. Sabitov. The volume as a metric invariant of polyhedra. *Discrete Comput. Geom.*, 20(4):405–425, 1998.
- [Sab01] I.Kh. Sabitov. Calculation of polyhedra. *Dokl. Akad. Nauk*, 377(2):161–164, 2001.
- [Sab02] I.Kh. Sabitov. Algorithmic solution of the problem of the isometric realization of two-dimensional polyhedral metrics. *Izv. Ross. Akad. Nauk Ser. Mat.*, 66(2):159–172, 2002.
- [Sau35] R. Sauer. Infinitesimale Verbiegungen zueinander projektiver Fläschen. *Math. Ann*, 111:71–82, 1935.
- [Ste16] E. Steinitz. Polyeder und Raumeinteilungen. *Encycl. Math. Wiss.*, 3:1–139, 1916.
- [Sto68] J.J. Stoker. Geometrical problems concerning polyhedra in the large. *Comm. Pure Appl. Math.*, 21:119–168, 1968.
- [Syd65] J.-P. Sydler. Conditions nécessaires et suffisantes pour l'équivalence des polyèdres de l'espace euclidien à trois dimensions. *Comment. Math. Helv.*, 40:43–80, 1965.
- [Thu97] W.P. Thurston. Three-dimensional geometry and topology. Recent version of the 1980 notes. <http://www.msri.org/publications/books/gt3m/>, 1997.
- [TW00] T.-S. Tay and W. Whiteley. A homological interpretation of skeletal rigidity. *Adv. in Appl. Math.*, 25(1):102–151, 2000.

Jean-Marc SCHLENKER

Laboratoire Émile Picard

UMR CNRS 5580

UFR MIG

Université Paul Sabatier

F-31062 Toulouse cedex 4

E-mail : schlenker@picard.ups-tlse.fr

NOUVELLES APPROCHES
DE LA PROPRIÉTÉ (T) DE KAZHDAN

par Alain VALETTE

INTRODUCTION

La propriété (T) pour un groupe localement compact G est une forme de rigidité de G en théorie des représentations : il est impossible de déformer la représentation triviale de dimension 1 de G parmi les représentations unitaires de G . Elle a été énoncée en 1967 par D. Kazhdan [35], qui s'intéressait des variétés riemanniennes M localement symétriques, irréductibles, de rang au moins 2, de volume fini. En d'autres termes, $M = \Gamma \backslash G / K$, où G est un groupe de Lie simple centre fini, de rang rel ≥ 2 , K est un sous-groupe compact maximal de G , et Γ est un réseau sans torsion de G . Les principaux résultats de l'article de Kazhdan sont les suivants :

- Le groupe fondamental $\Gamma = \pi_1(M)$ est de type fini (quand M n'est pas compact, ce résultat est loin d'être trivial!).
- Le premier nombre de Betti $b_1(M) = \dim_{\mathbb{R}} \text{Hom}(\Gamma, \mathbb{R})$ est nul.

Le principal mérite de Kazhdan est d'avoir vu que ces deux propriétés de M dépendent en fait de la structure des représentations unitaires du groupe de Lie G ambiant (voir les Théorèmes 1 et 2 ci-dessous pour les nonces précis).

La propriété (T) a trouvé une série d'applications, qui vont de la théorie des graphes à la théorie ergodique (voir [15]). J'en isolerai une seule : le théorème de Margulis sur les sous-groupes normaux des réseaux Γ en rang ≥ 2 (voir [38], Theorem 4.9) : un sous-groupe normal $N \triangleleft \Gamma$ est soit central dans G (et donc fini), soit d'indice fini dans Γ . Pour cela, on montre que, si N est infini, alors Γ/N est moyennable. Or Γ/N a aussi la propriété (T) (car Γ l'a), et un groupe dénombrable, qui est moyennable et a la propriété (T), est nécessairement fini¹. À ma connaissance, toutes les preuves connues de ce résultat passent par la propriété (T).

¹Pour la preuve, voir l'Exemple 1 de la section 1. Il est important que la preuve ne puisse pas être effective, puisque Γ est résiduellement fini, donc admet des quotients finis arbitrairement grands.

Le texte se présente comme suit : après un résumé de l'article original de Kazhdan [35] au chapitre 1, nous introduisons au chapitre 2 la caractérisation cohomologique de la propriété (T). Les résultats récents mentionnés dans ce texte sont quasiment tous des conséquences de la caractérisation cohomologique : actions par difféomorphismes du cercle (section 2.4 et Théorème 5), caractérisation de la propriété (T) par l'annulation de la 1-cohomologie réduite (section 3.1), lien avec les applications harmoniques et preuve de la propriété (T) pour $Sp(n, 1)$ (section 3.2), non invariance de la propriété (T) par quasi-isométries (section 3.3), critère spectral pour la propriété (T) (chapitre 4).

La propriété (T) de Kazhdan a été exposée chez Bourbaki dans l'exposé 343 de Delaroché-Kirillov [16], et partiellement dans l'exposé 778 de Pansu [41]. Les progrès depuis 1993 ont été tels que j'ai dû me livrer à des choix : faute de temps ou faute de goût, j'ai choisi de ne pas parler du lien entre propriété (T) et génération bornée (Shalom [49]), de l'invariance par équivalence de mesure de la propriété (T) (Furman [21]), de l'industrie des constantes de Kazhdan (voir les références chez Gelfand-Zuk [23]), ou des groupes alatoires (Gromov [26], Zuk [55]) - ce dernier aspect devrait faire l'objet d'un prochain Séminaire Bourbaki par J. Ghys.

1. DÉFINITIONS ET EXEMPLES DE BASE

Soient G un groupe localement compact, et π une représentation unitaire, fortement continue, de G sur un espace de Hilbert \mathcal{H} π .

DÉFINITION 1. - La représentation π possède presque des vecteurs invariants si, pour tout $\epsilon > 0$ et toute partie compacte $K \subset G$, il existe un vecteur $\xi \in \mathcal{H}$ π tel que

$$\max_{g \in K} \|\pi(g)\xi - \xi\| < \epsilon \|\xi\|.$$

Exemple 1 : 1) Soit G un groupe compact. Toute représentation π de G qui possède presque des vecteurs invariants, possède des vecteurs invariants non nuls. En effet, soit ξ un vecteur-unité de \mathcal{H} π tel que

$$\max_{g \in G} \|\pi(g)\xi - \xi\| < 1.$$

Posons $\eta = \int_G \pi(g)\xi \, dg$, où dg désigne la mesure de Haar normalisée sur G . Le vecteur η est clairement invariant. Pour montrer qu'il est non nul, on remarque que $\|\eta - \xi\| = \left\| \int_G (\pi(g)\xi - \xi) \, dg \right\| < 1$. Comme $\|\xi\| = 1$, on a $\eta \neq 0$.

2) Notons λ_G la représentation régulière gauche du groupe localement compact G sur l'espace de Hilbert $L^2(G)$. D'une part, on observe que λ_G a des vecteurs invariants non nuls si et seulement si G est compact. D'autre part, c'est un résultat classique de Hulanicki [33] et Reiter [45] que λ_G a presque des vecteurs invariants si et seulement si G est moyennable.

Donc, si G est moyennable non compact (par exemple $G = \mathbf{Z}$), la représentation λ G fournit un exemple de représentation ayant presque des vecteurs invariants, mais sans vecteur invariant non nul.

DEFINITION 2. - Un groupe localement compact G a la propriété (T), ou est un groupe de Kazhdan, si toute représentation de G qui possède presque des vecteurs invariants, possède des vecteurs invariants non nuls.

L'Exemple 1 montre d'une part que les groupes compacts ont trivialement la propriété (T), d'autre part que les groupes moyennables non compacts n'ont pas la propriété (T) (ceci a été vu dans l'introduction).

Les principales propriétés structurelles des groupes de Kazhdan sont rassemblées dans le résultat suivant, dû à Kazhdan [35] (voir aussi [16], [15]).

THEOREME 1. - Soit G un groupe de Kazhdan.

- i) G est compactement engendré.
- ii) Le quotient $G/[\overline{G, G}]$ de G par l'adhérence du sous-groupe des commutateurs est compact.
- iii) Si H est un sous-groupe fermé de co-volume fini dans G , alors H est un groupe de Kazhdan.

En particulier, soit Γ un réseau dans G , c'est-à-dire un sous-groupe discret de co-volume fini dans G . Si G a la propriété (T), le Théorème 1 implique que Γ , ainsi que tous ses sous-groupes d'indice fini, sont finiment engendrés avec un abélien fini.

Exemple 2 : Voyons comment cette remarque peut être exploitée pour montrer que certains groupes classiques n'ont pas la propriété (T). D'abord, le groupe libre \mathbf{F}_n sur n générateurs ($n \geq 1$) n'a pas la propriété (T) car son abélien est infini. Ensuite, $SL_2(\mathbf{Z})$ n'a pas la propriété (T) car il contient un sous-groupe libre d'indice fini (noter que l'abélien de $SL_2(\mathbf{Z})$ est fini!). En fait, $SL_2(\mathbf{R})$ n'a pas la propriété (T) car il contient $SL_2(\mathbf{Z})$ comme réseau.

Les principaux exemples de groupes de Kazhdan sont donnés par le résultat suivant, dû à Kazhdan [35] en rang ≥ 3 , et complété en rang 2 par Delaroché-Kirillov [16], Vaserstein [51], et Wang [53]. Un corps local est un corps commutatif localement compact non discret (penser $\mathbf{R}, \mathbf{C}, \mathbf{Q}_p$ et aux corps de séries de Laurent sur les corps finis).

THEOREME 2. - Soit K un corps local. Soit \mathbf{G} un groupe algébrique connexe, presque simple sur K , avec $K - \text{rang}(\mathbf{G}) \geq 2$. Le groupe $G = \mathbf{G}(K)$ des K -points de \mathbf{G} possède la propriété (T).

Exemple 3 : 1) Les groupes $SL_n(K)$ ($n \geq 3$) et $Sp_{2n}(K)$ ($n \geq 2$), ont la propriété (T).

2) Les groupes discrets $SL_n(\mathbf{Z})$ ($n \geq 3$) et $Sp_{2n}(\mathbf{Z})$ ($n \geq 2$), qui sont des réseaux respectivement dans $SL_n(\mathbf{R})$ et $Sp_{2n}(\mathbf{R})$, ont la propriété (T).

Les Thormes 1 et 2 impliquent les propri ts des varits riemanniennes localement symtriques qui ont t mentionnes dans l'introduction.

Le cas des groupes de rang 1 rserve une surprise : la situation n'est pas la mme selon que le corps \mathbb{K} est archimdien ou pas. Nous verrons au Corollaire 1 que toute action d'un groupe de Kazhdan sur un arbre possde un sommet \emptyset xe ou une arte \emptyset xe. Or, si \mathbb{K} est un corps local non archimdien et \mathbf{G} un \mathbb{K} -groupe algbrique connexe, presque simple avec $\mathbb{K} - rang(G) = 1$, le groupe $\mathbf{G}(\mathbb{K})$ agit proprement sur un arbre (savoir son immeuble de Bruhat-Tits, voir [8]). Donc $\mathbf{G}(\mathbb{K})$ n'a pas la propri t (T).

Considrons maintenant le cas archimdien. D'aprs la classi cation de Cartan (voir [32]), les groupes de Lie simples, connexes, de rang rel 1 sont localement isomorphes soit $SO(n, 1)$, soit $SU(n, 1)$, soit $Sp(n, 1)$ (avec chaque fois $n \geq 2$), soit en \emptyset n $F_{4(-20)}$. Gomtriquement, on ralise ces groupes comme groupes d'isomtries des espaces hyperboliques de dimension n respectivement sur \mathbb{R} , sur \mathbb{C} , sur l'algbre \mathbb{H} des quaternions de Hamilton, ou sur l'algbre Cay des octaves de Cayley (avec $n = 2$ dans ce dernier cas). Le rsultat suivant est d Kostant [36]. Nous esquisserons une preuve de la seconde moiti la section 3.2.

THORME 3. Soit G un groupe de Lie simple, connexe, de rang rel 1.

- i) Si G est localement isomorphe $SO(n, 1)$ ou $SU(n, 1)$ ($n \geq 2$), alors G n'a pas la propri t (T).
- ii) Si G est localement isomorphe $Sp(n, 1)$ ($n \geq 2$) ou $F_{4(-20)}$, alors G a la propri t (T).

Le groupe $Sp(n, 1)$ et ses rseaux exhibent un curieux mlange de rigidit et de non rigidit. Certains caractres les rapprochent des groupes de rang suprieur (propri t (T), super-rigidit, voir [13], [28]). D'autres attributs les rattachent au rang 1 (hyperbolicit des rseaux co-compacts [27], non-isolement de la representation triviale parmi les representations uniformment bornes, voir [14]).

Ces propri ts antagonistes ont t exploites : construction d'une in nit non d-nombrable de groupes de torsion non moyennables, par Gromov [27]; construction de C^* -algbres non quivalentes en K-thorie des algbres nuclaires, par Skandalis [50].

2. 1-COHOMOLOGIE ET ACTIONS AFFINES

2.1. D nitions

Soit π une representation unitaire du groupe localement compact G .

DEFINITION 3. a) L'espace des 1-cocycles valeurs dans π est

$$Z^1(G, \pi) = \{b : G \rightarrow \mathcal{H}_\pi : b \text{ continu}, b(gh) = \pi(g)b(h) + b(g) \forall g, h \in G\}.$$

b) L'espace des 1-cobords à valeurs dans π est

$$B^1(G, \pi) = \{b \in Z^1(G, \pi) : \exists \xi \in \mathcal{H}_\pi : b(g) = \pi(g)\xi - \xi, \forall g \in G\}.$$

c) Le groupe de 1-cohomologie à coefficients dans π est le quotient

$$H^1(G, \pi) = Z^1(G, \pi)/B^1(G, \pi).$$

d) Munissons $Z^1(G, \pi)$ de la topologie de la convergence uniforme sur les compacts. Le groupe de 1-cohomologie réduite à coefficients dans π est le quotient

$$\overline{H^1}(G, \pi) = Z^1(G, \pi)/\overline{B^1(G, \pi)},$$

où $\overline{B^1(G, \pi)}$ désigne la fermeture de $B^1(G, \pi)$.

Ces notions s'interprètent en termes d'actions isométriques à gauche de G sur \mathcal{H}_π .

- Si $b : G \rightarrow \mathcal{H}_\pi$ est une application continue, associons à $b(g)$ l'isométrie à gauche de \mathcal{H}_π :

$$\alpha(g)v = \pi(g)v + b(g) \quad (v \in \mathcal{H}_\pi).$$

L'application b est un 1-cocycle si et seulement si $\alpha(gh) = \alpha(g)\alpha(h)$ pour tous $g, h \in G$, c'est-à-dire si α est une action isométrique à gauche de G , de partie linéaire π .

- Le 1-cocycle b est un 1-cobord si et seulement si l'action à gauche α possède un point fixe, c'est-à-dire si α est conjugué à π par une translation. Le lemme du centre (voir [15]) dit qu'un cocycle b est un cobord si et seulement si b est borné comme fonction sur G .
- $H^1(G, \pi) = 0$ si et seulement si toute action isométrique à gauche de partie linéaire π possède un point fixe.
- $\overline{H^1}(G, \pi) = 0$ si et seulement si toute action isométrique à gauche de partie linéaire π possède presque des points fixes.

Les trois sections suivantes sont consacrées à des exemples où apparaissent naturellement des représentations ayant de la 1-cohomologie non nulle.

2.2. Représentations ayant presque des vecteurs invariants

Si π est une représentation unitaire de G , nous notons $\infty\pi$ la somme directe hilbertienne d'une infinité dénombrable de copies de π .

PROPOSITION 1. Soit G un groupe localement compact σ -compact. Soit π une représentation de G , possédant presque des vecteurs invariants, mais pas de vecteur invariant non nul. Alors $H^1(G, \infty\pi) \neq 0$.

Preuve : Soit $(K_n)_{n \geq 1}$ une suite croissante de parties compactes, recouvrant G . Comme π possede presque des vecteurs invariants, on trouve une suite $(\xi_n)_{n \geq 1}$ de vecteurs de norme 1 dans \mathcal{H}_π tels que

$$\max_{g \in K_n} \|\pi(g)\xi_n - \xi_n\| < 2^{-n}$$

pour tout $n \geq 1$. On pose alors, pour $g \in G$:

$$b(g) = \bigoplus_{n=1}^{\infty} n(\pi(g)\xi_n - \xi_n).$$

Cette somme converge dans $\mathcal{H}_\pi \oplus \mathcal{H}_\pi \oplus \dots$, uniformment sur tout compact de G . On a donc $b \in Z^1(G, \infty\pi)$. Pour montrer que b n'est pas un cobord, montrons que b n'est pas born sur G . Si b tait born par une constante R , on aurait pour tout $n \geq 1$ et tout $g \in G$:

$$n\|\pi(g)\xi_n - \xi_n\| \leq R.$$

Prenons n assez grand pour avoir, pour tout $g \in G$:

$$\|\pi(g)\xi_n - \xi_n\| \leq 1$$

c'est--dire

$$\frac{1}{2} \leq \operatorname{Re}\langle \pi(g)\xi_n | \xi_n \rangle.$$

Notons alors C l'enveloppe convexe ferme de $\pi(G)\xi_n$ dans \mathcal{H}_π : elle est invariante par $\pi(G)$, et pour tout $\eta \in C$ on a $\frac{1}{2} \leq \operatorname{Re}\langle \eta | \xi_n \rangle$. L'unique vecteur de norme minimale dans C est donc non nul, et invariant par $\pi(G)$, ce qui contredit nos hypothses. \triangleright

2.3. Actions sur des arbres

Soit X un arbre. Notons V l'ensemble des sommets, et $d(x, y)$ la distance entre les sommets x et y de X . Notons encore E l'ensemble des artes de X , et \mathbf{E} l'ensemble des artes orientes : chaque arte gomtrique $e \in E$ apparat avec deux orientations opposes $e^+, e^- \in \mathbf{E}$. Notons \mathcal{H}_X l'espace des fonctions antisymtriques et carr sommable sur \mathbf{E} :

$$\mathcal{H}_X = \{\xi \in \ell^2(\mathbf{E}) : \xi(e^+) = -\xi(e^-), \forall e \in E\}.$$

Si G est un groupe localement compact agissant par automorphismes sur X , on note σ_X la representation correspondante de G sur \mathcal{H}_X .

PROPOSITION 2.- Si G ne øxe ni sommet ni arte de X , alors $H^1(G, \sigma_X) \neq 0$.

Preuve : Dønissons une application $c : V \times V \rightarrow \mathcal{H}_X$: pour $x, y \in V, e^+ \in \mathbf{E}$, on pose $c(x, y)(e^+) = 0$ si e^+ n'est pas sur l'unique godsique joignant x y , et sinon :

$$c(x, y)(e^+) = \begin{cases} 1 & \text{si } e^+ \text{ pointe de } x \text{ vers } y; \\ -1 & \text{si } e^+ \text{ pointe de } y \text{ vers } x. \end{cases}$$

L'application c est G -quivariante :

$$c(gx, gy) = \sigma_X(g)(c(x, y)) \quad (g \in G; x, y \in V).$$

Comme les triangles dans un arbre sont d'gnrs (ce sont des tripodes), l'application c vrie la relation de Chasles :

$$c(x, y) + c(y, z) = c(x, z) \quad (x, y, z \in V).$$

Fixons alors un sommet de base x_0 dans V , et posons, pour $g \in G$:

$$b(g) = c(gx_0, x_0).$$

Grce aux deux propriets de c , on vrie immdiatement que $b : G \rightarrow \mathcal{H}$ est un 1-cocycle valeurs dans σ_X .

Raisonnons alors par contraposition, et supposons que $H^1(G, \sigma_X) = 0$. Le cocycle b est donc born sur G . Comme

$$\|c(x, y)\|_2^2 = 2d(x, y)$$

pour tous $x, y \in V$, on en dduit que l'orbite de x_0 sous G est borne. Un lemme lmentaire (Proposition 19 du Chapitre I de [47]) permet alors de conclure que G fixe soit un sommet, soit une arce de X . ▷

2.4. Actions par diceomorphismes sur le cercle

Notons S^1 le cercle-unit de \mathbf{R}^2 , paramtr par la longueur d'arc. Notons \mathcal{H} l'espace de Hilbert des noyaux K carr intgrable sur $S^1 \times S^1$ par rapport la mesure de Lebesgue, et antisymtriques : donc $K(\theta, \phi) = -K(\phi, \theta)$, pour presque tout $(\theta, \phi) \in S^1 \times S^1$.

Notons $\text{Dice}_+^{1+\alpha}(S^1)$ le groupe des diceomorphismes de S^1 de classe $1 + \alpha$, avec $\alpha \geq 0$. C'est le groupe des diceomorphismes f du cercle, de classe C^1 , tels que f' et $(f^{-1})'$ satisfont une condition de Hlder d'exposant α .

On fait agir $\text{Dice}_+^{1+\alpha}(S^1)$ sur \mathcal{H} au moyen de la representation unitaire σ : pour $K \in \mathcal{H}, g \in \text{Dice}_+^{1+\alpha}(S^1)$:

$$(\sigma(g)K)(\theta, \phi) = \sqrt{(g^{-1})'(\theta)(g^{-1})'(\phi)}K(g^{-1}.\theta, g^{-1}.\phi).$$

Si Γ est un groupe et $\Phi : \Gamma \rightarrow \text{Dice}_+^{1+\alpha}(S^1)$ un homomorphisme, notons $\sigma_\Phi = \sigma \circ \Phi$ la representation unitaire correspondante de Γ sur \mathcal{H} .

Considrons maintenant le noyau antisymtrique (mais pas carr intgrable!) donn par :

$$F(\theta, \phi) = \frac{1}{2 \tan(\frac{\theta-\phi}{2})}.$$

Pour $g \in \Gamma$, d'onnons formellement

$$b_\Phi(g) = \sigma_\Phi(g)F - F.$$

En observant qu'au voisinage de la diagonale on a

$$(1) \quad F(\theta, \phi) = \frac{1}{\theta - \phi} + K_0(\theta, \phi),$$

o K_0 est un noyau continu, on dmontre facilement le lemme suivant, d Pressley et Segal [43].

LEMME 1.- Pour $\alpha > \frac{1}{2}$, la fonction $b_\Phi(g)$ est carr intgrable sur $S^1 \times S^1$; donc $b_\Phi : \Gamma \rightarrow \mathcal{H}$ dñit un 1-cocycle dans $Z^1(\Gamma, \sigma_\Phi)$. \triangleright

Notons Δ la diagonale de $S^1 \times S^1$. La terminologie suivante se justiø par le fait que le quotient de $(S^1 \times S^1) \setminus \Delta$ par la volte $(\theta, \phi) \mapsto (\phi, \theta)$, s'identifiø l'espace des godsiques du disque de Poincar.

DEFINITION 4.- Un courant godsique est une mesure μ borlienne, positive, rgulire sur $(S^1 \times S^1) \setminus \Delta$, telle que

$$\mu([a, b] \times [c, d]) = \mu([c, d] \times [a, b])$$

chaque fois que a, b, c, d sont quatre points distincts, cycliquement ordonnés sur S^1 .

Exemple 4 : La mesure $d\mu_0(\theta, \phi) = \frac{d\theta d\phi}{4 \sin^2(\frac{\theta - \phi}{2})}$ est un courant godsique, invariant par l'action par homographies de $SU(1, 1) \simeq SL_2(\mathbb{R})$ sur S^1 , vu comme bord du disque de Poincar.

PROPOSITION 3.- Soit $\alpha > \frac{1}{2}$ et $\Phi : \Gamma \rightarrow \text{Dico}_+^{1+\alpha}(S^1)$. Si b_Φ appartient $B^1(\Gamma, \sigma_\Phi)$, alors $\Phi(\Gamma)$ laisse invariant un courant godsique μ , qui de plus satisfait :

$$(2) \quad \mu([a, a] \times [b, c]) = 0$$

$$(3) \quad \mu([a, b] \times (b, c]) = \infty$$

chaque fois que a, b, c sont trois points distincts, cycliquement ordonnés de S^1 .

Preuve :crivons $b_\Phi(g) = \sigma_\Phi(g)K - K$, pour $K \in \mathcal{H}$. Comme b_Φ est aussi donn par $b_\Phi(g) = \sigma_\Phi(g)F - F$, on en tire, pour tout $g \in \Gamma$:

$$\sigma_\Phi(g)(F - K) = F - K.$$

En d'autres termes :

$$\sqrt{(\Phi(g)^{-1})'(\theta)(\Phi(g)^{-1})'(\phi)}(F - K)(\Phi(g)^{-1}.\theta, \Phi(g)^{-1}.\phi) = (F - K)(\theta, \phi).$$

En levant au carr :

$$(\Phi(g)^{-1})'(\theta)(\Phi(g)^{-1})'(\phi)(F - K)^2(\Phi(g)^{-1}.\theta, \Phi(g)^{-1}.\phi) = (F - K)^2(\theta, \phi).$$

Ceci exprime l'invariance par $\Phi(\Gamma)$ du courant godsique

$$d\mu(\theta, \phi) = (F - K)^2(\theta, \phi)d\theta d\phi.$$

Comme μ est absolument continu par rapport à la mesure de Lebesgue $d\theta d\phi$, la propriété (2) est immédiate. Pour vérifier la propriété (3), on utilise l'équation (1) et le fait que $\frac{1}{\theta-\phi}$ n'est pas carré intégrable au voisinage de la diagonale Δ . ▷

Notons encore, sans démonstration, le résultat suivant, dû à A. Navas [40]. Il généralise le fait classique qu'une homographie du cercle qui fixe 3 points, est l'identité.

PROPOSITION 4. Soit h un homomorphisme préservant l'orientation de S^1 . Si h fixe trois points de S^1 et laisse invariant un courant godsonique possédant les propriétés (2) et (3), alors h est l'identité de S^1 . ▷

2.5. Thorme de Delorme-Guichardet

La caractérisation cohomologique suivante de la propriété (T) est due à Delorme [17] et Guichardet [29].

THORME 4. Soit G un groupe localement compact σ -compact. Les notions suivantes sont équivalentes :

- i) G a la propriété (T).
- ii) Pour toute représentation unitaire π de G , on a $H^1(G, \pi) = 0$.
- iii) Toute action isométrique adéquate de G sur un espace de Hilbert possède un point fixe.

Propos de la preuve : L'équivalence (ii) \Leftrightarrow (iii) résulte des commentaires faits après la Définition 3. L'implication (ii) \Rightarrow (i) résulte de la Proposition 1. Pour la preuve de la réciproque, voir par exemple [15], Chapitre 4. ▷

De la Proposition 2 découle alors immédiatement le résultat suivant, dû indépendamment à Alperin [1] et Watatani [54].

COROLLAIRE 1. Toute action d'un groupe de Kazhdan sur un arbre fixe un sommet ou une arête. ▷

L'application suivante du Thorme de Delorme-Guichardet est un résultat récent de A. Navas [40]. Il s'agit d'une nouvelle obstruction à la propriété (T) : essentiellement, un groupe infini de difféomorphismes du cercle ne peut avoir la propriété (T). Ceci généralise des résultats de Reznikov ([46], Thorme 1.7 du Chapitre 2). Dans le cas des réseaux dans les groupes de Lie simples de rang au moins 2, on retrouve des résultats de Ghys [24] et Burger-Monod [9] (toutefois valables sous une hypothèse de régularité un peu plus faible).

THORME 5. Soit $\alpha > \frac{1}{2}$. Si Γ est un groupe dénombrable ayant la propriété (T), et $\Phi : \Gamma \rightarrow \text{Dice}_+^{1+\alpha}(S^1)$ est un homomorphisme, alors $\Phi(\Gamma)$ est un groupe cyclique fini.

tapes de la preuve :

1er pas : L'action de $\Phi(\Gamma)$ est libre. En effet, soit $x_0 \in S^1$, et soit $g \in \Gamma$ tel que $\Phi(g)$ fixe x_0 . Nous devons montrer que $\Phi(g)$ est l'identité.

L'argument qui suit, aussi élégant qu'astucieux, est dû à D. Witte. On considère l'application $p : S^1 \rightarrow S^1 : z \mapsto z^3$, vue comme un revêtement triple de S^1 . La théorie classique des revêtements fournit une extension centrale

$$0 \rightarrow \mathbb{Z}/3\mathbb{Z} \rightarrow \tilde{\Gamma} \xrightarrow{q} \Gamma \rightarrow 1,$$

ainsi qu'un homomorphisme $\tilde{\Phi} : \tilde{\Gamma} \rightarrow \text{Dicoe}_+^{1+\alpha}(S^1)$ qui relève Φ . D'une part, comme $\tilde{\Gamma}$ a la propriété (T), le groupe $\tilde{\Phi}(\tilde{\Gamma})$ laisse invariant un courant godsiatique possédant les propriétés (2) et (3). D'autre part, $\mathbb{Z}/3\mathbb{Z}$ agit sur S^1 par multiplication par les racines cubiques de l'unité, et les trois images inverses de g dans $\tilde{\Gamma}$ agissent sur la fibre $p^{-1}(x_0)$ par permutations cycliques. L'une d'entre elles, appelons-la \tilde{g} , fixe donc chacun des trois points de $p^{-1}(x_0)$. Par la Proposition 4, $\tilde{\Phi}(\tilde{g})$ est l'identité de S^1 , donc $\Phi(g)$ également.

2me pas : $\Phi(\Gamma)$ est abélien fini. Pour cela, on invoque un résultat classique de Hlder : un groupe d'homomorphismes préservant l'orientation de S^1 , et agissant librement sur S^1 , est nécessairement abélien (voir par exemple [31], Théorème 3.1.6). Comme $\Phi(\Gamma)$ est abélien et possède la propriété (T), $\Phi(\Gamma)$ est fini.

3me pas : $\Phi(\Gamma)$ est cyclique. En effet un sous-groupe fini F de $\text{Dicoe}_+^{1+\alpha}(S^1)$ est nécessairement cyclique : puisque F fixe la mesure $\frac{1}{|F|} \sum_{g \in F} g^*(d\theta)$, qui est équivalente à la mesure de Lebesgue $d\theta$, le groupe F est conjugué à un sous-groupe du rotateur de $d\theta$. Mais celui-ci n'est autre que le groupe $SO(2)$ des rotations, et tout sous-groupe fini de $SO(2)$ est cyclique.

Notons que, dans cette preuve, l'hypothèse $\alpha > \frac{1}{2}$ sert uniquement à invoquer le lemme 1. ▷

Exemple 5 : Soit G le groupe de R.J. Thompson, défini comme suit. C'est le groupe des homomorphismes f de \mathbb{R}/\mathbb{Z} , préservant l'orientation, linéaires par morceaux, tels que :

- f' a un nombre fini de points de discontinuité, en des rationnels dyadiques ;
- les pentes de f sont des puissances de 2 ;
- $f(0)$ est un rationnel dyadique.

Le groupe G est un groupe simple, infini, de présentation finie. Ghys et Sergiescu [25] ont démontré que G est conjugué à un groupe de difféomorphismes C^∞ de S^1 , et ont demandé si G a la propriété (T). Leur résultat, joint au Théorème 5, montre que la réponse est négative. ²

²Récemment, D. Farley [19] a montré un résultat plus fort : le groupe G de Thompson a la propriété de Haagerup, ou encore est a-T-menable au sens de Gromov, c'est-à-dire qu'il admet une action isométrique, métriquement propre sur un espace de Hilbert séparable. Pour des détails sur cette notion, voir [12].

3. GROUPES COMPACTEMENT ENGENDRÉS

3.1. Un thorme de Shalom

Commenons par un exemple.

Exemple 6 : Soit $\Gamma = \bigoplus_{n \in \mathbb{N}} \mathbb{Z}/2\mathbb{Z}$ la somme directe d'une infinité d'nombrable de copies du groupe deux lments. D'une part, le groupe Γ n'est pas de type fini, donc n'a pas la proprié (T). D'autre part, on vrie aisement que, pour toute représentation unitaire irréductible de Γ , c'est-à-dire tout caractre $\chi : \Gamma \rightarrow \{\pm 1\}$, on a $H^1(\Gamma, \chi) = 0$.

L'exemple 6 montre que, dans le thorme de Delorme-Guichardet (Thorme 4) on ne peut pas se restreindre à l'annulation de la 1-cohomologie coefficients dans les représentations irréductibles pour caractriser la proprié (T). Vershik et Karpuchev [52] ont demandé si, néanmoins, une telle caractrisation n'était pas possible pour les groupes gnration compacte. La rponse - affirmative - est donnée par le rsultat suivant, de Y. Shalom ([48], Thormes 0.2 et 6.1).

THORME 6. - Soit G un groupe localement compact, compactement engendr, sparable. Les propriés suivantes sont équivalentes :

- i) G a la proprié (T).
- ii) Pour toute représentation unitaire π de G , on a $H^1(G, \pi) = 0$.
- iii) Pour toute représentation unitaire irréductible σ de G , on a $H^1(G, \sigma) = 0$.
- iv) Pour toute représentation unitaire irréductible σ de G , on a $H^1(G, \sigma) = 0$.

Parties faciles de la preuve : L'implication (i) \Rightarrow (iv) rsulte du Thorme 4, si l'on remarque qu'un groupe compactement engendr est σ -compact. L'implication (iv) \Rightarrow (iii) est triviale. L'implication (iii) \Rightarrow (ii) utilise la thorie de la rduction : comme G est sparable, toute représentation unitaire de G peut s'crire comme intgrale directe de représentations irréductibles. On invoque alors un rsultat de Guichardet ([30], Chapitre 3, 2) : l'annulation de la 1-cohomologie rduite est prserve par intgrale directe.

Pour la preuve de l'implication (ii) \Rightarrow (i), osons une partie gnratrice compacte Q de G . Si π est une représentation unitaire de G sur un espace de Hilbert \mathcal{H} , on dñit une fonction continue

$$\delta : \mathcal{H} \rightarrow \mathbb{R}^+ : \xi \mapsto \max_{g \in Q} \|\pi(g)\xi - \xi\|.$$

Le lemme suivant est l'ingrdient principal pour oñir la preuve du Thorme 6.

LEMME 2. - Supposons que π ait presque des vecteurs invariants, mais pas de vecteur

invariant non nul. Pour tout $M > 1$, il existe un vecteur $v_M \in \mathcal{H}$ tel que $\delta(v_M) = 1$ et $\delta(u) > \frac{1}{2}$ pour tout $u \in \mathcal{H}$ tel que $\|u - v_M\| < M$.

Preuve : On va montrer que, pour tout $M > 1$, il existe $r > 0$ et $w_M \in \mathcal{H}$ tels que $\delta(w_M) = \frac{r}{M}$ et $\delta(u) > \frac{r}{2M}$ pour tout $u \in \mathcal{H}$ avec $\|u - w_M\| < r$. Il suffira alors de poser $v_M = \frac{Mw_M}{r}$ pour conclure.

En effet, comme π possède presque des vecteurs invariants, on trouve d'abord un vecteur $w_1 \in \mathcal{H}$ tel que $\|w_1\| > 1$ et $\delta(w_1) = \frac{1}{2M}$.

Si $\delta(u) > \frac{1}{4M}$ pour tout $u \in \mathcal{H}$ tel que $\|u - w_1\| < \frac{1}{2}$, alors $r = \frac{1}{2}$ et $w_M = w_1$ font l'affaire.

Sinon, il existe $u \in \mathcal{H}$ tel que $\|u - w_1\| < \frac{1}{2}$ et $\delta(u) \leq \frac{1}{4M}$. Par continuité de δ , on trouve $w_2 \in \mathcal{H}$ tel que $\|w_2 - w_1\| < \frac{1}{2}$ et $\delta(w_2) = \frac{1}{4M}$.

Si $\delta(u) > \frac{1}{8M}$ pour tout $u \in \mathcal{H}$ tel que $\|u - w_2\| < \frac{1}{4}$, alors $r = \frac{1}{4}$ et $w_M = w_2$ font l'affaire.

Sinon, il existe $u \in \mathcal{H}$ tel que $\|u - w_2\| < \frac{1}{4}$ et $\delta(u) \leq \frac{1}{8M}$. Par continuité de δ , on trouve $w_3 \in \mathcal{H}$ tel que $\|w_3 - w_2\| < \frac{1}{4}$ et $\delta(w_3) = \frac{1}{8M}$.

Cette construction par récurrence s'arrête en un nombre fini de pas. En effet, dans le cas contraire on construit une suite $(w_i)_{i \geq 1}$ dans \mathcal{H} , telle que $\|w_{i+1} - w_i\| < \frac{1}{2^i}$ et $\delta(w_{i+1}) = \frac{1}{2^{i+1}M}$ pour tout $i \geq 1$. Cette suite est une suite de Cauchy dans \mathcal{H} , elle converge vers un vecteur $w \in \mathcal{H}$, non nul (car $\|w_1\| > 1$) et tel que $\delta(w) = 0$ (par continuité de δ). Donc w est fixe par tout élément de Q . Comme Q engendre G , le vecteur w est invariant non nul, ce qui contredit notre hypothèse sur π . \triangleright

Il reste à montrer la preuve du Théorème 6 : Il reste à démontrer l'implication

(ii) \Rightarrow (i), ce qui se fait par contrapositive. Si G n'a pas la propriété (T), on trouve une représentation unitaire π de G , qui possède presque des vecteurs invariants mais pas de vecteurs invariants non nuls. Par le lemme 2, pour tout $m \in \mathbb{N}$ on trouve un vecteur v_m tel que $\delta(v_m) = 1$ et $\delta(u) > \frac{1}{2}$ pour tout u dans la boule de rayon m centrée en v_m .

Heuristiquement, la preuve consiste à considérer la suite des boules B_m centrées en v_m et de rayon m , munies de la restriction de l'action $\tilde{\cdot}$ de G (bien sûr, les B_m ne sont pas G -invariantes, mais 'presque' : comme $\delta(v_m) = 1$, les éléments de Q déplacent très peu le centre, par rapport au rayon). L'idée est d'extraire de cette suite une sous-suite convergente dans l'espace des G -espaces métriques : l'espace-limite sera un espace de Hilbert muni d'une action isométrique adéquate α de G , de partie linéaire σ . La condition $\delta(u) > \frac{1}{2}$, pour tout u dans la réunion des B_m , montre que α n'a pas presque des points fixes. Par l'interprétation de la 1-cohomologie réduite donnée dans la section 2.1, on en tire que $\overline{H^1}(G, \sigma) \neq 0$.

Cette heuristique se traduit en un argument d'analyse fonctionnelle classique, grâce à la notion de fonction conditionnellement de type négatif. Une fonction continue $\psi : G \rightarrow \mathbb{R}$ est conditionnellement de type négatif si $\psi(1) = 0$, $\psi(g) = \psi(g^{-1})$ pour tout $g \in G$, et

pour tous $n \in \mathbb{N}$, $g_1, \dots, g_n \in G$; $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ tels que $\sum_{i=1}^n \lambda_i = 0$:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \psi(g_i^{-1} g_j) \leq 0.$$

Les fonctions conditionnellement de type négatif sur G forment un cône convexe, fermé pour la topologie de la convergence uniforme sur les parties compactes de G . Si α est une action isométrique à gauche de G sur un espace de Hilbert \mathcal{H} , et $v \in \mathcal{H}$, la fonction $\psi(g) = \|\alpha(g)v - v\|^2$ est conditionnellement de type négatif. Cet exemple est en fait général : une construction de type Gelfand-Naimark-Segal (voir [15], chapitre 5) montre qu’une fonction ψ conditionnellement de type négatif sur G , on associe un espace de Hilbert \mathcal{H}_ψ et une action isométrique à gauche α_ψ de G sur \mathcal{H}_ψ , telle que l’orbite $\alpha_\psi(G)(0)$ engendre un sous-espace dense, et $\psi(g) = \|\alpha_\psi(g)(0)\|^2$. Shalom a montré (lemme 6.5 et corollaire 6.6 de [48]) que l’action α_ψ ne possède pas presque des points fixes si et seulement s’il existe $\epsilon > 0$ tel que, pour tout $n \in \mathbb{N}$, pour tous nombres réels positifs a_1, \dots, a_n avec $\sum_{i=1}^n a_i = 1$, et tous $g_1, \dots, g_n \in G$ il existe $g \in Q$ tel que :

$$(4) \quad \sum_{i=1}^n \sum_{j=1}^n a_i a_j (\psi(g_i^{-1} g g_j) - \psi(g_i^{-1} g_j)) \geq \epsilon.$$

Si $\psi(g) = \|\alpha(g)v - v\|^2$, et $u = \sum_{i=1}^n a_i \alpha(g_i)v$, l’équation (4) est équivalente :

$$(5) \quad \|\alpha(g)u - u\|^2 \geq \epsilon.$$

Reprenons la preuve de l’implication (ii) \Rightarrow (i) du Théorème 6, en supposant le groupe G discret ³. Posons $\psi_m(g) = \|\pi(g)v_m - v_m\|^2$: la fonction ψ_m est conditionnellement de type négatif sur G (on voit l’action linéaire π comme une action à gauche). Tout élément $g \in G$ peut s’écrire comme un mot de longueur k sur $Q \cup Q^{-1}$; par l’inégalité triangulaire, on en tire

$$(6) \quad \sqrt{\psi_m(g)} \leq k\delta(v_m) = k.$$

La suite $(\psi_m)_{m>1}$ est donc bornée sur les parties finies de G . Le théorème d’Ascoli-Arzelà permet d’extraire de cette suite une sous-suite convergente vers une fonction conditionnellement de type négatif ψ . Il reste à montrer que α_ψ n’a pas presque des points fixes.

Pour ce faire, nous allons montrer que l’inégalité (4) est satisfaite avec $\epsilon = \frac{1}{4}$. Si ce n’était pas le cas, on trouverait $a_1, \dots, a_n \in [0, 1]$, avec $\sum_{i=1}^n a_i = 1$, et $g_1, \dots, g_n \in G$ tels que pour tous $g \in Q$:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j (\psi(g_i^{-1} g g_j) - \psi(g_i^{-1} g_j)) < \frac{1}{4}.$$

³Le cas général exige de rappeler le lemme 2, pour assurer que la famille de fonctions $(g \mapsto \|\pi(g)v_m - v_m\|)_{m>1}$ est équicontinue sur les parties compactes de G .

La même inégalité est donc satisfaite par les fonctions ψ_m , pour une infinité d'indices m .

Par l'inégalité (5), elle se récrit, avec $u_m = \sum_{i=1}^n a_i \pi(g_i) v_m$:

$$\|\pi(g)u_m - u_m\|^2 < \frac{1}{4}$$

pour tout $g \in Q$, ou encore $\delta(u_m) < \frac{1}{2}$. Soit k tel que g_1, \dots, g_n s'écrivent comme des mots de longueur au plus k sur $Q \cup Q^{-1}$. Pour $m > k$, l'inégalité (6) montre que les $\pi(g_1)v_m, \dots, \pi(g_n)v_m$ sont dans la boule B_m , et donc $u_m \in B_m$. Mais le lemme 2 implique alors que $\delta(u_m) > \frac{1}{2}$: en prenant m assez grand, on arrive ainsi à une contradiction. \triangleright

Shalom a aussi obtenu une conséquence remarquable du Théorème 6, ou plutôt de sa preuve :

COROLLAIRE 2. Tout groupe de Kazhdan d'ordre fini est quotient d'un groupe de Kazhdan de présentation finie.

Preuve : Soit Γ un groupe de Kazhdan d'ordre fini. Comme Γ est de type fini, on peut l'écrire comme quotient d'un groupe libre d'un certain rang m fini :

$$\Gamma = \mathbf{F}_m / N.$$

Soit r_1, r_2, \dots une numération des éléments du sous-groupe normal N . Notons N_n le sous-groupe normal de \mathbf{F}_m engendré par $\{r_1, \dots, r_n\}$, et posons $\Gamma_n = \mathbf{F}_m / N_n$. On va montrer qu'il existe n tel que Γ_n n'a pas la propriété (T).

Si ce n'était pas le cas, on trouverait pour tout n une représentation unitaire π_n de Γ_n , possédant presque des vecteurs invariants, mais sans vecteur invariant non nul. On voit π_n comme une représentation du groupe libre \mathbf{F}_m . Soit Q une base de \mathbf{F}_m . Le lemme 2 fournit, pour tout n , un vecteur v_n dans l'espace de Hilbert de π_n , tel que $\delta(v_n) = 1$ et $\delta(u) \geq \frac{1}{2}$ pour tout u dans la boule centrée en v_n et de rayon n . Pour $g \in \mathbf{F}_m$, posons $\psi_n(g) = \|\pi_n(g)v_n - v_n\|^2$. Comme dans la preuve du Théorème 6, on peut extraire de la suite $(\psi_n)_{n \geq 1}$ une sous-suite qui converge ponctuellement sur \mathbf{F}_m vers une fonction conditionnellement de type négatif ψ , telle que l'action isométrique associée à ψ n'a pas presque des points fixes, et en particulier n'a pas de point fixe. D'autre part, la fonction ψ_n s'annule sur N_n , donc ψ s'annule sur N (qui est la réunion croissante des N_n). Ceci veut dire que l'action isométrique associée à ψ factorise par Γ , ce qui contredit la propriété (T) pour Γ . \triangleright

Exemple 7 : Notons $\mathbf{F}_p[X]$ (resp. $\mathbf{F}_p((X))$) l'anneau des polynômes (resp. le corps des séries formelles) coefficients dans le corps fini \mathbf{F}_p d'ordre p premier. Le groupe $\Gamma = SL_3(\mathbf{F}_p[X])$ est isomorphe à $\Gamma = SL_3(\mathbf{F}_p[\frac{1}{X}])$, lequel est un réseau dans $SL_3(\mathbf{F}_p((X)))$. Par les Théorèmes 1 et 2, le groupe Γ a la propriété (T). D'autre part, Behr [4] a montré que Γ n'est pas de présentation finie⁴. Un problème ouvert consiste à écrire explicitement

⁴Notons que, pour $n \geq 4$, le groupe $SL_n(\mathbf{F}_p[X])$ est de présentation finie ! C'est un résultat de Rehmann et Soul [44].

un groupe de Kazhdan de présentation finie qui se surjecte sur Γ . Le groupe $SL_3(\mathbb{Z}[X])$ se qualifie-t-il pour cela (à ma connaissance, on ne sait pas si $SL_3(\mathbb{Z}[X])$ est de présentation finie, ni s'il a la propriété (T)) ?

3.2. Propriété (T) et applications harmoniques

La possibilité d'exploiter la théorie des applications harmoniques (et spécialement la super-rigidité géométrique) pour établir la propriété (T) a déjà été discutée dans le Séminaire Bourbaki 778 de P. Pansu [41]. Nous allons commencer par exposer un résultat non publié de Y. Shalom, qui lie la propriété (T) d'un groupe G aux applications harmoniques valeurs dans un espace de Hilbert, G -quivariantes par rapport à une action admissible isométrique de G .

DEFINITION 5. Soit G un groupe localement compact, et K un sous-groupe compact.

a) La paire (G, K) est une paire de Gelfand si l'algèbre de convolution $C_c(G//K)$ des fonctions continues, K -bi-invariantes, à support compact, est commutative.

b) Si (G, K) est une paire de Gelfand, une représentation unitaire irréductible π de G est sphérique si elle admet des vecteurs K -fixes non nuls⁵.

Pour les représentations sphériques, Delorme [17] a obtenu un résultat d'annulation de la 1-cohomologie :

THORME 7. Soit (G, K) une paire de Gelfand. Si π est une représentation sphérique non triviale de G , alors $H^1(G, \pi) = 0$. ▷

Voici alors le résultat de Shalom.

THORME 8. Soit (G, K) une paire de Gelfand, où G est un groupe de Lie connexe avec $G/[G, G]$ compact. Si G n'a pas la propriété (T), il existe une application F harmonique, non constante, G -quivariante de G/K vers un espace de Hilbert \mathcal{H} muni d'une action isométrique admissible de G .

Preuve : Commençons par construire l'application F . Comme G n'a pas la propriété (T), par le Théorème 6 il existe une représentation unitaire irréductible π de G , sur un espace de Hilbert \mathcal{H} , telle que $H^1(G, \pi) \neq 0$. Comme $G/[G, G]$ est compact, π n'est pas la représentation triviale de dimension 1. Soit $b \in Z^1(G, \pi)$ un cocycle qui n'est pas un cobord ; notons α l'action isométrique admissible de G sur \mathcal{H} associée à b . Comme K a la propriété (T) (ou simplement en moyennant sur K par rapport à la mesure de Haar dk normalisée) on trouve un point $v_0 \in \mathcal{H}$ fixe par $\alpha(K)$. On peut alors considérer l'application orbitale $\tilde{F} : G \rightarrow \mathcal{H} : g \mapsto \alpha(g)(v_0)$, qui factorise en une application

⁵Dans ce cas, l'espace des vecteurs K -fixes est de dimension 1.

$F : G/K \rightarrow \mathcal{H} : gK \mapsto \alpha(g)(v_0)$ non constante, G -quivariante. Nous allons montrer en trois pas que F est harmonique.

- 1er pas : Le point $v = v_0$ est l'unique point de \mathcal{H} fixe par $\alpha(K)$. En effet, s'il existait un deuxième point $v = v_1$ fixe par $\alpha(K)$, le vecteur non nul $v = v_0 - v_1$ serait fixe par la partie linéaire $\pi(K)$, et π serait sphérique, ce qui contredirait le Théorème 7.

- 2me pas : Pour tous $g, h \in G$, on a $\int_K \tilde{F}(gkh) dk = \tilde{F}(g)$. En effet, pour tout $v \in \mathcal{H}$, on a, par le premier pas : $\int_K \alpha(k)v dk = v_0$ (car le membre de gauche est fixe par $\alpha(K)$). Donc :

$$\int_K \tilde{F}(gkh) dk = \alpha(g) \int_K \alpha(k)\alpha(h)v_0 dk = \alpha(g)v_0 = \tilde{F}(g).$$

Pour $x_0 = g_0K$ et x des points de G/K , on peut écrire :

$$(7) \quad \int_K F(g_0kg_0^{-1}x) dk = F(x_0),$$

qui veut dire que la moyenne de F sur toute orbite du générateur de $x = x_0$ est égale à la valeur de F en x_0 (on est donc très proche de la caractérisation des fonctions harmoniques par la propriété de la moyenne).

- 3me pas : Pour $u \in \mathcal{H}$, posons $\tilde{F}_u(g) = \langle \tilde{F}(g)|u \rangle$ et $F_u(x) = \langle F(x)|u \rangle$. Montrons que F est harmonique au sens fort, c'est-à-dire que $D(F_u) = 0$ pour tout $u \in \mathcal{H}$ et tout opérateur D sur $C^\infty(G/K)$, G -invariant, nul sur les constantes. Reprenons l'équation (7) : pour tout $g_0 \in G$, la fonction $x \mapsto \int_K F_u(g_0kg_0^{-1}x) dk$ est constante, donc par G -invariance de D :

$$0 = D\left(\int_K F_u(g_0kg_0^{-1}x) dk\right) = \int_K (DF_u)(g_0kg_0^{-1}x) dk;$$

pour $x = x_0$, on en tire $(DF_u)(x_0) = 0$, donc DF_u est identiquement nulle. ▷

Pour faire le lien entre propriété (T) et super-rigidité géométrique, nous avons un résultat de J. Jost [34].

THÉORÈME 9. Soit X un espace riemannien symétrique irréductible du type non compact, distinct de l'espace hyperbolique réel $\mathbb{H}^n(\mathbb{R})$ ou complexe $\mathbb{H}^n(\mathbb{C})$; soit Γ un sous-groupe co-compact du groupe des isométries de X . Soit N une variété riemannienne complète simplement connexe, courbure sectionnelle négative (≤ 0), éventuellement de dimension infinie, sur laquelle Γ agit par isométries. Soit $F : X \rightarrow N$ une application harmonique Γ -quivariante. Ou bien F est constante, ou bien F est (après multiplication de la matrice de N par un scalaire positif) un plongement isométrique totalement géodésique.

L'exemple le plus simple de variété riemannienne complète, simplement connexe, courbure sectionnelle négative, qui soit de dimension infinie, est évidemment un espace de Hilbert. Dans ce cas, l'application F du Théorème 9 ne peut pas être un plongement isométrique totalement géodésique. Donc F est constante. Nous voyons que le Théorème 9,

combin au Thorme 8, fournit une preuve (plutt coteuse) de la propriit (T) pour les groupes de Lie simples connexes de centre \emptyset ni, non localement isomorphes $SO(n, 1)$ ou $SU(n, 1)$.

Terminons cette section en expliquant une preuve rcente, due M. Gromov ([26], 3.7.D'), de la propriit (T) pour les groupes $Sp(n, 1)$ et $F_4(-20)$, de rang rel 1 (voir Thorme 3(ii)).

Soit X un espace riemannien symtrique irrductible de rang 1 de type non compact : X est donc un espace hyperbolique $\mathbb{H}^n(\mathbb{K})$ de dimension n sur \mathbb{K} , avec $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$ ou Cay (et $n = 2$ dans ce cas).

Fixons quelques notations. G est la composante connexe du neutre du groupe des isomtries de X ; K est un sous-groupe compact maximal de G , et $x_0 \in X$ un point dont le \emptyset xateur dans G est exactement K . On pose encore $d = \dim_{\mathbb{K}}(\mathbb{K})$, $m_1 = d(n-1)$, $m_2 = d-1$ (m_1, m_2 sont les paramtres du groupe nilpotent qui apparat dans la dcomposition d'Iwasawa de G).

Notons r la distance godsiq de x x_0 dans X . Une fonction f valeurs relles sur X est radiale si elle ne dpend que de r , c'est--dire s'il existe une autre fonction ϕ telle que $f(x) = \phi(r)$. Normalisons la mtrique de sorte que l'inclusion $\mathbb{H}^{n-1}(\mathbb{K}) \hookrightarrow \mathbb{H}^n(\mathbb{K})$ soit un plongement isomtrique totalement godsiq. Le laplacien sur une fonction radiale $f(x) = \phi(r)$ est alors donn par l'expression

$$(8) \quad \Delta f = -\phi'' - m(r)\phi',$$

o $m(r) = m_1 \coth r + 2m_2 \coth 2r$ (voir [18]).

J'ai eu besoin du lemme suivant pour comprendre la page 37 de [26].

LEMME 3.- Soit F une application G -quivariante de X vers un espace de Hilbert \mathcal{H} muni d'une action isomtrique aEne de G , telle que $F(x_0) = 0$. Alors :

i) Si F est non constante, il existe une constante $\lambda > 0$ telle que λF est localement isomtrique, c'est--dire $\lambda \|dF_{x_0}(Y)\| = \|Y\|$ pour tous $x \in X, Y \in T_x X$.

ii) Pour tout $x \in X$, on a $Re\langle \Delta F(x) | F(x) \rangle \geq 0$.

Preuve du point (i) : Notons π la partie linaira de l'action aEne de G sur \mathcal{H} . La dierentielle dF_{x_0} est un oprateur d'entrelacement entre la representation d'isotropie de K sur $T_{x_0} X$ et la restriction de π K . Les deux formes quadratiques sur $T_{x_0} X$:

$$Y \mapsto \|Y\|^2; Y \mapsto \|dF_{x_0}(Y)\|^2$$

sont K -invariantes. Comme la representation d'isotropie est irrductible, le lemme de Schur assure que ces deux formes quadratiques sont proportionnelles : il existe $\lambda > 0$ tel que $\lambda \|dF_{x_0}(Y)\| = \|Y\|$ pour tout $Y \in T_{x_0} X$. La transitivit de l'action de G sur X , et la G -quivariance de F , permettent de conclure. ▷

La preuve du point (ii) est plus calculatoire, et rutilise le Thorme 7 ; pour les dtails, voir [5], 3.6.18.

L'ide de Gromov pour tablir la propriit (T) pour $Sp(n, 1)$ et F (4(-20) est d'tudier la croissance des applications G -quivariantes $X \rightarrow \mathcal{H}$ et de montrer que, parmi celles-ci, les applications harmoniques ont la croissance la plus rapide. La proposition suivante rsume les ides de [26], 3.7.D '.

PROPOSITION 5.- Soit $F : X \rightarrow \mathcal{H}$ une application G -quivariante, localement isomtrique, avec $F(x_0) = 0$. Soit ϕ la fonction sur \mathbf{R}^+ d'one par $\phi(r) = \|F(x)\|^2$. Alors, pour $r \rightarrow \infty$:

$$\phi(r) \leq \frac{2 \dim X}{m_1 + 2m_2} r + o(r),$$

avec galit si F est harmonique.

Preuve : Si M est une varit riemannienne et $h : M \rightarrow \mathcal{H}$ est une application vers un espace de Hilbert, un calcul facile montre que, pour $x \in M$:

$$(9) \quad \Delta \|h\|^2(x) = 2Re\langle \Delta h(x) | h(x) \rangle - 2 \|dh_x\|_{HS}^2,$$

o $\|dh_x\|_{HS} = \sqrt{Tr((dh_x)^* dh_x)}$ est la norme de Hilbert-Schmidt de la dierentielle dh_x . En particulier, si h est localement isomtrique, on a $\|dh_x\|_{HS}^2 = \dim M$.

Appliquons ceci avec $M = X$, $h = F$. En combinant les quations (8) et (9), on obtient :

$$\phi'' + m(r)\phi' = -2Re\langle \Delta F(x) | F(x) \rangle + 2 \dim X;$$

c'est une quation dierentielle ordinaire du second ordre en ϕ , qu'on peut rsoudre par la mthode de variation des constantes. Comme $Re\langle \Delta F(x) | F(x) \rangle \geq 0$ (voir le lemme 3 (ii)), la solution ϕ sera domine par la solution de l'quation obtenue en annulant le terme $-2Re\langle \Delta F(x) | F(x) \rangle$. Un calcul direct donne alors le rsultat. ▷

La constante $\frac{2 \dim X}{m_1 + 2m_2}$ se calcule aisement dans les dierents cas :

X	$\frac{2 \dim X}{m_1 + 2m_2}$
$\mathbf{H}^n(\mathbf{R})$	$2 + \frac{2}{n-1}$
$\mathbf{H}^n(\mathbf{C})$	2
$\mathbf{H}^n(\mathbf{H})$	$2 - \frac{2}{2n+1}$
$\mathbf{H}^1(\text{Cay})$	$\frac{8}{7}$
$\mathbf{H}^2(\text{Cay})$	$\frac{16}{11}$

On voit ici la principale différence entre les quatre familles d'espaces hyperboliques. Pour $\mathbb{H}^n(\mathbb{R})$, la constante $\frac{2 \dim X}{m_1 + 2m_2}$ décroît avec la dimension ; pour $\mathbb{H}^n(\mathbb{C})$ elle ne dépend pas de la dimension ; pour $\mathbb{H}^n(\mathbb{H})$ ou $\mathbb{H}^n(\text{Cay})$, elle augmente avec la dimension.

Preuve du Théorème 3 (ii) : Nous voulons montrer que, pour $n \geq 2$, le groupe $G = Sp(n, 1)$ a la propriété (T). Supposons par l'absurde qu'il ne l'a pas. Par le Théorème 8, il existe une application G -quivariante, harmonique, non constante, F de $X = \mathbb{H}^n(\mathbb{H})$ vers un espace de Hilbert. Quitte à conjuguer l'action naturelle de G par une translation, nous pouvons supposer $F(x_0) = 0$; par le lemme 3 (ii), nous pouvons supposer que F est localement isométrique.

Soit $\phi(r) = \|F(x)\|^2$. L'idée est de restreindre F à une droite hyperbolique quaternionienne $L = \mathbb{H}^1(\mathbb{H})$ passant par x_0 ; le groupe d'isométries de L est $G' = Sp(1, 1)$. La restriction $F|_L$ est localement isométrique, G' -quivariante (mais pas nécessairement harmonique). Comme L est totalement géodésique dans X , on a $\|F|_L(x)\|^2 = \phi(r)$ pour $x \in L$. Par deux applications de la Proposition 5, on a, pour $r \rightarrow \infty$:

$$\phi(r) \leq \frac{4}{3}r + o(r) < (2 - \frac{2}{2n+1})r + o(r) = \phi(r),$$

ce qui est une contradiction. La preuve pour $F_{4(-20)}$ est tout fait semblable. ▷

3.3. Extensions centrales

Nous allons exposer ici la preuve de Shalom d'un résultat de Serre (voir [15], Théorème 12 du Chapitre 2), qui permet de relever la propriété (T) des extensions centrales. Le lemme suivant est facile.

LEMME 4. - Soit Z un sous-groupe fermé central du groupe localement compact G . Soit π une représentation unitaire irréductible de G . Si la restriction de $\pi|_Z$ est non triviale, alors $H^1(G, \pi) = 0$.

Preuve : Par le lemme de Schur, il existe un caractère $\chi : Z \rightarrow S^1$ tel que $\pi(z) = \chi(z)Id$ pour tout $z \in Z$. Soit $z_0 \in Z$ tel que $\chi(z_0) \neq 1$. Soit $b \in Z^1(G, \pi)$; montrons que b est un cobord. Pour $g \in G$, la relation de cocycle donne :

$$\pi(g)b(z_0) + b(g) = b(gz_0) = b(z_0g) = \chi(z_0)b(g) + b(z_0)$$

ou encore :

$$b(g) = \pi(g) \left(\frac{b(z_0)}{\chi(z_0) - 1} \right) - \frac{b(z_0)}{\chi(z_0) - 1},$$

donc $b \in B^1(G, \pi)$. ▷

THÉORÈME 10. - Soit G un groupe localement compact, compactement engendré, séparé. On suppose de plus que l'abélianisé par $G/[G, G]$ est compact. Soit Z un sous-groupe fermé central de G . Si G/Z a la propriété (T), alors G l'a également.

Preuve : Par le Thorme 6, il suÉt de montrer que, si π est une representation unitaire irreductible de G , alors $H^1(G, \pi) = 0$. Si π est non triviale sur Z , cela rsulte du lemme 4. Supposons donc π triviale sur Z , et voyons π comme une representation de G/Z . Soit $b \in Z^1(G, \pi)$. Si la restriction de b Z est identiquement nulle, alors b factorise par G/Z , et la propriÉt (T) pour G/Z assure que b est un cobord.

Il reste montrer que b doit s'annuler sur Z . Si ce n'tait pas le cas, on trouverait $z_0 \in Z$ tel que $b(z_0) \neq 0$ et donc, pour tout $g \in G$:

$$\pi(g)b(z_0) + b(g) = b(gz_0) = b(z_0g) = b(g) + b(z_0),$$

ou encore

$$\pi(g)b(z_0) = b(z_0).$$

La representation π admet donc un vecteur øxe non nul. Comme elle est irreductible, cela implique que π est la representation triviale en dimension 1, et que b est un homomorphisme continu de G vers \mathbb{C} . Comme $G/[G, G]$ est compact, b doit tre identiquement nul, ce qui contredit notre hypothse $b(z_0) \neq 0$. ▷

Du Thorme 10, on dduit immdiatement :

COROLLAIRE 3. Soit G un groupe de Lie semi-simple connexe, et Γ un rseau dans G . Notons \tilde{G} le revtement universel de G , et $\tilde{\Gamma}$ l'image inverse de Γ dans G .

- i) G a la propriÉt (T) si et seulement si \tilde{G} l'a ;
- ii) Γ a la propriÉt (T) si et seulement si $\tilde{\Gamma}$ l'a. ▷

Ce Corollaire montre que, pour les groupes de Lie semi-simples, la propriÉt (T) est un invariant d'isomorphisme local ⁶.

Le Corollaire 3 est spcialement intressant pour un groupe de Lie G simple, groupe fondamental cyclique inøni ⁷ : si le rang rel de G est ≥ 2 , alors \tilde{G} est un exemple de groupe de Kazhdan centre inøni. C'est le cas pour $G = Sp_{2n}(\mathbb{R})$ ($n \geq 2$), $G = SU(p, q)$ ($2 \leq p \leq q$), $G = SO(2, m)$ ($m \geq 3$), ...

Exploitions cette dernire remarque pour donner une rponse - ngative - une question pose dans [15] : pour les groupes de type øni, la propriÉt (T) n'est pas un invariant de quasi-isomtrie. Il est diÉcile d'attribuer ce rsultat avec prcision : au printemps 2000, plusieurs exemples ont surgi un peu partout... Je dirai simplement que la construction suivante m'a t expliquée par C. Pittet.

PROPOSITION 6. Soit Γ un rseau co-compact dans un groupe de Lie G simple, de

⁶Ceci est *faux* pour les groupes de Lie quelconques : penser au cercle et la droite relle !

⁷On sait que ceci se produit exactement quand l'espace symtrique associ G est un domaine hermitien symtrique irreductible.

rang réel au moins 2, groupe fondamental cyclique infini. Soit \tilde{G} le revêtement universel de G , et $\tilde{\Gamma}$ l'image inverse de Γ dans \tilde{G} . Le groupe $\tilde{\Gamma}$, qui a la propriété (T), est quasi-isométrique au groupe $\Gamma \times \mathbb{Z}$, qui n'a pas la propriété (T).

Preuve : Par le Corollaire 3, le groupe $\tilde{\Gamma}$ a la propriété (T), et bien sûr le groupe $\Gamma \times \mathbb{Z}$ ne l'a pas.

Soit $G = ANK$ une décomposition d'Iwasawa de G , avec K un sous-groupe compact maximal. Munissons G/K de sa structure riemannienne d'espace riemannien symétrique, et G, K de structures riemanniennes invariantes à gauche; relevons ces dernières aux revêtements universels \tilde{G} et \tilde{K} . Le groupe $\tilde{\Gamma}$ agit librement, isométriquement, avec quotient compact, sur \tilde{G} : les espaces métriques $\tilde{\Gamma}$ et \tilde{G} sont donc quasi-isométriques. De même, $\Gamma \times \mathbb{Z}$ agit proprement isométriquement, avec quotient compact, sur $G/K \times \tilde{K}$: les espaces métriques $\Gamma \times \mathbb{Z}$ et $G/K \times \tilde{K}$ sont donc quasi-isométriques.

Il suffit alors de montrer que \tilde{G} et $G/K \times \tilde{K}$ sont quasi-isométriques. Et pour ce faire, il suffit de s'assurer que G et $G/K \times K$ sont bi-Lipschitz-équivalents.

Pour voir ceci, identifions $AN \backslash G/K$ de façon AN -quivariante, et considérons l'application

$$\psi : G \rightarrow AN \times K : ank \mapsto (an, k);$$

c'est un difféomorphisme AN -quivariant. Fixons $x \in G$; comme deux normes sur \mathbb{R}^n sont bi-Lipschitz-équivalentes, il existe une constante $C_x > 0$ telle que, pour tout vecteur Y tangent à G en x :

$$\frac{1}{C_x} \|Y\| \leq \|d\psi_x(Y)\| \leq C_x \|Y\|.$$

Par compacité de K , on trouve $C > 0$ tel que, pour tout $k \in K$ et tout vecteur Y tangent à G en xk :

$$\frac{1}{C} \|Y\| \leq \|d\psi_{xk}(Y)\| \leq C \|Y\|.$$

Enfin, comme ψ est AN -quivariante, cette constante C est indépendante de x , ce qui montre que ψ est bi-lipschitzien. ▷

Par contraste avec la Proposition 6, l'annulation de la 1-cohomologie à coefficients dans la représentation régulière à gauche est un invariant de quasi-isométrie pour les groupes de type fini, voir [6].

4. LE CRITÈRE SPECTRAL

4.1. L'approche

Soit X une variété compacte, ou un CW-complexe fini. Notons \tilde{X} le revêtement universel de X , et $\Gamma = \pi_1(X)$ le groupe fondamental de X .

Dans [41], Pansu propose le programme suivant pour démontrer que Γ a la propriété (T). Soit π une représentation unitaire de Γ , sur un espace de Hilbert \mathcal{H} : on veut montrer

que $H^1(\Gamma, \pi) = 0$. Pour cela, on considère le produit $\tilde{X} \times \mathcal{H}_\pi$, muni de l'action diagonale de Γ . Le quotient $E_\pi = \tilde{X} \times_\Gamma \mathcal{H}_\pi$ est un \mathcal{O}_X -module plat, de module \mathcal{H}_π au-dessus de X . La cohomologie des groupes assure alors que :

$$H^1(\Gamma, \pi) = H^1(X, E_\pi);$$

le membre de droite se comprenant comme la 1-cohomologie des formes différentielles à valeurs dans E_π si X est une variété, et comme la 1-cohomologie cellulaire à coefficients dans E_π si X est un CW-complexe.

Si X est une variété, on peut faire appel à la théorie de Hodge (ceci n'est pas trivial, puisque les modules de E_π sont en général de dimension infinie! - voir Mok [39], Corollaire 0.1 pour les détails) : Γ a la propriété (T) si et seulement si, pour toute représentation unitaire π de Γ , il n'y a pas de 1-forme harmonique non nulle sur X à valeurs dans E_π . Si X est un espace riemannien symétrique irréductible du type non compact, distinct de l'espace hyperbolique réel $\mathbb{H}^n(\mathbb{R})$ ou complexe $\mathbb{H}^n(\mathbb{C})$, la formule de Bochner-Weitzenböck montre que le laplacien sur les 1-formes à valeurs dans E_π est borné inférieurement par une constante strictement positive, donc a un noyau trivial : c'est ainsi que Mok ([39], Corollaire 0.2) démontre le Théorème 2 (pour les groupes de Lie simples réels) et le Théorème 3(ii). Il est à noter que la borne inférieure sur le laplacien est indépendante de la représentation π , et ne dépend que du tenseur de courbure de X .

Dans le cas où \tilde{X} est l'immeuble de Bruhat-Tits associé à un groupe algébrique simple \mathbf{G} sur un corps local non archimédien \mathbb{K} , Garland [22] avait dès 1973 fait marcher le programme esquissé ci-dessus : si Γ est un réseau co-compact dans $\mathbf{G}(\mathbb{K})$, alors $H^1(\Gamma, \pi) = 0$ pour toute représentation unitaire π de dimension infinie de Γ : l'hypothèse de finitude de la dimension s'explique par le recours à la théorie de Hodge (voir aussi l'exposé de Borel [7]). P. Pansu [42], dans le cas des immeubles de type \tilde{A}_2 , et W. Ballmann et J. Świątkowski [2] ainsi que A. Zuk [56] indépendamment, dans le cas plus général d'un complexe simplicial simplement connexe de dimension 2, ont pu limiter l'usage de la théorie de Hodge. La condition locale de courbure se lit alors sur les liens de sommets, qui sont des graphes, plus particulièrement sur les valeurs propres du laplacien combinatoire sur les liens.

DEFINITION 6. Soit $L = (V, E)$ un graphe fini. Le laplacien combinatoire de L est l'opérateur auto-adjoint $\Delta_L : \ell^2(V) \rightarrow \ell^2(V)$ défini par :

$$(\Delta_L f)(x) = f(x) - \frac{1}{\deg x} \sum_{y \text{ voisin de } x} f(y)$$

($f \in \ell^2(V), x \in V$). Les valeurs propres de L sont celles de Δ_L :

$$0 = \lambda_0(L) \leq \lambda_1(L) \leq \dots \leq 2$$

(avec $0 < \lambda_1(L)$ si et seulement si L est connexe).

Le resultat de Ballman-Swiatkowski [2] et Zuk [56] s'nonce :

THORME 11. Soit \tilde{X} un complexe simplicial localement fini, simplement connexe, de dimension 2, tel que pour chaque sommet $v \in \tilde{X}$ le link L_v satisfait $\lambda_1(L_v) > \frac{1}{2}$. Si Γ est un groupe agissant proprement et avec quotient compact sur \tilde{X} , alors Γ a la propriete (T). ▷

Exemple 8 : 1) Soit \tilde{X} le pavage du plan euclidien par des triangles quilatraux. Les links de sommets sont des cycles de longueur 6, pour lesquels on a $\lambda_1 = \frac{1}{2}$. D'autre part, le groupe des translations de \tilde{X} est isomorphe \mathbf{Z}^2 , donc n'a certainement pas la propriete (T). Cet exemple montre que l'hypothese spectrale dans le Thorme 11 ne peut tre amliore.

2) Un groupe Γ s'appelle groupe \tilde{A}_2 s'il admet une action simplement transitive, permutant cycliquement le type des sommets, sur les sommets d'un immeuble pais de type \tilde{A}_2 . Ces groupes ont ete etudies et caractrisees par Cartwright-Mantero-Steger-Zappa [10] : la plupart peuvent tre realises comme reseaux co-compacts dans PGL_3 d'un corps local non archimkien K , mais pas tous : l'existence d'immeubles de type \tilde{A}_2 non classiques permet de construire des groupes \tilde{A}_2 qui ne sont pas des reseaux dans $PGL_3(K)$. Cartwright-Mlotkowski-Steger [11] ont montr, par un remarquable calcul direct, que tout groupe \tilde{A}_2 a la propriete (T) : pour la premiere fois, la propriete (T) a ete tablie pour un groupe denombrable sans faire appel la thorie des reseaux dans les groupes algebriques simples sur les corps locaux ⁸. Le Thorme 11 fournit une preuve simple du resultat de Cartwright-Mlotkowski-Steger : en effet un link de sommet d'un immeuble de type \tilde{A}_2 est le graphe d'incidence points-droites d'un plan projectif sur un corps fini : le calcul du spectre de ce graphe (voir Feit-Higman [20]) montre que le critere spectral est satisfait.

3) Dans [3], S. Barr construit un immeuble pais de type \tilde{A}_2 , qui admet un groupe d'automorphismes Γ quotient compact, mais pas de groupe d'automorphismes transitif sur les sommets. Un tel immeuble est donc non classique. Le calcul de [10] ne s'applique pas au groupe Γ ; par contre, le Thorme 11 s'y applique, et montre que Γ est un exemple geometrique completement neuf de groupe ayant la propriete (T).

4.2. La propriete (T) sur ordinateur ?

Le Thorme 11 a ouvert une possibilite qui paraissait inimaginable il y a 10 ans : celle de verifier la propriete (T) sur ordinateur ! En effet, si \tilde{X} est un complexe simplicial de dimension 2 donn explicitement, avec des links pas trop grands et en petit nombre (isomorphisme pres), on peut envisager de verifier le critere spectral en faisant calculer par la machine la plus petite valeur propre non nulle du laplacien combinatoire des links de \tilde{X} .

⁸Comparer avec la preuve de Shalom [49] de la propriete (T) pour $SL_n(\mathbf{Z})$ ($n \geq 3$), qui n'utilise pas le fait que $SL_n(\mathbf{Z})$ est un reseau dans $SL_n(\mathbf{R})$.

D'autre part, tout groupe de présentation onie admet une action propre, quotient compact, sur un complexe simplicial simplement connexe de dimension 2 (savoir le revêtement universel de la subdivision barycentrique du 2-complexe de présentation). Le Thorme 11 suggre que la proprié (T) pourrait se lire sur une présentation du groupe. . . Ballmann et Swiatkowski ([2], Corollaire 2) ont donn les premiers exemples de présentations de ce type :

THORME 12. Soient H un groupe onie, $S \subset H - \{1\}$ une partie gnratrice de H , et $H = \langle S | R \rangle$ une présentation de H . On suppose que le graphe de Cayley $L = \mathcal{G}(H, S)$ a un tour de taille au moins gal 6, et satisfait $\lambda_1(L) > \frac{1}{2}$. Alors le groupe Γ donn par la présentation

$$\Gamma = \langle S \cup \{\tau\} | R \cup \{\tau^2\} \cup \{(s\tau)^3 : s \in S\} \rangle$$

est un groupe inonnie possdant la proprié (T). \triangleright

En fait, sous la seule hypothse de tour de taille ≥ 6 , le groupe Γ donn par la présentation du Thorme 12 agit proprement, avec quotient onie, sur un complexe simplicial contractile de dimension 2, dont tous les links de sommets sont isomorphes $L = \mathcal{G}(H, S)$ ([2], Theorem 2). Ceci montre que Γ est inonnie. Le Thorme 12 dcoule alors du Thorme 11. Pour des exemples de graphes de Cayley de groupes onies satisfaisant les deux hypothses du Thorme 12, voir [37].

A. Zuk est parvenu abstraire les conditions du Thorme 11 pour donner une condition suEsante pour la proprié (T), qui ne dpend que d'une présentation d'un groupe (mieux : qui ne dpend que de certaines relations).

Soit Γ un groupe engendr par une partie S onie, symtrique ($S = S^{-1}$), avec $1 \notin S$. On dñit un graphe $L(S)$ dont l'ensemble des sommets est S , et dont l'ensemble des artes est $\{\{s, t\} : s, t, s^{-1}t \in S\}$. Si $\Gamma = \langle T | R \rangle$ est une présentation de Γ , et $S = T \cup T^{-1}$, on voit que $L(S)$ ne dpend que des relations de longueur 3. Zuk dmontre alors ([55], Theorem 1) :

THORME 13. Si le graphe $L(S)$ satisfait $\lambda_1(L(S)) > \frac{1}{2}$, alors Γ a la proprié (T). \triangleright

La condition $\lambda_1(L(S)) > \frac{1}{2}$ implique que $L(S)$ est connexe. Remarquons qu'il n'est pas restrictif de supposer $L(S)$ connexe : en coet, en remplaant S par la runion des mots de longueur 1 et de longueur 2 sur S , on obtient une partie gnratrice S' avec $L(S')$ connexe. Notons encore que, malgr certains reliquats cohomologiques, la preuve du Thorme 13 ne repose plus sur le thorme de Delorme-Guichardet, mais sur la dñition originale de la proprié (T).

4.3. Applications

La condition suffisante du Thorme 13 s'avre particulirement maniable. Par exemple, il rsulte du Corollaire 2 (et de sa preuve) que, si $\Gamma = \langle T | R \rangle$ est une prsentation d'un groupe de Kazhdan, il est possible d'extraire de l'ensemble R des relations un nombre fini r_1, \dots, r_n de relations, telles que $\tilde{\Gamma} = \langle T | r_1, \dots, r_n \rangle$ possde la propriert (T). D'o la question : peut-on expliciter les relations qu'on peut supprimer de R , tout en assurant que le groupe correspondant a la propriert (T) ? Zuk ([55], Theorem 5) donne une rponse en ce qui concerne les relations de longueur 3 :

THORME 14. - Soit Γ un groupe de type fini. Soit S une partie gnratrice finie, symtrique, avec $1 \notin S$. On suppose que $\lambda_1(L(S)) > \frac{1}{2}$. Soit $t \in \mathbb{N}$ tel que $t \leq \frac{1}{11}(\lambda_1(L(S)) - \frac{1}{2}) \min\{\deg(s) : s \in L(S)\}$. En supprimant t relations de longueur 3 de toute prsentation de Γ sur la partie gnratrice S , on obtient un groupe qui possde la propriert (T). \triangleright

Zuk en dduit aussi la gnricit des groupes de Kazhdan parmi les groupes donns par une prsentation relations de longueur 3. Plus prcisment, on considre les prsentations P sur m gnrateurs s_1, \dots, s_m , $2mv$ relations de la forme

$$(s_j^{\pm 1} \sigma_1^i(s_j^{\pm 1}) \sigma_2^i(s_j^{\pm 1}))_{1 \leq i \leq v; 1 \leq j \leq m}$$

o $\sigma_1^1, \sigma_2^1, \dots, \sigma_1^v, \sigma_2^v$ sont $2v$ permutations de l'ensemble $\{s_1, \dots, s_m, s_1^{-1}, \dots, s_m^{-1}\}$ $2m$ lments. On note $\Gamma(P)$ le groupe d'ni par cette prsentation, et $\mathcal{P}(m, v)$ l'ensemble des prsentations de ce type (deux prsentations tant dierentes si les permutations sont dierentes). On a alors ([55], Theorem 12) :

THORME 15. - Pour v assez grand :

$$\lim_{m \rightarrow \infty} \frac{\text{card}\{P \in \mathcal{P}(m, v) : \Gamma(P) \text{ est de Kazhdan}\}}{\text{card } \mathcal{P}(m, v)} = 1.$$

\triangleright

Remerciements : Merci Yehuda Shalom pour d'utiles conversations, et mes co-auteurs Bachir Bekka et Pierre de la Harpe pour leur relecture attentive d'une premire version de ce texte. La prsentation suivie ici doit beaucoup [5].

REFERENCES

[1] R. ALPERIN - Locally compact groups acting on trees and property (T), Monatsh. Math. 93 (1982), 261-265.

- [2] W. BALLMANN and J. SWIATKOWSKI \checkmark On L^2 -cohomology and property (T) for automorphism groups of polyhedral cell complexes, *Geom. funct. anal.* 7 (1997), 615-645.
- [3] S. BARR \checkmark Immeubles de Tits triangulaires exotiques, *Ann. Fac. Sci. Toulouse* 9 (2000), 575-603.
- [4] H. BEHR \checkmark $SL_3(F_q[t])$ is not finitely presentable, *Homological group theory*, Proc. Symp., Durham 1977, Lond. Math. Soc. Lect. Notes Ser. 36 (1979), 213-224.
- [5] B. BEKKA, P. de la HARPE and A. VALETTE \checkmark Kazhdan's property (T), Preprint, Juillet 2002.
- [6] B. BEKKA and A. VALETTE \checkmark Group cohomology, harmonic functions and the first L^2 -Betti number, *Potential analysis* 6 (1997), 313-326.
- [7] A. BOREL \checkmark Cohomologie de certains groupes discrets et laplacien p -adique (d'après H. Garland), in *Sm. Bourbaki (1973/74)*, Exp. n $^\circ$ 437, Springer Lect. Notes in Math. 431 (1975), 12-35.
- [8] F. BRUHAT et J. TITS \checkmark Groupes réductifs sur un corps local I ; données radicielles values, *Publ. Math. IHS* 41 (1972), 5-251.
- [9] M. BURGER and N. MONOD \checkmark Bounded cohomology of lattices in higher rank Lie groups, *J. Eur. Math. Soc.* 1 (1999), 199-235.
- [10] D. CARTWRIGHT, A. MANTERO, T. STEGER and A. ZAPPA \checkmark Groups acting simply transitively on the vertices of a building of type \tilde{A}_2 , *Geometriae Dedicata* 47 (1993), 143-166.
- [11] D. CARTWRIGHT, W. MLOTKOWSKI and T. STEGER \checkmark Property (T) and \tilde{A}_2 groups, *Ann. Inst. Fourier* 44 (1993), 213-248.
- [12] P.-A. CHERIX, M. COWLING, P. JOLISSAINT, P. JULG and A. VALETTE \checkmark Groups with the Haagerup property (Gromov's a-T-menability), *Progress in Math.*, Birkhuser (2001).
- [13] K. CORLETTE \checkmark Archimedean superrigidity and hyperbolic rigidity, *Ann. of Math.* 135 (1992), 165-182.
- [14] M. COWLING \checkmark Sur les coefficients des représentations des groupes de Lie simples, in *Analyse harmonique sur les groupes de Lie II*, Springer Lect. Notes in Math. 739 (1979), 132-178.
- [15] P. de la HARPE et A. VALETTE \checkmark La propriété (T) de Kazhdan pour les groupes localement compacts, *Astrisque* 175, Soc. Math. de France, 1989.
- [16] C. DELAROCHE et A. KIRILLOV \checkmark Sur les relations entre l'espace dual d'un groupe et la structure de ses sous-groupes fermés (d'après D.A. Kazhdan), in *Sm. Bourbaki (1967/68)*, Exp. n $^\circ$ 343, réimpression Soc. Math. de France, Collection Hors Série, vol. 10 (1995), 507-528.

- [17] P. DELORME ~ 1-cohomologie des representations unitaires des groupes de Lie semi-simples et resolubles, Bull. Soc. Math. France 105 (1977), 281~336.
- [18] J. FARAUT ~ Analyse harmonique sur les paires de Guelfand et les espaces hyperboliques, in Analyse harmonique, Cours du CIMPA, Nice, 1983, 315~446.
- [19] D. FARLEY ~ Proper isometric actions of Thompson's groups on Hilbert space, Preprint, 2002.
- [20] W. FEIT and G. HIGMAN ~ The nonexistence of certain generalized polygons, J. Algebra 1 (1964), 114~131.
- [21] A. FURMAN ~ Gromov's measure equivalence and rigidity of higher rank lattices, Ann. of math. 150 (1999), 1059~1081.
- [22] H. GARLAND ~ p -adic curvature and the cohomology of discrete subgroups of p -adic groups, Ann. of math. 97 (1973), 375~423.
- [23] T. GELANDER and A. ZUK ~ Dependence of Kazhdan constants on generating subsets, Israel J. of Math. 129 (2002), 93~98.
- [24] . GHYS ~ Actions de reseaux sur le cercle, Invent. Math. 137 (1999), 199-231.
- [25] . GHYS and V. SERGIIESCU ~ Sur un groupe remarquable de diffeomorphismes du cercle, Comment. Math. Helv. 62 (1987), 185~239.
- [26] M. GROMOV ~ Random walk in random groups, Preprint IHS, Jan. 2002.
- [27] M. GROMOV ~ Hyperbolic groups, in Essays in group theory (S.M. Gersten, ed.), M.S.R.I. Publ. 8, Springer-Verlag, 75-263, 1987.
- [28] M. GROMOV and R. SCHOEN ~ Harmonic maps into singular spaces and p -adic superrigidity for lattices in groups of rank one, Inst. Hautes etudes Sci. Publ. Math. 76 (1992), 165~246.
- [29] A. GUICHARDET ~ Sur la cohomologie des groupes topologiques II, Bull. Sci. Math. 96 (1972), 305~332.
- [30] A. GUICHARDET ~ Cohomologie des groupes topologiques et des algebres de Lie, Cedric - F. Nathan, 1980.
- [31] G. HECTOR and U. HIRSCH ~ The geometry of foliations, Part B, Vieweg, 1983.
- [32] S. HELGASON ~ Differential geometry and symmetric spaces, Academic Press, 1963.
- [33] A. HULANICKI ~ Means and Flner condition on locally compact groups, Studia Math. 24 (1966), 87~104.
- [34] J. JOST ~ Equilibrium maps between metric spaces, Calc. Var. 2 (1994), 173~204.
- [35] D. KAZHDAN ~ Connection of the dual space of a group with the structure of its closed subgroups, Funct. Anal. and its Appl. 1 (1967), 63~65.

- [36] B. KOSTANT ~ On the existence and irreducibility of certain series of representations, Bull. Amer. Math. Soc. 75 (1969), 627~642.
- [37] A. LUBOTZKY ~ Discrete groups, expanding graphs and invariant measures, Progress in Math. 125, Birkhuser, 1994.
- [38] G.A. MARGULIS ~ Discrete subgroups of semisimple Lie groups, Springer-Verlag, Ergeb. Math. Grenzgeb. 3 Folge, Bd. 17, 1991.

- [39] N. MOK ~ Harmonic forms with values in locally constant Hilbert bundles, J. Fourier analysis and appl. (1995), 433~453. Kahane special issue.
- [40] A. NAVAS ~ Actions de groupes de Kazhdan sur le cercle, paratre aux Ann. Scient. c. Norm. Sup., 4me srie, t. 35 (2002).
- [41] P. PANSU ~ Sous-groupes discrets des groupes de Lie : rigidit, arithmticit, in Sm. Bourbaki (1993/94), Exp. n ° 778, Astrisque 227 (1995), 69-105.
- [42] P. PANSU ~ Formules de Matsushima, de Garland, et propriit (T) pour des groupes agissant sur des espaces symtriques ou des immeubles, Bull. Soc. math. France 126 (1998), 107~139.
- [43] A. PRESSLEY and G. SEGAL ~ Loop groups, Oxford Univ. Press, 1986.
- [44] U. REHMANN and C. SOUL ~ Finitely presented groups of matrices, in Algebraic K-theory, Springer Lect. Notes in Math. 551 (1976), 164~169.
- [45] H. REITER ~ Some properties of locally compact groups, Nederl. Akad. Wetensch. Indag. Math. 27 (1965), 697~701.
- [46] A. REZNIKOV ~ Analytic topology of groups, actions, strings and varieties, Preprint, Jan. 2000.
- [47] J.-P. SERRE ~ Arbres, amalgames, SL_2 , Astrisque 46, Soc. Math. France, 1977.
- [48] Y. SHALOM ~ Rigidity of commensurators and irreducible lattices, Invent. Math. 141 (2000), 1~54.
- [49] Y. SHALOM ~ Bounded generation and Kazhdan property (T), Publ. Math. IHS 90 (2001), 145~168.
- [50] G. SKANDALIS ~ Une notion de nuclearit en K -thorie, K-theory 1 (1988), 549-573.
- [51] L.N. VASERSTEIN ~ Groups having the property (T), Funct. Anal. and its Appl. 2 (1968), 174 p.
- [52] A.M. VERSHIK and S.I. KARPUSHEV ~ Cohomology of groups in unitary representations, the neighbourhood of the identity and conditionally positive deõnite functions, Math. USSR Sbornik 47 (1984), 513~526.
- [53] S.P. WANG ~ The dual space of semi-simple Lie groups, Amer. J. Math. 23 (1969), 921~937.
- [54] Y. WATATANI ~ Property (T) of Kazhdan implies property (FA) of Serre, Math. Japon. 27 (1981), 97~103.
- [55] A. ZUK ~ Property (T) and Kazhdan constants for discrete groups, Preprint, juin 2002.

- [56] A. ZUKOVICH La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres,
C.R. Acad. Sci. Paris 323 (1996), 453-458.

Alain VALETTE

Université de Neuchâtel

Institut de Mathématiques

Rue Emile-Argand 11

CH-2007 NEUCHÂTEL

(Suisse)

E-mail : alain.valette@unine.ch

**POINTS RATIONNELS ET GROUPES FONDAMENTAUX :
APPLICATIONS DE LA COHOMOLOGIE p -ADIQUE**
[d'après P. Berthelot, T. Ekedahl, H. Esnault, etc.]

par **Antoine Chambert-Loir**

Dans cet exposé, je présente trois résultats concernant les variétés algébriques en caractéristique positive :

a) *Deux variétés propres et lisses sur \mathbf{F}_q qui sont birationnelles ont même nombre de points rationnels modulo q (cf. T. Ekedahl, [25]).*

b) *Sur un corps fini, une variété propre et lisse qui est de Fano, ou bien géométriquement faiblement unirationnelle, ou plus généralement rationnellement connexe par chaînes, a un point rationnel (H. Esnault, [26]).*

c) *Sur un corps algébriquement clos de caractéristique $p > 0$, le groupe fondamental d'une variété propre et lisse qui est de Fano, ou bien géométriquement faiblement unirationnelle, ou plus généralement rationnellement connexe par chaînes, est un groupe fini d'ordre premier à p (cf. T. Ekedahl, [25]).*

Le point commun des démonstrations est un contrôle des valuations p -adiques des valeurs propres de Frobenius. Elles gagnent donc à être présentées dans le cadre d'une théorie cohomologique p -adique. La cohomologie rigide, développée par P. Berthelot, fournit l'outil idéal pour cela. Elle a connu récemment des progrès importants et je décris le formalisme auquel elle donne lieu. Les énoncés ci-dessus s'obtiennent en contrôlant les *pentés* des F -isocristaux que fournit la cohomologie rigide.

Je voudrais remercier P. Berthelot, J.-L. Colliot-Thélène, O. Debarre, P. Deligne, H. Esnault, B. Kahn, Y. Laszlo, D. Petrequin et J-P. Serre pour leurs conseils, suggestions ou corrections.

1. AUTOUR DU THÉORÈME DE CHEVALLEY-WARNING

Je commence cet exposé par un énoncé élémentaire et assez ancien, dû à C. Chevalley et E. Warning [68].

THÉORÈME 1.1. — Soit \mathbf{F} un corps fini, notons q son cardinal et p sa caractéristique. Soient F_1, \dots, F_r des polynômes en x_1, \dots, x_n , à coefficients dans \mathbf{F} et de degrés d_1, \dots, d_r . Si $n > d_1 + \dots + d_r$, le nombre de solutions dans \mathbf{F}^n du système

$$(1.2) \quad F_1(x_1, \dots, x_n) = \dots = F_r(x_1, \dots, x_n) = 0$$

est multiple de p .

La démonstration classique est très simple et repose sur le fait que pour $x \in \mathbf{F}$, x^{q-1} vaut 1 si $x \neq 0$ et 0 sinon. Ainsi, le nombre de solutions du système est congru modulo p à l'expression

$$\sum_{\mathbf{x} \in \mathbf{F}^n} \prod_{i=1}^r (1 - F_i(\mathbf{x})^{q-1}).$$

Le produit

$$\prod_{i=1}^r (1 - F_i(\mathbf{x})^{q-1})$$

est un polynôme de degré $\leq (q-1) \sum d_i$. Soit

$$c_{\mathbf{m}} x_1^{m_1} \dots x_n^{m_n}$$

un de ses monômes non nuls. On a donc $m_1 + \dots + m_n \leq (q-1) \sum d_i < n(q-1)$ si bien que nécessairement, l'un des entiers m_i , disons m_1 , est strictement inférieur à $q-1$ et

$$\sum_{\mathbf{x} \in \mathbf{F}^n} c_{\mathbf{m}} x_1^{m_1} \dots x_n^{m_n} = c_{\mathbf{m}} \prod_{i=1}^n \sum_{t \in \mathbf{F}} t^{m_i} = 0$$

puisque

$$\sum_{t \in \mathbf{F}} t^m = \begin{cases} -1 & \text{si } (q-1) \text{ divise } m \text{ et } m \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

Ce théorème a été généralisé par J. Ax [1] et N. Katz [39] :

THÉORÈME 1.3. — Si b désigne le plus petit entier tel que

$$b \geq \frac{n - \sum d_i}{\max d_i},$$

le nombre de solutions du système (1.2) est divisible par q^b .

Leurs démonstrations sont assez savantes. Celle qu'a proposée récemment D. Wan dans [67] lorsque $\mathbf{F} = \mathbf{F}_p$ est en revanche élémentaire et tout à fait dans l'esprit de celle du théorème 1.1. Commençons par introduire l'anneau \mathbf{Z}_p des entiers p -adiques. C'est un anneau de valuation discrète complet, de caractéristique 0, de corps résiduel \mathbf{F}_p ; son idéal maximal est engendré par p ; il contient les p racines de l'équation $x^p - x = 0$ (notons S cet ensemble). L'idée de base est maintenant que pour $x \in \mathbf{Z}_p$,

$x^{p^n(p-1)}$ est congru modulo p^n à 1 si $x \not\equiv 0 \pmod{p}$, et à 0 si $x \equiv 0 \pmod{p}$. Le nombre de solutions du système (1.2) est ainsi congru modulo p^n à l'expression

$$\sum_{\mathbf{x} \in S^n} \prod_{i=1}^r (1 - F_i(\mathbf{x})^{p^n(p-1)}).$$

Par récurrence sur r , il suffit de montrer que

$$\sum_{\mathbf{x} \in S^n} \prod_{i=1}^r F_i(\mathbf{x})^{p^n(p-1)} \equiv 0 \pmod{p^b},$$

ce que fait Wan en développant $F_i^{p^n(p-1)}$ puis en constatant que la valuation des divers symboles du multinôme et celle de la somme de puissances qui apparaissent s'ajoutent pour dépasser b . (C'est tout de même assez délicat.)

Lorsque \mathbf{F} n'est pas le corps à p éléments, on peut suivant C. Moreno et O. Moreno [58] effectuer une réduction des scalaires (à la Weil) de \mathbf{F} à \mathbf{F}_p : si on fixe une base de \mathbf{F} sur \mathbf{F}_p , de cardinal a , on se ramène à un système de ar équations en nr variables de degrés d_1, \dots, d_r (a fois). La nouvelle quantité $(n - \sum d_i) / \max(d_i)$ est égale à a fois l'ancienne. À cause de la partie entière, on en déduit une divisibilité un peu plus faible que celle affirmée par le théorème 1.3.

2. FONCTIONS ZÊTA ET COHOMOLOGIES DE WEIL

Soit \mathbf{F} un corps fini, notons q son cardinal. Dans [69], A. Weil avait introduit pour tout système d'équations polynomiales à coefficients dans \mathbf{F} , voire tout \mathbf{F} -schéma V de type fini, la série génératrice

$$Z_V(t) = \exp \left(\sum_{a \geq 0} |V(\mathbf{F}^{(a)})| \frac{t^a}{a} \right)$$

où $\mathbf{F}^{(a)}$ désigne l'unique extension de \mathbf{F} de degré a . B. Dwork [24] a démontré que c'est une fraction rationnelle (première conjecture de Weil). Ses zéros et ses pôles sont nécessairement des entiers algébriques. En fait, la congruence du théorème 1.3 implique qu'ils sont divisibles par q^b dans l'anneau des entiers algébriques.

Les deux autres conjectures de Weil ont été démontrées par A. Grothendieck (rationalité et équation fonctionnelle) dans SGA 5 et P. Deligne (analogue de l'hypothèse de Riemann, [23]), grâce à l'introduction de la cohomologie (étale) ℓ -adique, où ℓ est un nombre premier *distinct* de p . L'utilité d'une telle théorie cohomologique avait déjà été pressentie dans l'article de Weil. En effet, V dispose d'un endomorphisme de Frobenius F et l'ensemble $V(\mathbf{F}^{(a)})$ n'est autre que le lieu des points fixes de F^a ; il faudrait pouvoir disposer alors d'une formule des traces de Lefschetz. Les conditions nécessaires sur une telle cohomologie ont été rapidement formalisées sous le nom de

cohomologie de Weil : la théorie cohomologique doit vérifier la formule de Künneth, la dualité de Poincaré, fournie par une classe fondamentale, et les sous-variétés (non nécessairement lisses) doivent posséder une classe de cohomologie, de manière compatible au produit d'intersection et à la classe fondamentale. Tout ceci est exposé en détail dans [47], en lien avec les théorèmes de Lefschetz faible et difficile, les conjectures « standard » et la conjecture de Hodge.

Lorsque $\ell \neq p$, la cohomologie étale ℓ -adique est effectivement une cohomologie de Weil. Lorsque $\ell = p$, ce n'est pas vrai : $H^i(X, \mathbf{Q}_p)$ est nul dès que $i > \dim X$ ce qui viole la dualité de Poincaré.

Finalement, par voie ℓ -adique, on sait que la fonction zêta d'un \mathbf{F} -schéma géométriquement connexe, propre, lisse de dimension d , est de la forme suivante :

$$Z_X(t) = \frac{P_1(t) \cdots P_{2d-1}(t)}{P_0(t) \cdots P_{2d}(t)},$$

où $P_i(t) \in \mathbf{Z}[t]$, avec $P_0(t) = 1 - t$, $P_{2d}(t) = 1 - q^{2d}t$; si $P_i(t) = \prod_{j=1}^{b_i} (1 - a_{ij}t)$, on sait aussi que les a_{ij} sont des entiers algébriques de valeur absolue archimédienne $q^{i/2}$ (hypothèse de Riemann) et, si $\ell \neq p$, de valeur absolue ℓ -adique nulle.

L'interprétation des valuations p -adiques des a_{ij} , et notamment de congruences du type fourni par les théorèmes 1.1 et 1.3 justifie la recherche d'une théorie cohomologique de Weil qui soit p -adique.

Il y a un autre intérêt — sur lequel insiste Kedlaya — des cohomologies p -adiques que je vais décrire maintenant : elles sont par nature plus calculables que ne l'est la cohomologie étale. Il est par exemple remarquable que soient apparus récemment trois *algorithmes* efficaces (Sato [62], Kedlaya [43], Lauder-Wan [50]) pour calculer le nombre de points de certaines variétés algébriques sur un corps fini, et tous trois sont de nature p -adique. Celui de Kedlaya repose sur l'explicitation de la cohomologie de Monsky-Washnitzer, celui de Lauder-Wan est inspiré de la démonstration par Dwork de la rationalité de la fonction zêta. De même, c'est par des techniques p -adiques que Bombieri [12] a étudié le degré des numérateurs et dénominateurs de fonctions rationnelles associées à des sommes d'exponentielles.

3. COHOMOLOGIES p -ADIQUES

Mais avant cela, il faut peut-être dire pourquoi il n'existe pas de cohomologie de Weil sur la catégorie des variétés algébriques projectives lisses sur un corps algébriquement clos de caractéristique positive à valeur dans la catégorie des \mathbf{Q} -espaces vectoriels de dimension finie. Pour toute cohomologie de Weil à valeurs dans les K -vectoriels (K est un corps commutatif), le H^1 d'une courbe elliptique E est de dimension 2. De plus, pour tout endomorphisme non nul α de E , l'endomorphisme $\alpha^* : H^1(E) \rightarrow H^1(E)$ est injectif, d'où une injection de l'anneau (opposé à celui) des endomorphismes de

E dans l'anneau des matrices 2×2 à coefficients dans K . En caractéristique 0, tout irait bien (d'ailleurs, la cohomologie singulière est une cohomologie de Weil), mais en caractéristique finie, il existe des courbes elliptiques *supersingulières* dont l'anneau des endomorphismes est plus gros que prévu : c'est une algèbre de quaternions sur \mathbf{Q} . Comme une telle algèbre ne se plonge pas dans $M_2(\mathbf{Q})$, il n'y a pas de cohomologie de Weil à coefficients dans \mathbf{Q} . M. Deuring a même montré que l'algèbre de quaternions $\text{End}(E) \otimes \mathbf{Q}$ est ramifiée en p et l'infini, c'est-à-dire que $\text{End}(E) \otimes \mathbf{Q}_p$ et $\text{End}(E) \otimes \mathbf{R}$ sont des corps gauches. Cela empêche aussi $K = \mathbf{Q}_p$ ou $K = \mathbf{R}$. (Cet argument est dû à J.-P. Serre, cf. [32].)

La théorie que nous utiliserons dans cet exposé est la *cohomologie rigide*, construite par P. Berthelot. Elle unifie deux théories disjointes qui sont la cohomologie de Monsky-Washnitzer [57], valable pour les variétés affines et lisses, et la cohomologie cristalline [32, 2], qui a de bonnes propriétés pour les variétés propres et lisses. Cette théorie est encore éparpillée dans la littérature, et un petit guide de lecture ne sera peut-être pas inutile. Il faudrait aussi citer les travaux de Y. André, F. Baldassarri, P. Berthelot, B. Chiarellotto, G. Christol, R. Crew, B. Dwork, Z. Mebkhout, P. Robba, N. Tsuzuki...

Je trouve l'introduction que propose Berthelot dans [3] très agréable à lire ; le gros article [4] fournit des détails importants dans la construction (les fameux « théorèmes de fibrations »). L'article [6] est très important : outre la démonstration du fait que la cohomologie rigide d'une variété lisse est de dimension finie, on y trouve les théorèmes de comparaison avec les théories de Monsky-Washnitzer et cristalline. Concernant ce théorème de finitude, citons aussi l'article [55] de Z. Mebkhout qui propose une démonstration de la finitude de la cohomologie de Monsky-Washnitzer indépendante des théorèmes de de Jong sur l'existence d'altérations. Notons que la finitude de la cohomologie de Monsky-Washnitzer n'était pas connue avant ces deux articles, à l'exception du H^0 et du H^1 par P. Monsky. La finitude de la cohomologie rigide dans le cas général est démontrée (indépendamment) dans l'article [31] de E. Grosse-Klönne et dans celui [66] de N. Tsuzuki dans lequel il établit un théorème de descente cohomologique propre.

La dualité de Poincaré et la formule de Künneth sont établies dans la note [5] de Berthelot. Enfin, l'existence de classes de Chern est démontrée par D. Petrequin [60].

Qu'est-ce que la cohomologie rigide ? Disons tout d'abord que c'est une sorte de cohomologie de de Rham. Fixons quelques notations : soit k un corps de caractéristique $p > 0$, fixons alors un anneau de valuation discrète complet \mathcal{V} de corps résiduel k , dont nous noterons K le corps des fractions, supposé de caractéristique zéro, et π un générateur de l'idéal maximal de \mathcal{V} . Nous supposons que \mathcal{V} admet un endomorphisme σ qui se réduit modulo π en l'automorphisme de Frobenius $\sigma : x \mapsto x^p$ de k ; alors, σ s'étend en un endomorphisme de K . Dans tout ce qui va suivre, on peut se limiter au

cas important où le corps k est supposé parfait et où \mathcal{V} est l'anneau des vecteurs de Witt de k .

Soit X un k -schéma et essayons de définir la cohomologie rigide de k . Ce seront des K -espaces vectoriels. Le cas idéal est celui où X est la *réduction* modulo p d'un V -schéma propre et lisse \mathcal{X} . Dans ce cas, \mathcal{X} dispose d'une cohomologie de de Rham définie algébriquement : l'hypercohomologie du complexe des formes différentielles sur \tilde{X} ; ce sont des \mathcal{V} -modules de type fini. Un point crucial, déjà à la base de l'existence de la cohomologie de Monsky–Washnitzer, est que ces modules ne dépendent pas du choix de \mathcal{X} — autrement dit, si \mathcal{X}' est un autre \mathcal{V} -schéma propre et lisse de même réduction modulo p , on a un isomorphisme canonique $H_{DR}^*(\mathcal{X}) \simeq H_{DR}^*(\mathcal{X}')$.

Du point de vue topologique, une variété plongée dans un espace lisse est rétracte par déformation d'un voisinage tubulaire assez petit, et il est possible de définir sa cohomologie à l'aide de celle de ces voisinages. En caractéristique zéro, Hartshorne [34] avait étudié une cohomologie de de Rham pour des variétés singulières définie suivant ces lignes ; le rôle du voisinage tubulaire y est joué par le complété formel de l'espace ambiant le long de la variété dont il s'agit de définir la cohomologie.

La définition de la cohomologie rigide « naïve » suit cette approche si ce n'est qu'il faut plonger et relever. Supposons donc que X est un sous-schéma d'un schéma propre et lisse P , réduction d'un \mathcal{V} -schéma \mathcal{P} ; l'exemple important est bien entendu l'espace projectif. À \mathcal{P} est associée une variété analytique *rigide*, notée \mathbf{P} — c'est une structure plus riche que la structure analytique p -adique naïve sur $\mathcal{P} \otimes K$ qui donne lieu à une théorie des faisceaux non triviale (penser que la topologie de K est totalement discontinue!).

Un point de \mathbf{P} se *spécialise* en un point de P : dans le cas de l'espace projectif, il suffit de chasser les dénominateurs pour qu'un point soit à coordonnées homogènes dans \mathcal{V} , non toutes multiples de π , puis de réduire modulo π ces coordonnées homogènes. Dans \mathbf{P} , on peut alors définir le *tube* de X comme l'ensemble des points de \mathbf{P} qui se réduisent en un point de X . Ce tube, noté usuellement $]X[$, est une variété analytique rigide (pas forcément quasi-compacte). Par exemple, si $X = \{0\}$ et $\mathcal{P} = \mathbf{P}^1$, $]X[$ s'identifie au « disque unité ouvert », formé des $x \in K$ tels que $|x| < 1$. Si $X = \mathbf{A}^1$ et $\mathcal{P} = \mathbf{P}^1$, $]X[$ est alors le « disque unité fermé », formé des $x \in K$ tels que $|x| \leq 1$.

La cohomologie rigide naïve de X est la cohomologie de de Rham du tube de X . D'après un théorème de fibration, elle ne dépend pas du choix de \mathcal{P} . Si l'on peut prendre $P = X$, c'est-à-dire si X est propre, lisse et relevable, la cohomologie définie n'est autre que la cohomologie de de Rham tensorisée par K . Si X est propre et lisse, on retrouve la cohomologie cristalline de X tensorisée par K . Ce sont en particulier des K -espaces vectoriels de dimension finie.

En revanche, si X n'est pas propre, cela ne suffit pas. Prenons l'exemple de la droite affine $X = \mathbf{A}^1$ et de son tube $]X[= \{x ; |x| \leq 1\}$. Le complexe de de Rham dont la

cohomologie rigide naïve est la cohomologie est donné par

$$K\{x\} \rightarrow K\langle x \rangle, \quad f = \sum a_n x^n \mapsto f' = \sum n a_n x^{n-1},$$

où $K\{x\}$ est l'anneau des séries entières à coefficients dans K dont les coefficients tendent vers 0 (de sorte qu'elles convergent sur le disque fermé). On a bien $H_{\text{naïf}}^0(\mathbf{A}^1) = K$, mais $H_{\text{naïf}}^1(\mathbf{A}^1)$ n'est pas nul puisque la série $f = \sum p^n x^{p^n-1}$ n'a pas de primitive dans $K\{x\}$. En fait, $H_{\text{naïf}}^1(\mathbf{A}^1)$ est même de dimension infinie.

Monsky et Washnitzer ont remarqué que la situation s'arrange notablement si l'on remplace l'anneau $K\{x\}$ par celui des fonctions qui convergent dans un disque un peu plus gros que le disque unité (on dit qu'elles *surconvergent*). Notons $K\langle x \rangle^\dagger$ cet anneau : il est formé des séries $\sum a_n x^n$ telles que $\limsup \log|a_n|/\log n < 0$. Le complexe de de Rham surconvergent de la droite affine a alors la cohomologie attendue car si $f = \sum a_n x^n$ converge sur le disque $|x| \leq \lambda$, avec $\lambda > 1$, ses primitives convergent sur le disque ouvert $|x| < \lambda$, donc sur tout disque fermé $|x| \leq \lambda'$ avec $1 < \lambda' < \lambda$.

Pour définir la cohomologie rigide en général, il faut ainsi introduire ce que Berthelot appelle des *voisinages stricts* du tube (analogues des disques fermés $|x| \leq \lambda$ pour le disque unité) et la cohomologie du complexe de de Rham formé des formes différentielles surconvergentes, c'est-à-dire qui convergent dans un voisinage strict non précisé de $]X[$.

Outre la cohomologie rigide $H_{\text{rig}}^i(X/K)$, Berthelot définit aussi une cohomologie à support $H_{\text{rig},Z}^*(X/K)$ (pour $Z \subset X$) et une cohomologie à support propre $H_{\text{rig},c}^*(X/K)$. Ils sont de dimension finie (voir les références plus haut). Ce sont les analogues algébriques de la cohomologie d'une paire et de la cohomologie à support propre.

Ces espaces de cohomologie sont compatibles à l'extension des scalaires si K' est une extension isométrique de K , d'anneau de valuation \mathcal{V}' , de corps résiduel k' , il existe un isomorphisme canonique

$$K' \otimes_K H_{\text{rig}}^*(X/K) \xrightarrow{\sim} H_{\text{rig}}^*(X'/K'),$$

où $X' = k' \otimes_k X$. (Voir [6], prop. 1.8, pour le cas d'une extension finie, le cas général est plus difficile et est affirmé à la fin de [5].)

Disons un mot des functorialités dont disposent ces cohomologies. La cohomologie rigide est naturellement contravariante pour les morphismes de k -schémas. La cohomologie à support propre n'est contravariante que pour les morphismes propres, et est covariante pour les immersions ouvertes. Sur la cohomologie rigide des variétés lisses, la dualité de Poincaré permet d'en déduire une functorialité covariante (avec un décalage de deux fois la dimension) pour les morphismes propres (« morphismes de Gysin »).

Le morphisme de Frobenius F_X n'est pas un morphisme de k -schéma, sauf si $k = \mathbf{F}_p$, mais il se factorise en

$$X \xrightarrow{F_{X/k}} k \otimes_{\sigma} X \rightarrow X$$

où le premier morphisme est k -linéaire et le second est le morphisme de changement de base par le Frobenius σ de k . Grâce à la compatibilité des scalaires, si σ est un endomorphisme de \mathcal{V} qui induit le Frobenius modulo π , on en déduit un endomorphisme σ -linéaire de la cohomologie rigide :

$$F: H_{\text{rig}}^*(X/K) \rightarrow H_{\text{rig}}^*(X/K), \quad F(ax) = \sigma(a)F(x),$$

et de même pour les cohomologies à support et à support propre.

Les espaces de cohomologie rigide s'insèrent dans des suites exactes d'excision familières : si U est un ouvert de X et Z le fermé complémentaire, on a des suites exactes

$$H_{\text{rig,c}}^i(U/K) \rightarrow H_{\text{rig,c}}^i(X/K) \rightarrow H_{\text{rig,c}}^i(Z/K) \xrightarrow{[+1]}$$

et

$$H_{\text{rig,Z}}^i(X/K) \rightarrow H_{\text{rig}}^i(X) \rightarrow H_{\text{rig}}^i(U) \xrightarrow{[+1]}.$$

Celles-ci sont d'ailleurs équivariantes pour les divers morphismes de Frobenius (voir [15], th. 2.4).

Dans tout ceci, je n'ai en fait parlé que des « coefficients constants ». L'analogie des faisceaux localement constants est fourni par les F -cristaux surconvergentes : ce sont des fibrés vectoriels sur un voisinage strict du tube munis d'une connexion intégrable et d'une structure de Frobenius. La cohomologie à support propre d'un F -isocrystal surconvergent et la formule des traces de type Lefschetz sont étudiées dans les articles [27, 28] d'Étessé et Le Stum. La finitude de la cohomologie rigide d'un F -isocrystal surconvergent, la dualité de Poincaré et la formule de Künneth sont démontrées dans la prépublication [44] de K. Kedlaya.

La théorie des \mathcal{D} -modules arithmétiques de Berthelot est censée fournir une catégorie de coefficients stable par les six opérations de Grothendieck, mais à ma connaissance, ceci n'est pas encore démontré.

4. F-ISOCRISTAUX

(Pour ce paragraphe, l'article [41] de Katz est un *must*.) Supposons pour simplifier que k est un corps parfait de caractéristique $p > 0$, \mathcal{V} l'anneau des vecteurs de Witt de k et K son corps des fractions. Soit σ l'automorphisme de \mathcal{V} qui relève l'automorphisme de Frobenius de k ; il s'étend à K .

DÉFINITION 4.1. — *Un F -isocrystal sur K est un K -espace vectoriel de dimension finie muni d'un endomorphisme σ -linéaire injectif.*

Par exemple, soit $\alpha \in \mathbf{Q}^*$ un rationnel non nul, notons $\alpha = r/d$ avec $(r, d) = 1$ et $d > 0$, et soit $M_\alpha = K^d$, de base (e_1, \dots, e_d) , muni de l'endomorphisme σ -linéaire F donné par

$$F(e_1) = e_2, \dots, F(e_{d-1}) = e_d, F(e_d) = p^r e_1.$$

De même, les espaces de cohomologie rigide d'un schéma de type fini sont naturellement des F -isocristaux (l'injectivité n'est pas évidente et sera établie au paragraphe suivant.)

Si (M, F) est un F -isocristal, on peut exprimer F dans une K -base de M à l'aide d'une matrice $A \in M_n(K)$ ($n = \dim_K M$). Il y a aussi une notion évidente de somme directe, de produit tensoriel, extérieur, symétrique, d'homomorphisme de F -isocristaux.

Si le corps k est algébriquement clos, la catégorie des F -isocristaux a été élucidée par J. Dieudonné et Yu. Manin [51] : tout F -isocristal est somme directe de F -isocristaux (simples) du type M_α . Cela permet de définir les *pentés* d'un F -isocristal M : ce sont les rationnels α tels que M_α soit un sous-objet de M . Si $M \simeq \bigoplus M_\alpha^{n_\alpha}$, la multiplicité de la pente α dans M est par définition égale à $n_\alpha \dim M_\alpha$.

Si le corps k n'est pas algébriquement clos, les pentés d'un F -isocristal sont celles du F -isocristal obtenu après tensorisation par le corps des fractions de $W(\bar{\mathbf{F}}_p)$.

Quel que soit le corps k , pour tout rationnel α , on peut définir facilement le plus grand sous- F -isocristal $M^{\geq \alpha}$ de M dont les pentés sont $\geq \alpha$. Fixons une base de M , d'où on déduit une norme ultramétrique $\|\cdot\|$ sur M . L'ensemble $M^{\geq \alpha}$ des $x \in M$ tels que la suite $(\|F^n(x)\|p^{\alpha n})_n$ soit bornée est un sous- K -espace vectoriel de M , stable par F ; il ne dépend pas de la base choisie. Cela définit une filtration décroissante de M par des sous- F -isocristaux, exhaustive (si les coefficients d'une matrice de F sont de valuation $\geq r$, $M = M^{\geq r}$, plus un raisonnement analogue pour F^{-1}).

Pour synthétiser les pentés d'un F -isocristal, il est commode de définir son *polygone de Newton*. Si les pentés de (M, F) sont les rationnels $\alpha_1 \leq \dots \leq \alpha_n$, comptées avec multiplicités (donc $\dim M = n$), c'est par définition l'unique fonction $\text{Nwt}_M : [0; n] \rightarrow \mathbf{R}$ qui est affine par morceaux, continue, vérifie $\text{Nwt}_M(0) = 0$ et est de pente α_j sur l'intervalle $[j; j+1]$.

Lorsque le corps k est fini, on a une autre caractérisation des pentés. Supposons pour simplifier que σ^a soit l'identité, ce qui est vérifié dans le cas important où $k = \mathbf{F}_{p^a}$ et $\mathcal{V} = W(k)$. L'application F^a est alors K -linéaire et la décomposition de Jordan fournit une autre caractérisation des pentés :

PROPOSITION 4.2. — *Soit M un F -isocristal et supposons que $\sigma^a = \text{id}$. Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres de l'application K -linéaire F^a . Les pentés de M sont les $\text{ord}_p(\lambda_i)/a$.*

Combinons cette proposition avec la formule des traces de Lefschetz en cohomologie rigide : En effet, si X est un schéma séparé de type fini sur un corps fini \mathbf{F}_q , avec

$q = p^a$, on a pour tout entier $n \geq 0$,

$$|X(\mathbf{F}_q)| = \sum_{i=0}^{2 \dim X} (-1)^i \operatorname{Tr}(F^{an} | H_{\text{rig},c}^i(X/K)).$$

Ainsi, on constate que *minorer* les pentes de la cohomologie rigide à support propre fournit des congruences p -adiques pour sa fonction zêta. Précisément : contrôler la partie de pente 0 implique des congruences modulo p pour $|X(\mathbf{F}_q)|$, contrôler la partie de pentes < 1 des congruences modulo q .

Dans cette veine, il faut citer un résultat fondamental, dû à Mazur [53, 54] moyennant des hypothèses restrictives, et Ogus [7, chap. 8] en général. Soit X un k -schéma propre et lisse ; sa cohomologie rigide (cristalline en fait) fournit un F -isocrystal $H_{\text{cris}}^m(X/W(k)) \otimes \operatorname{Frac} W(k)$ dont nous noterons $\operatorname{Nwt}_X^{(m)}$ le polygone de Newton. Par ailleurs, X a des nombres de Hodge $h_i = \dim_k H^{m-i}(X, \Omega_{X/k}^i)$. Définissons le m -ième polygone de Hodge de X , $\operatorname{Hdg}_X^{(m)}$, comme la fonction continue affine par morceaux qui vaut 0 en 0, est de pente 0 sur l'intervalle $[0; h_0]$, de pente 1 sur l'intervalle $[h_0; h_0 + h_1]$, etc.

THÉORÈME 4.3. — *On a l'inégalité $\operatorname{Nwt}_X^{(m)} \geq \operatorname{Hdg}_X^{(m)}$.*

On en déduit des théorèmes de type Chevalley-Warning et notamment une autre démonstration du théorème 1.3 de Ax-Katz *dans le cas lisse et homogène*.

COROLLAIRE 4.4 (à la Chevalley-Warning). — *Soit X une intersection complète lisse de dimension d dans \mathbf{P}^{n-1} , définie sur le corps fini \mathbf{F}_q . Alors, le nombre de points de $X(\mathbf{F}_{q^s})$ est égal au nombre de points de $\mathbf{P}^d(\mathbf{F}_{q^s})$ modulo q^{bs} où b est le plus petit entier $i \geq 0$ tel que $\dim H^{d-i}(\Omega^i) \neq 0$. (Si d_1, \dots, d_r sont les degrés des hypersurfaces qui définissent X , l'entier b est égal au plus petit entier supérieur ou égal à $(n - \sum d_i) / \max(d_i)$.)*

5. PENTES DE LA COHOMOLOGIE RIGIDE

Je donne dans ce paragraphe deux résultats généraux concernant les pentes de la cohomologie rigide à support propre. Le premier décrit la partie de pente 0, le second précise les pentes possibles.

THÉORÈME 5.1. — *Pour tout schéma X , séparé et de type fini sur un corps algébriquement clos de caractéristique $p > 0$, on a un isomorphisme canonique*

$$H_{\text{ét}}^i(X, \mathbf{Q}_p) \otimes K \simeq H_{\text{rig},c}^i(X/K)^0,$$

où l'exposant 0 signifie qu'on considère la partie de pente 0 dans le F -isocrystal donné par la cohomologie rigide.

Lorsque X est propre et lisse, ce théorème est dû à Bloch (lorsque p est assez grand) et Illusie (pour tout p , [35], II, 5.4). Dans le cas général, c'est un résultat d'Étesse et Le Stum ([28], prop. 6.3).

Ils se déduisent dans ce cas des propriétés du complexe de de Rham-Witt via une généralisation de la suite exacte d'Artin-Schreier

$$0 \rightarrow (\mathbf{Z}/p\mathbf{Z})_X \rightarrow \mathcal{O}_X \xrightarrow{1-F} \mathcal{O}_X \rightarrow 0$$

(suite exacte de faisceaux étales sur X). Plus généralement, si $W_n\mathcal{O}_X$ désigne le faisceau des vecteurs de Witt de longueur n sur X , on a une suite exacte de faisceaux étales sur X :

$$0 \rightarrow (\mathbf{Z}/p^n\mathbf{Z})_X \rightarrow W_n\mathcal{O}_X \xrightarrow{1-F} W_n\mathcal{O}_X \rightarrow 0$$

qui induit par passage à la cohomologie, puis passage à la limite une suite exacte en cohomologie étale :

$$0 \rightarrow H_{\text{ét}}^i(X, \mathbf{Z}_p) \rightarrow H^i(X, W\mathcal{O}_X) \xrightarrow{1-F} H^i(X, W\mathcal{O}_X) \rightarrow 0$$

qui identifie $H_{\text{ét}}^i(X, \mathbf{Q}_p) \otimes K$ au plus grand sous- F -isocristal de pente 0 dans le F -isocristal $H^i(X, W\mathcal{O}_X) \otimes_{W(k)} K$. (Compte tenu du fait que $H^i(X, W\mathcal{O}_X)$ est pour tout entier i un $W(k)$ -module de type fini, la surjectivité de $1 - F$ est une propriété générale des F -cristaux, cf. par exemple [35], II, 5.3.) La théorie du complexe de de Rham-Witt de Bloch et Illusie, et en particulier la dégénérescence de la suite spectrale des pentes ([35], II, 3.5), implique que ce dernier espace vectoriel est le plus grand sous- F -isocristal de pentes $[0; 1[$ dans $H_{\text{rig}}^i(X/K)$. Le résultat s'ensuit si X est propre et lisse.

Dans le cas général, Étesse et Le Stum combinent des suites exactes d'Artin-Schreier, le calcul syntomique de la cohomologie cristalline (initié par Fontaine et Messing dans [29]), la cohomologie cristalline « de niveau variable » et un théorème de Berthelot selon lequel cette dernière permet de calculer la cohomologie rigide.

THÉORÈME 5.2. — *Soit k un corps parfait de caractéristique $p > 0$ et soit X un k -schéma de type fini de dimension d .*

a) *pour tout i , $H_{\text{rig},c}^i(X/K)$ est un F -isocristal dont les pentes appartiennent à l'intervalle $[\max(0, i - d), \min(i, d)]$;*

b) *si X est lisse, c'est encore vrai de $H_{\text{rig}}^i(X/K)$.*

Le b) est conséquence du a) compte tenu de la dualité de Poincaré en cohomologie rigide : il existe un morphisme trace

$$H_{\text{rig},c}^{2d}(X/K) \xrightarrow{\text{Tr}} K$$

tel que $\text{Tr} \circ F = p^d F$ qui induit par *cup-produit* des isomorphismes (X est lisse)

$$H_{\text{rig}}^i(X/K) \simeq H_{\text{rig},c}^{2d-i}(X/K)^\vee(-d)$$

où le *twist* $(-d)$ signifie que le Frobenius est multiplié par p^d . Notons au passage que cela implique l'injectivité des Frobenius sur la cohomologie rigide (*resp.* la cohomologie rigide à support propre).

Avant de montrer le a), faisons quelques remarques.

1) On peut supposer que le corps k est algébriquement clos et que X est réduit. On peut aussi supposer que X est connexe car la cohomologie rigide d'une réunion disjointe est la somme directe des cohomologies rigides.

2) Si X est projective et lisse, le théorème de comparaison entre cohomologies rigide et cristalline implique que $H_{\text{rig}}^i(X/K)$ est un F-isocristal à pentes positives ou nulles. Par dualité de Poincaré, elles sont donc $\leq d$. Le théorème de Lefschetz faible en cohomologie cristalline implique alors que les pentes appartiennent à l'intervalle $[0; i]$, donc à l'intervalle $[i - d; i]$ via la dualité de Poincaré.

3) Si U est un ouvert dense de X et Z le fermé complémentaire, la suite exacte longue d'excision

$$\rightarrow H_{\text{rig,c}}^{i-1}(Z/K) \rightarrow H_{\text{rig,c}}^i(U/K) \rightarrow H_{\text{rig,c}}^i(X/K) \rightarrow$$

et l'hypothèse que l'assertion a) est vérifiée en dimension $< \dim X$ montrent qu'il est équivalent de démontrer l'énoncé a) pour X et pour U .

4) Si $U' \rightarrow U$ est un revêtement étale de k -variétés lisses, la cohomologie rigide à support propre de U' admet celle de U comme facteur direct, si bien qu'il suffit de démontrer l'assertion a) pour U' .

Démontrons maintenant a) par récurrence sur la dimension de X . D'après le théorème d'altération de de Jong ([36], th. 4.1), il existe un ouvert dense $U \subset X$, une k -variété projective et lisse X' , un ouvert $U' \subset X'$ et un revêtement étale $U' \rightarrow U$.

D'après 2), l'assertion a) est vraie pour X' . D'après 3), elle est donc vraie pour U' et la remarque 4) entraîne sa véracité pour U , donc aussi pour X grâce à 3).

Remarque 5.3 (Références bibliographiques). — La démonstration est celle suggérée par Berthelot dans [6], remarque 3.9 et se trouve aussi dans l'article [17] de B. Chiarelotto et B. Le Stum. En suivant cette approche, ces auteurs ont aussi élucidé la structure des poids (c'est-à-dire des valeurs absolues archimédiennes des valeurs propres de Frobenius) sur la cohomologie rigide à support propre d'une variété sur un corps fini (cf. [15] et [16]). Ils doivent faire usage du théorème de Katz-Messing : dans [42], ces derniers déduisent des conjectures de Weil et du théorème de Lefschetz difficile en cohomologie ℓ -adique les théorèmes correspondants en cohomologie cristalline. Signalons aussi que Kedlaya a récemment adapté à la cohomologie rigide (voir [45]). la démonstration par Laumon des conjectures de Weil.

On peut en fait démontrer un analogue de ce théorème sur un corps fini, via la cohomologie étale ℓ -adique, et c'est ainsi que procède Ekedahl dans [25]. Cela nécessite de montrer au préalable que les valeurs propres de Frobenius sur la cohomologie ℓ -adique à support propre sont des entiers algébriques, ce qui est fait par Deligne (§ 5

de l'exposé [40]). On peut alors, par un dévissage analogue, étudier leurs valuations p -adiques.

Enfin, signalons un article de M. Kim [46] dans lequel la cohomologie rigide est remplacée par celle d'un complexe de de Rham-Witt à pôles logarithmiques.

6. LES THÉORÈMES D'EKEDAHL ET ESNAULT

Dans ce paragraphe, je présente deux théorèmes dus à T. Ekedahl [25] et H. Esnault [26] qui permettent de contrôler la partie de pentes < 1 dans la cohomologie rigide. Leur utilité apparaîtra aux paragraphes suivants, lorsque nous déduirons de cette partie de pente < 1 des renseignements géométriques. Dans [38], B. Kahn redémontre ces énoncés analogues à l'aide des *motifs birationnels*, notion qu'il a introduite avec R. Sujatha.

Si (M, F) est un F -isocrystal et si $\alpha \in \mathbf{Q}$, notons $M^{>\alpha}$ et $M^{<\alpha}$ les plus grands sous- F -isocristaux de M dont les pentes soient respectivement strictement supérieures et inférieures à α .

THÉORÈME 6.1 (Ekedahl). — *Soit k un corps parfait de caractéristique positive. Soient X et Y deux k -schémas de type fini, intègres de même dimension d .*

a) *S'il existe une application rationnelle dominante $X \rightarrow Y$, il existe pour tout i , une injection de F -isocristaux (canonique) de $H_{\text{rig},c}^i(Y)^{>d-1}$ dans $H_{\text{rig},c}^i(X)^{>d-1}$;*

a') *S'il existe une application rationnelle dominante $X \rightarrow Y$ et si X et Y sont lisses, $H_{\text{rig}}^i(Y)^{<1}$ s'injecte canoniquement dans $H_{\text{rig}}^i(X)^{<1}$.*

b) *Si X et Y sont birationnels, alors pour tout i , $H_{\text{rig},c}^i(X)^{>d-1}$ et $H_{\text{rig},c}^i(Y)^{>d-1}$ sont des F -isocristaux isomorphes.*

b') *Si X et Y sont lisses et birationnels, alors pour tout i , les F -isocristaux $H_{\text{rig}}^i(X)^{<1}$ et $H_{\text{rig}}^i(Y)^{<1}$ sont isomorphes.*

Les énoncés a') et b') se déduisent des énoncés a) et b) par dualité de Poincaré. Il est par ailleurs clair que a) implique b), ce qui nous ramène à démontrer l'énoncé a).

Il existe des ouverts lisses et denses $U \subset X$ et $V \subset Y$ sur lesquels l'application rationnelle $X \rightarrow Y$ induit un morphisme fini et plat $f: U \rightarrow V$. Les suites exactes d'excision associées aux ouverts U et V et l'étude des pentes de la cohomologie rigide à support propre impliquent que l'on a des isomorphismes (induits par les inclusions)

$$H_{\text{rig},c}^i(U/K)^{>d-1} \rightarrow H_{\text{rig},c}^i(X/K)^{>d-1}, \quad H_{\text{rig},c}^i(V/K)^{>d-1} \rightarrow H_{\text{rig},c}^i(Y/K)^{>d-1}.$$

Par ailleurs, sont attachés à f deux morphismes de F -isocristaux

$$f_*: H_{\text{rig},c}^i(U/K) \rightarrow H_{\text{rig},c}^i(V/K) \quad \text{et} \quad f^*: H_{\text{rig},c}^i(V/K) \rightarrow H_{\text{rig},c}^i(U/K)$$

tels que $f_* \circ f^* = \text{deg}(f)$. Il en résulte que f^* est injectif et le résultat s'en déduit.

THÉORÈME 6.2 (Esnault). — Soit k un corps de caractéristique $p > 0$. Soit X une variété propre et lisse telle que, notant Ω une clôture algébrique du corps des fonctions $k(X)$, on ait $\mathrm{CH}_0(X_\Omega)_{\mathbf{Q}} = \mathbf{Q}$. Alors, pour tout $i > 0$, $H_{\mathrm{rig}}^i(X/K)^{<1} = 0$.

Une classe importante de variétés où la condition $\mathrm{CH}_0(X_\Omega)_{\mathbf{Q}} = \mathbf{Q}$ est vérifiée est fournie par les variétés (géométriquement) rationnellement connexes par chaînes : il s'agit de variétés propres X pour lesquelles deux points quelconques sont joints par une chaîne de courbes rationnelles. Les variétés faiblement unirationnelles sont évidemment rationnellement connexes par chaînes ; l'énoncé b) redonne ainsi l'un des résultats principaux de la note [25]. D'après un théorème dû indépendamment à F. Campana et à J. Kollár, Y. Miyaoka et S. Mori, les variétés de Fano, c'est-à-dire les variétés projectives lisses dont le fibré anticanonique est ample, sont rationnellement connexes par chaînes (voir [20], prop. 5.16.) Si le corps de base est de caractéristique zéro, par deux points quelconques d'une variété propre et lisse qui est rationnellement connexe par chaînes, passe en fait une courbe rationnelle.

Pour démontrer le théorème 6.2, on peut supposer que k est algébriquement clos. D'après un théorème de Bloch ([8, 9], je rappelle aussi l'argument plus bas), il existe un ouvert dense U de X et un point $x_0 \in X(k)$ tels que sur $U \times X$, notant Δ_X la diagonale de $X \times X$,

$$(6.3) \quad \Delta_X = U \times [x_0] \quad \text{dans } \mathrm{CH}_d(U \times X)_{\mathbf{Q}}.$$

Autrement dit, le graphe $\Gamma_j \subset U \times X$ de l'immersion fermée $j : U \rightarrow X$ est linéairement équivalent à $U \times [x_0]$ dans $\mathrm{CH}_d(U \times X)_{\mathbf{Q}}$.

Remarquons qu'une classe de cohomologie $\alpha \in H_{\mathrm{rig}}^{2d}(U \times X)$ définit une correspondance sur la cohomologie rigide à support propre :

$$\alpha_* : H_{\mathrm{rig},c}^i(U) \xrightarrow{q^*} H_{\mathrm{rig},c}^i(U \times X) \xrightarrow{\cap \alpha} H_{\mathrm{rig},c}^{i+2d}(U \times X) \xrightarrow{p_*} H_{\mathrm{rig},c}^i(X).$$

L'application q^* existe car X est propre, l'accouplement $\cap \alpha$ vient de celui entre cohomologie rigide à support propre et cohomologie rigide (défini dans [5]) et l'application p_* est la transposée, via la dualité de Poincaré, de l'application p^* sur la cohomologie rigide. Le graphe de j a une classe $\gamma(\Gamma_j)$ dans $H_{\mathrm{rig}}^{2d}(U \times X)$, construite dans [60] et l'on a pour tout i ,

$$\gamma(\Gamma_j)_* = j_* : H_{\mathrm{rig},c}^i(U) \rightarrow H_{\mathrm{rig},c}^i(X).$$

Mais la décomposition (6.3) affirme que modulo l'équivalence rationnelle, $\Gamma_j = p^*[x_0]$, donc, en passant aux classes de cycles ([60], prop. 6.10),

$$\gamma(\Gamma_j) = p^*\gamma([x_0]), \quad \gamma([x_0]) \in H_{\mathrm{rig}}^{2d}(X)$$

Par suite, pour toute classe $\xi \in H_{\mathrm{rig},c}^i(U)$,

$$\begin{aligned} j_*(\xi) &= \gamma(\Gamma_j)_*(\xi) = p_*(q^*(\xi) \cap p^*(\gamma([x_0]))) \\ &= p_*(q^*(\xi)) \cap \gamma([x_0]). \end{aligned}$$

Comme $p_*(q^*(\xi)) \in H_{\text{rig},c}^{i-2d}(X)$, il est nul pour $i < 2d$ et finalement, les homomorphismes $j_*: H^i_{\text{rig},c}(U) \rightarrow H^i_{\text{rig},c}(X)$ sont nuls pour $0 \leq i < 2d$. (En degré $2d$, c'est un isomorphisme.) D'après le théorème 6.1, ils induisent des isomorphismes sur les parties de pentes $> d - 1$. Par conséquent, pour tout i , le F-isocristal $H_{\text{rig},c}^i(X)^{>d-1}$ est nul.

Par dualité de Poincaré, $H_{\text{rig}}^i(X)^{<1} = 0$ si $i > 0$.

Donnons pour terminer la démonstration de l'existence d'une décomposition (6.3). L'hypothèse est que $\text{CH}_0(X_\Omega)_{\mathbf{Q}} = \mathbf{Q}$, où Ω est une clôture algébrique de $k(X)$. On suppose toujours que le corps k est algébriquement clos. Soit x_0 un point géométrique de X et soit η son point générique; ils définissent deux points de $X(k(X))$ et leur différence α est un 0-cycle sur $X_{k(X)}$. Par hypothèse, leur différence est de torsion sur X_Ω , donc sur une extension finie K de $k(X)$. En prenant des normes, on voit que α est déjà de torsion dans $\text{CH}_0(X_{k(X)})$. Cela signifie qu'il existe un ouvert U de X tel que le d -cycle $\bar{\alpha} = X \times [x_0] - \Delta_X$ est de torsion sur $U \times X$.

Remarque 6.4. — Si $\dim X \leq 3$, et si X est séparablement unirationnelle, ou bien de Fano, on sait que les groupes de cohomologie $H^i(X, \mathcal{O}_X)$ sont nuls pour $i \geq 1$ (Nygaard [59], Shepherd-Barron [64]). On peut alors en déduire que pour $i \geq 1$, $H^i(X, W\mathcal{O}_X) = 0$, d'où une autre approche au théorème 6.2.

Remarque 6.5. — Dans [26], H. Esnault énonce son théorème sous l'hypothèse, apparemment plus forte, que $\text{CH}_0(X_\Omega) = \mathbf{Z}$; elle est en fait équivalente. Notons $\text{CH}_0(X_\Omega)^0$ le noyau de l'application degré $\text{CH}_0(X_\Omega) \rightarrow \mathbf{Z}$ et soit $\text{alb}: X \rightarrow A$ l'application d'Albanese de X_Ω . Comme A est engendrée par l'image de X , l'application $\text{CH}_0(X_\Omega)^0 \rightarrow A(\Omega)$ déduite de alb est surjective. Si $\text{CH}_0(X_\Omega)_{\mathbf{Q}} = \mathbf{Q}$, $\text{CH}_0(X_\Omega)^0$ est un groupe de torsion, $A(\Omega)$ aussi, et donc $A = 0$. Par ailleurs, un théorème de Roitman [61] complété par Milne [56] affirme que l'application $\text{CH}_0(X_\Omega)^0 \rightarrow A(\Omega)$ induit un isomorphisme sur les sous-groupes de torsion. Il en résulte que $\text{CH}_0(X_\Omega)^0$ est sans torsion, donc nul. Ainsi, $\text{CH}_0(X_\Omega) = \mathbf{Z}$.

7. RETOUR SUR LA FONCTION ZÊTA

Comme je l'ai déjà mentionné au § 5, la formule des traces de Lefschetz en cohomologie rigide à support propre montre que la partie de la cohomologie de pente $< \alpha$ fournit une congruence modulo q^α pour le nombre de points rationnels.

Par suite, les théorèmes 6.1 et 6.2 impliquent le théorème suivant :

THÉORÈME 7.1. — *Soit X et Y deux variétés propres, lisses et géométriquement connexes sur le corps fini \mathbf{F}_q . Soit Ω une clôture algébrique de $\mathbf{F}_q(X)$.*

- a) *Si X et Y sont birationnelles, alors $|X(\mathbf{F}_q)| \equiv |Y(\mathbf{F}_q)| \pmod{q}$;*
- b) *Si $\text{CH}_0(X_\Omega)_{\mathbf{Q}} = \mathbf{Q}$, alors $|X(\mathbf{F}_q)| \equiv 1 \pmod{q}$.*

Même si l'énoncé a) ne figure pas explicitement dans la note [25], c'en est une conséquence immédiate. D'un autre côté, il est évident si X et Y sont liées par un éclatement de centre lisse. Il le serait plus généralement pour tout couple de variétés birationnelles, si l'on disposait d'un théorème de factorisation faible en caractéristique positive. Dans leur article récent [49], G. Lachaud et M. Perret ont fait marcher cette approche en dimension 3.

Puisque les variétés de Fano sur un corps algébriquement clos sont rationnellement connexes par chaînes, il en résulte ainsi le théorème, conjecturé par Lang et Manin [52] :

COROLLAIRE 7.2 (Esnault). — *Soit X une variété de Fano sur un corps fini \mathbf{F}_q . Alors, $|X(\mathbf{F}_q)| \equiv 1 \pmod{q}$. Il est en particulier non nul.*

On en déduit aussi le résultat :

PROPOSITION 7.3. — *Soit X une variété propre, lisse et géométriquement connexe sur un corps fini \mathbf{F}_q . Supposons que X soit géométriquement dominée par une k -variété Y , propre, lisse et connexe, de même dimension, telle que $H_{\text{ét}}^i(Y, \mathbf{Q}_p) = 0$ si $i \neq 0$. Alors, pour toute extension finie \mathbf{F} de \mathbf{F}_q , $|X(\mathbf{F})| \equiv 1 \pmod{p}$.*

Remarque 7.4. — Le théorème d'Esnault peut être mis en parallèle avec plusieurs résultats récents. Soit k le corps des fonctions d'une courbe projective et lisse sur un corps algébriquement clos. Graber, Harris, Starr [30], et de Jong et Starr [37] ont montré qu'une k -variété projective, lisse qui est séparablement rationnellement connexe a un point rationnel. Un tel corps k , de même qu'un corps fini, est C_1 , donc de dimension cohomologique au plus 1. Cependant, Colliot-Thélène et Madore ont construit dans [18] une surface cubique sur \mathbf{Q} (une telle surface est séparablement rationnellement connexe) et un corps de dimension cohomologique 1 sur laquelle elle n'a pas de point rationnel. La question de savoir si une variété séparablement rationnellement connexe (voire rationnellement connexe par chaînes) sur un corps C_1 admet un point rationnel reste ouverte.

Remarque 7.5. — Contrairement aux théorèmes d'Ax et Katz, la démonstration du théorème 6.2 nécessite une hypothèse de lissité. Bloch, Esnault et Levine [10] ont proposé de remplacer la décomposition de la diagonale dans le groupe de Chow (formule (6.3)) par une décomposition analogue dans un groupe de *cohomologie motivique* adéquat. Leur condition entraîne une minoration des pentes de la cohomologie rigide et ils ont montré qu'elle est vérifiée dans le cas des hypersurfaces de degré $d \leq n$ éventuellement singulières de l'espace projectif \mathbf{P}^n .

8. SIMPLE CONNEXITÉ DE CERTAINES VARIÉTÉS

On peut aussi appliquer ces considérations pour étudier le groupe fondamental de certaines variétés algébriques, notamment les variétés unirationnelles.

Rappelons que J-P. Serre a démontré dans [63] que le groupe fondamental d'une telle variété est trivial, pourvu que le corps de base soit *de caractéristique zéro*.

LEMME 8.1. — *Soit X une variété projective lisse, géométriquement connexe, sur un corps de caractéristique zéro. On suppose que X est unirationnelle, ou que X est de Fano, ou, Ω désignant la clôture algébrique du corps des fonctions de X , que $\mathrm{CH}_0(X_\Omega)_\mathbf{Q} = \mathbf{Q}$. Alors, $\chi(X, \mathcal{O}_X) = 1$.*

Dans le premier cas, on a en effet $H^0(X, \Omega_X^i) = 0$, pour $i > 0$, comme on le voit en tirant une i -forme de X à l'espace projectif : elle y sera régulière hors d'un lieu de codimension 2, donc partout, donc nulle. En caractéristique zéro, cet espace a même dimension que $H^i(X, \mathcal{O}_X)$, d'où l'assertion. Dans le second cas, si $i > 0$, $H^i(X, \mathcal{O}_X)$ est dual de $H^{d-i}(X, \omega_X^{-1})$, donc est nul par le théorème d'annulation de Kodaira.

Le dernier cas se démontre en décomposant la diagonale mais en théorie de Hodge. Par le même argument que dans la preuve du théorème 6.2, il existe un ouvert dense U de X tel que l'application $H_{DR}^i(X) \rightarrow H_{DR}^i(U)$ soit nulle pour $i > 0$. D'autre part, la théorie de Hodge mixte de P. Deligne fournit une factorisation

$$H_{DR}^i(X) \rightarrow H_{DR}^i(U) \rightarrow H^i(X, \mathcal{O}_X),$$

l'application composée étant surjective (cf. [22], (3.2.13), (ii)). Il en résulte que $H^i(X, \mathcal{O}_X) = 0$ pour $i > 0$.

COROLLAIRE 8.2. — (En caractéristique zéro.) *Une variété propre et lisse qui est unirationnelle, ou de Fano, ou rationnellement connexe par chaînes, n'a pas de revêtement étale fini non trivial.*

Si $f: Y \rightarrow X$ est un tel revêtement, remarquons que l'hypothèse implique que Y est aussi unirationnelle (*resp.* de Fano). Par suite, on a $\chi(Y, \mathcal{O}_Y) = 1$. Or, le théorème de Riemann-Roch implique que $f_* \mathrm{ch}(\mathcal{O}_Y) = \mathrm{deg}(f) \mathrm{ch}(\mathcal{O}_X)$, si bien que $\chi(Y, \mathcal{O}_Y) = \mathrm{deg}(f) \chi(X, \mathcal{O}_X)$. Nécessairement, $\mathrm{deg}(f) = 1$.

Remarque 8.3. — a) Si le corps de base est \mathbf{C} , le groupe fondamental topologique de telles variétés est aussi trivial.

b) Soit X une surface d'Enriques sur le corps des nombres complexes. Bloch, Kas et Lieberman montrent dans [11] que sur X , tout zéro-cycle et de degré nul est rationnellement équivalent à 0. L'hypothèse $\mathrm{CH}_0(X_\Omega)_\mathbf{Q} = \mathbf{Q}$ ne suffit donc pas à assurer la validité du corollaire précédent.

En caractéristique $p > 0$, les choses sont plus compliquées. Tout d'abord, il existe des surfaces unirationnelles non simplement connexes (Shioda, [65]) : si $p \neq 5$ et $p \not\equiv 1 \pmod{5}$, la surface d'équation

$$X_0^5 + X_1^5 + X_2^5 + X_3^5 = 0$$

dans \mathbf{P}^3 est unirationnelle mais possède une action libre du groupe des racines 5-ièmes de l'unité (par $X_i \rightarrow \zeta^i X_i$). La surface de Godeaux obtenue par quotient est alors unirationnelle mais n'est pas simplement connexe : son groupe fondamental est $\mathbf{Z}/5\mathbf{Z}$. Cet exemple montre aussi qu'en caractéristique positive, une variété rationnellement connexe par chaînes X ne vérifie pas forcément $H^i(X, \mathcal{O}_X) = 0$.

Toutefois, on peut démontrer qu'une variété propre, définie sur un corps algébriquement clos qui est normale et rationnellement connexe par chaînes, a un groupe fondamental fini (Kollàr [48], voir aussi Campana [13], ainsi que la note [14]). De plus, les variétés dites *séparablement rationnellement connexes* sont simplement connexes. (Cela résulte du théorème de de Jong et Starr, cf. l'exposé [21] d'O. Debarre).

Le résultat suivant a été démontré par T. Ekedahl ([25]) dans le cas unirationnel.

PROPOSITION 8.4. — *Soit k un corps algébriquement clos de caractéristique $p > 0$ et soit X une variété propre, lisse sur k . Si X est de Fano, ou si X est rationnellement connexe par chaînes, son groupe fondamental est un groupe fini d'ordre premier à p .*

D'après le lemme 5.1, si X est unirationnelle, *resp.* de Fano, *resp.* rationnellement connexe par chaînes, sur un corps algébriquement clos k de caractéristique p , on a $\chi_{\text{ét},c}(X, \mathbf{Q}_p) = 1$. Mais un revêtement étale d'un tel X est aussi unirationnel (*resp.* de Fano, *resp.* rationnellement connexe par chaînes). Soit ainsi un revêtement étale $Y \rightarrow X$, galoisien de groupe G et soit P un p -sous-groupe de Sylow de G , de sorte que le revêtement $Y \rightarrow Y/P$ est étale galoisien de groupe P . Les variétés Y et Y/P sont rationnellement connexes par chaînes, si bien que $\chi_{\text{ét}}(Y, \mathbf{Q}_p) = \chi_{\text{ét}}(Y/P, \mathbf{Q}_p) = 1$. Le lemme 8.5 ci-dessous affirme que $\chi_{\text{ét}}(Y, \mathbf{Q}_p) = |P|\chi_{\text{ét}}(Y/P, \mathbf{Q}_p)$. Par suite, $P = \{1\}$ et tout revêtement étale de X est d'ordre premier à p . La proposition résulte alors de ce que le groupe fondamental de X est fini.

Pour que la démonstration de la prop. 8.4 soit complète, il reste à démontrer la formule d'Euler-Poincaré en cohomologie étale p -adique suivante, établie par R. Crew dans [19].

LEMME 8.5. — *Soit k un corps algébriquement clos de caractéristique p et soit $f: Y \rightarrow X$ un revêtement étale galoisien de k -schémas séparés de type fini, de degré une puissance de p . On a alors les formules suivantes entre caractéristiques d'Euler-Poincaré en cohomologie étale à support propre :*

$$\chi_{\text{ét},c}(Y, \mathbf{Q}_p) = \deg(f)\chi_{\text{ét},c}(X, \mathbf{Q}_p) \quad \text{et} \quad \chi_{\text{ét},c}(Y, \mathbf{Z}/p\mathbf{Z}) = \deg(f)\chi_{\text{ét},c}(X, \mathbf{Z}/p\mathbf{Z}).$$

En considérant la suite exacte

$$0 \rightarrow \mathbf{Z}_p \rightarrow \mathbf{Z}_p \rightarrow \mathbf{Z}/p \rightarrow 0$$

et en utilisant le fait que la cohomologie étale à support propre à coefficients dans \mathbf{Z}_p est représentée par un complexe parfait de \mathbf{Z}_p -modules, on démontre que

$$\chi_{\text{ét,c}}(X, \mathbf{Q}_p) = \chi_{\text{ét,c}}(X, \mathbf{Z}/p\mathbf{Z}) \quad \text{et} \quad \chi_{\text{ét,c}}(X, \mathbf{Q}_p) = \chi_{\text{ét,c}}(X, \mathbf{Z}/p\mathbf{Z}),$$

si bien que la seconde formule implique la première.

Le faisceau étale $f_*(\mathbf{Z}/p\mathbf{Z})$ sur X est localement constant et correspond à une représentation de son groupe fondamental sur un \mathbf{F}_p -espace vectoriel de dimension $d = \deg(f)$, représentation qui provient d'une représentation de $\text{Gal}(Y/X)$. Comme ce groupe est un p -groupe, cette représentation est extension successive de représentations triviales. Autrement dit, $f_*(\mathbf{Z}/p\mathbf{Z})$ est extension successive de d faisceaux $\mathbf{Z}/p\mathbf{Z}$. Par suite

$$\chi_{\text{ét,c}}(Y, \mathbf{Z}/p\mathbf{Z}) = \chi_{\text{ét,c}}(X, f_*(\mathbf{Z}/p\mathbf{Z})) = d\chi_{\text{ét,c}}(X, \mathbf{Z}/p\mathbf{Z}).$$

RÉFÉRENCES

- [1] J. AX – « Zeroes of polynomials over finite fields », *Amer. J. Math.* **86** (1964), p. 255–261.
- [2] P. BERTHELOT – *Cohomologie cristalline des schémas de caractéristique $p > 0$* , Lecture Notes in Math., no. 407, Springer Verlag, 1974.
- [3] ———, « Géométrie rigide et cohomologie des variétés algébriques de caractéristique p », *Mém. Soc. Math. France* (1986), no. 23, p. 7–32, Introductions aux cohomologies p -adiques (Luminy, 1984).
- [4] ———, « Cohomologie rigide et cohomologie rigide à supports propres. Première partie », Prépublication, IRMAR, Université Rennes 1, 1996.
- [5] ———, « Dualité de Poincaré et formule de Künneth en cohomologie rigide », *C. R. Acad. Sci. Paris Sér. I Math.* **325** (1997), no. 5, p. 493–498.
- [6] ———, « Finitude et pureté cohomologique en cohomologie rigide », *Invent. Math.* **128** (1997), no. 2, p. 329–377, Avec un appendice en anglais par A. J. de Jong.
- [7] P. BERTHELOT & A. OGUS – *Notes on crystalline cohomology*, Math. Notes, no. 21, Princeton Univ. Press, 1978.
- [8] S. BLOCH – *Lectures on algebraic cycles*, Duke University Mathematics Series, IV, Duke University Mathematics Department, Durham, N.C., 1980.
- [9] ———, « On an argument of Mumford in the theory of algebraic cycles », in *Journées de Géométrie Algébrique d'Angers, Juillet 1979/Algebraic Geometry, Angers, 1979*, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980, p. 217–221.

- [10] S. BLOCH, H. ESNAULT & M. LEVINE – « Decomposition of the diagonal and eigenvalues of Frobenius for Fano hypersurfaces », 2003, *arXiv:math.AG/0302109*.
- [11] S. BLOCH, A. KAS & D. I. LIEBERMAN – « Zero cycles on surfaces with $p_g = 0$ », *Compositio Math.* **33** (1976), no. 2, p. 135–145.
- [12] E. BOMBIERI – « On exponential sums in finite fields. II », *Invent. Math.* **47** (1978), no. 1, p. 29–39.
- [13] F. CAMPANA – « Remarques sur les groupes de Kähler nilpotents », *Ann. Sci. École Norm. Sup.* **28** (1995), p. 307–316.
- [14] A. CHAMBERT-LOIR – « À propos du groupe fondamental des variétés rationnellement connexes par chaînes », 2003, *arXiv:math.AG/0303051*.
- [15] B. CHIARELLOTTO – « Weights in rigid cohomology applications to unipotent F -isocrystals », *Ann. Sci. École Norm. Sup.* **31** (1998), no. 5, p. 683–715.
- [16] B. CHIARELLOTTO & B. LE STUM – « F -isocristaux unipotents », *Compositio Math.* **116** (1999), no. 1, p. 81–110.
- [17] ———, « Pentes en cohomologie rigide et F -isocristaux unipotents », *Manuscripta Math.* **100** (1999), no. 4, p. 455–468.
- [18] J.-L. COLLIOT-THÉLÈNE & D. A. MADORE – « Surfaces de Del Pezzo sans point rationnel sur un corps de dimension cohomologique un », 2003.
- [19] R. M. CREW – « Etale p -covers in characteristic p », *Compositio Math.* **52** (1984), no. 1, p. 31–45.
- [20] O. DEBARRE – *Higher-dimensional algebraic geometry*, Universitext, Springer-Verlag, New York, 2001.
- [21] ———, « Variétés rationnellement connexes », à paraître dans *Astérisque* (2003), Séminaire Bourbaki, Vol. 2001/02, exposé 905.
- [22] P. DELIGNE – « Théorie de Hodge II », *Publ. Math. Inst. Hautes Études Sci.* **40** (1972), p. 5–57.
- [23] ———, « La conjecture de Weil, I », *Publ. Math. Inst. Hautes Études Sci.* **43** (1974), p. 273–307.
- [24] B. DWORK – « On the rationality of the zeta function of an algebraic variety », *Amer. J. Math.* **82** (1960), p. 631–648.
- [25] T. EKEDAHL – « Sur le groupe fondamental d’une variété unirationnelle », *C. R. Acad. Sci. Paris Sér. I Math.* **297** (1983), no. 12, p. 627–629.
- [26] H. ESNAULT – « Varieties over a finite field with trivial Chow group of 0-cycles have a rational point », *Invent. Math.* **151** (2003), no. 1, p. 187–191, *arXiv:math.AG/0207022*.
- [27] J.-Y. ÉTESSE & B. LE STUM – « Fonctions L associées aux F -isocristaux surconvergens. I. Interprétation cohomologique », *Math. Ann.* **296** (1993), no. 3, p. 557–576.
- [28] ———, « Fonctions L associées aux F -isocristaux surconvergens. II. Zéros et pôles unités », *Invent. Math.* **127** (1997), no. 1, p. 1–31.

- [29] J.-M. FONTAINE & W. MESSING – « p -adic periods and p -adic étale cohomology », Contemporary mathematics, no. 67, p. 179–207, Contemporary mathematics **67**, Amer. Math. Soc., Arcata, 1985, 1987.
- [30] T. GRABER, J. HARRIS & J. STARR – « Families of rationally connected varieties », *J. Amer. Math. Soc.* **16** (2003), no. 1, p. 57–67.
- [31] E. GROSSE-KLÖNNE – « Finiteness of de Rham cohomology in rigid analysis », *Duke Math. J.* **113** (2002), no. 1, p. 57–91.
- [32] A. GROTHENDIECK – « Crystals and the de Rham cohomology of schemes », in *Dix exposés sur la cohomologie des schémas* [33], p. 306–358.
- [33] A. GROTHENDIECK et al. – *Dix exposés sur la cohomologie des schémas*, Adv. Stud. in pure Math., North-Holland, 1968.
- [34] R. HARTSHORNE – « On the De Rham cohomology of algebraic varieties », *Publ. Math. Inst. Hautes Études Sci.* (1975), no. 45, p. 5–99.
- [35] L. ILLUSIE – « Complexe de de Rham–Witt et cohomologie cristalline », *Ann. Sci. École Norm. Sup.* **12** (1979), p. 501–661.
- [36] A. J. DE JONG – « Smoothness, semistability and alterations », *Publ. Math. Inst. Hautes Études Sci.* **83** (1996), p. 51–93.
- [37] A. J. DE JONG & J. STARR – « Every rationally connected variety over the function field of a curve has a rational point », 2002, <http://www-math.mit.edu/~dejong/papers/familyofcurves3.dvi>.
- [38] B. KAHN – « Number of points of function fields over finite fields », 2002, *arXiv:math.NT/0210202*.
- [39] N. M. KATZ – « On a theorem of Ax », *Amer. J. Math.* **93** (1971), p. 485–499.
- [40] ———, « Le niveau de la cohomologie des intersections complètes », in *Groupes de monodromie en géométrie algébrique*, Lecture Notes in Math., no. 288-340, Springer Verlag, 1972-73, SGA 7, p. 363–399.
- [41] ———, « Slope filtration of F-crystals », in *Journées de Géométrie algébrique de Rennes*, Astérisque, no. 63, Soc. Math. France, 1979, p. 113–164.
- [42] N. M. KATZ & W. MESSING – « Some consequences of the Riemann hypothesis for varieties over finite fields », *Invent. Math.* **23** (1974), p. 73–77.
- [43] K. S. KEDLAYA – « Counting points on hyperelliptic curves using Monsky–Washnitzer cohomology », *J. Ramanujan Math. Soc.* **16** (2001), no. 4, p. 323–338.
- [44] ———, « Finiteness of rigid cohomology with coefficients », 2002, *arXiv:math.AG/0208027*.
- [45] ———, « Fourier transforms and p -adic Weil II », 2002, *arXiv:math.NT/0210149*.
- [46] M. KIM – « A vanishing theorem for Fano varieties in positive characteristic », 2002, *arXiv:math.AG/0201183*.
- [47] S. KLEIMAN – « Algebraic cycles and the Weil conjectures », in *Dix exposés sur la cohomologie des schémas* [33], p. 359–386.

- [48] J. KOLLÁR – « Shafarevich maps and plurigenera of algebraic varieties », *Invent. Math.* **113** (1993), p. 177–216.
- [49] G. LACHAUD & M. PERRET – « Un invariant birationnel des variétés de dimension 3 sur un corps fini », *J. Algebraic Geometry* **9** (2000), no. 3, p. 451–458.
- [50] A. G. B. LAUDER & D. WAN – « Computing zeta functions of Artin-Schreier curves over finite fields », *LMS J. Comput. Math.* **5** (2002), p. 34–55 (electronic).
- [51] YU. I. MANIN – « The theory of commutative formal groups over fields of finite characteristic », *Russian Math. Surveys* **18** (1963), p. 1–80.
- [52] ———, « Notes on the arithmetic of Fano threefolds », *Compositio Math.* **85** (1993), no. 1, p. 37–55.
- [53] B. MAZUR – « Frobenius and the Hodge filtration », *Bull. Amer. Math. Soc.* **78** (1972), p. 653–667.
- [54] ———, « Frobenius and the Hodge filtration (estimates) », *Ann. of Math.* **98** (1973), p. 58–95.
- [55] Z. MEBKHOUT – « Sur le théorème de finitude de la cohomologie p -adique d’une variété affine non singulière », *Amer. J. Math.* **119** (1997), no. 5, p. 1027–1081.
- [56] J. S. MILNE – « Zero cycles on algebraic varieties in nonzero characteristic : Roïtman’s theorem », *Compositio Math.* **47** (1982), no. 3, p. 271–287.
- [57] P. MONSKY & G. WASHNITZER – « Formal cohomology I », *Ann. of Math.* **88** (1968), p. 181–217.
- [58] O. MORENO & C. J. MORENO – « Improvements of the Chevalley-Waring and the Ax-Katz theorems », *Amer. J. Math.* **117** (1995), no. 1, p. 241–244.
- [59] N. NYGAARD – « On the fundamental group of a unirational 3-fold », *Invent. Math.* **44** (1978), no. 1, p. 75–86.
- [60] D. PETREQUIN – « Classes de Chern et classes de cycles en cohomologie rigide », *Bull. Soc. Math. France* **131** (2003), p. 59–121.
- [61] A. A. ROÏTMAN – « The torsion of the group of 0-cycles modulo rational equivalence », *Ann. of Math.* **111** (1980), no. 3, p. 553–569.
- [62] T. SATOH – « The canonical lift of an ordinary elliptic curve over a finite field and its point counting », *J. Ramanujan Math. Soc.* **15** (2000), no. 4, p. 247–270.
- [63] J.-P. SERRE – « On the fundamental group of a unirational variety », *J. London Math. Soc.* **34** (1959), p. 481–484.
- [64] N. I. SHEPHERD-BARRON – « Fano threefolds in positive characteristic », *Compositio Math.* **105** (1997), no. 3, p. 237–265.
- [65] T. SHIODA – « An example of unirational surface in characteristic p », *Math. Ann.* **211** (1974), p. 233–236.
- [66] N. TSUZUKI – « Cohomological descent of rigid cohomology for proper coverings », *Invent. Math.* **151** (2003), no. 1, p. 101–133.
- [67] D. Q. WAN – « A Chevalley-Waring approach to p -adic estimates of character sums », *Proc. Amer. Math. Soc.* **123** (1995), no. 1, p. 45–54.

- [68] E. WARNING – « Bemerkung zur vorstehenden Arbeit von Herr Chevalley », *Abh. Math. Sem. Univ. Hamburg* **11** (1936), p. 76–83.
- [69] A. WEIL – « Number of solutions of equations in finite fields », *Bull. Amer. Math. Soc.* **55** (1949), p. 397–508.

Antoine Chambert-Loir
Centre de mathématiques (CMAT)
École polytechnique
91128 Palaiseau Cedex
E-mail : chambert@math.polytechnique.fr

PROGRÈS RÉCENTS SUR L'HYPOTHÈSE DU CONTINU
[d'après Woodin]

par **Patrick DEHORNOY**

Une profusion de résultats conceptuellement profonds et techniquement difficiles ont été accumulés en théorie des ensembles depuis l'introduction des méthodes de forcing et de structure fine dans les années 1960. Ce rapport est consacré aux travaux récents de Woodin, qui non seulement ont constitué des percées techniques remarquables, mais ont aussi renouvelé le cadre conceptuel en améliorant l'intelligibilité globale de la théorie et en soulignant son unité profonde. Pour la première fois apparaissent une explication autre qu'empirique de la hiérarchie des grands cardinaux, et, surtout, une perspective réaliste de décider l'hypothèse du continu, en l'occurrence dans une direction négative :

CONJECTURE (Woodin, 1999). — Toute théorie des ensembles compatible avec l'existence de grands cardinaux et rendant invariantes par forcing les propriétés des ensembles héréditairement de cardinal au plus \aleph_1 implique que l'hypothèse du continu soit fausse.

Les travaux de Woodin arrivent très près de cette conjecture, qu'ils établissent pour une part substantielle de la hiérarchie des grands cardinaux. La question reste de savoir si cette part substantielle est en fait toute la hiérarchie des grands cardinaux. Dans tous les cas — et c'est ce qui légitime de parler de ces travaux maintenant, sans attendre une possible solution de la partie ouverte de la conjecture ci-dessus — les résultats de Woodin contribuent à montrer que le problème du continu et, plus généralement, la notion d'infini non dénombrable ne sont pas intrinsèquement vagues et inaccessibles à l'analyse, mais peuvent faire l'objet d'une véritable théorie conceptuelle allant bien au-delà de l'exploration formelle des conséquences d'axiomes plus ou moins arbitraires.

Le texte présent, qui doit beaucoup aux articles d'exposition [19, 20], s'efforce d'expliquer et de mettre en perspective les énoncés de quatre résultats de Woodin, apparaissant ici comme les théorèmes 5.8, 6.4, 6.7, et 7.2. Il semble hors de portée de donner une idée des démonstrations, dont la partie publiée occupe une bonne fraction des 900 pages de [18], et dont la partie la plus récente n'est décrite que dans [21].

Je remercie tous les théoriciens des ensembles qui m'ont apporté des commentaires et des suggestions, notamment Joan Bagaria, Matthew Foreman, Alexander Kechris, John Steel, Hugh Woodin, et, particulièrement, Stevo Todorćević.

1. UNE AFFAIRE TERMINÉE ?

L'hypothèse du continu (**HC**) est l'affirmation <Tout sous-ensemble infini de \mathbf{R} est en bijection soit avec \mathbf{N} , soit avec \mathbf{R} >, et le problème du continu est la question, soulevée par Cantor vers 1890, <L'hypothèse du continu est-elle vraie?>. Premier de la liste de Hilbert en 1900, le problème du continu a suscité des recherches tout au long du vingtième siècle. Une fois réuni un vaste consensus sur le système de Zermelo-Fraenkel (**ZF**, ou **ZFC** quand l'axiome du choix est inclus) comme point de départ axiomatique d'une théorie des ensembles, la *première* étape dans l'étude du problème du continu est la question <**HC**, ou sa négation $\neg\mathbf{HC}$, est-elle prouvable à partir de **ZFC**?>.

La réponse tient en deux résultats, tournants majeurs de la théorie des ensembles tant par leur importance propre que par les démonstrations qui ont permis de les établir :

THÉORÈME 1.1 (Gödel, 1938). — *Si **ZFC** est non contradictoire, il n'existe pas de preuve de $\neg\mathbf{HC}$ à partir de **ZFC**.*

THÉORÈME 1.2 (Cohen, 1963). — *Si **ZFC** est non contradictoire, il n'existe pas de preuve de **HC** à partir de **ZFC**.*

Il peut être tentant de retenir que le problème du continu ne peut être résolu et qu'à défaut d'être fermé, il est du moins sans intérêt, tout nouvel effort étant voué à l'échec. Cette conclusion est erronée. On peut certes juger le problème inopportun si on accorde peu d'intérêt aux objets qu'il met en jeu : sous-ensembles compliqués de \mathbf{R} , bons ordres dont l'existence relève de l'axiome du choix¹. Par contre, on doit voir que la question, si elle n'est pas écartée *a priori*, n'est pas fermée mais ouverte par les résultats de Gödel et Cohen : ainsi que le démontre le corpus accumulé, le système **ZFC** n'épuise pas notre intuition des ensembles, et la conclusion ne doit pas être que l'hypothèse du continu n'est ni vraie, ni fausse², mais, simplement, que le système **ZFC** est incomplet, et qu'il s'agit de le compléter.

Des analogies sont évidentes : que l'axiome des parallèles ne soit pas conséquence des autres axiomes d'Euclide n'a pas clos la géométrie, mais, au contraire, a permis l'émergence des géométries non euclidiennes, et a ouvert la question de reconnaître, parmi toutes les géométries possibles, la plus pertinente pour décrire le monde physique. De même, les résultats de Gödel et de Cohen montrent que plusieurs univers sont possibles à partir de **ZFC**, et ouvrent donc l'étude des divers univers possibles — c'est-à-dire, de façon équivalente, des divers systèmes axiomatiques complétant **ZFC** — et la question de reconnaître, parmi ceux-ci, le(s) plus pertinent(s) pour décrire le monde mathématique.

Diverses questions préliminaires se posent, de ce que peut être un bon axiome, et, surtout, de ce que peut signifier *résoudre* un problème tel que le problème du continu sur

¹En fait, il existe aussi des versions plus effectives de **HC** ne portant que sur des objets *définissables*.

²voire est indécidable en quelque sens mystérieux

la base d'axiomes additionnels³. On reviendra sur ces questions dans la section 2 à la lueur du cas de l'arithmétique. Divers axiomes susceptibles de compléter **ZFC** interviendront dans la suite de cet exposé. Pour le moment, mentionnons simplement les axiomes de grands cardinaux, qui, intuitivement, sont les plus naturels, et dont le rôle est central. Ces axiomes affirment l'existence d'infinis d'ordre supérieur, dépassant les infinis qui les précèdent à la façon dont l'infini dépasse le fini. Ils constituent une itération du principe de départ de la théorie des ensembles qui est précisément de postuler l'existence d'ensembles infinis⁴. L'une des raisons du succès des axiomes de grands cardinaux est leur efficacité pour décider un grand nombre d'énoncés non prouvables à partir de **ZFC**, cf. [9]. Le point important ici est qu'il semble raisonnable de tenir ces axiomes pour vrais, ou, au moins, de ne tenir pour plausibles que des axiomes **A compatibles** avec l'existence de grands cardinaux au sens où aucun axiome de grand cardinal ne contredit **A**.

2. ARITHMÉTIQUE, INCOMPLÉTUDE, ET FORCING

Notons V la collection de tous les ensembles⁵. De même que le but ultime de l'arithmétique serait de déterminer tous les énoncés satisfaits dans la structure $(\mathbb{N}, +, \times)$, celui de la théorie des ensembles serait de déterminer tous les énoncés satisfaits dans la structure (V, \in) . Ce but étant inaccessible, une possibilité est de se restreindre à des structures plus simples du type (H, \in) , où H est un certain fragment de la collection des ensembles. La filtration par la cardinalité est alors naturelle :

DÉFINITION 2.1. — *Pour k entier, on note H_k l'ensemble de tous les ensembles A héréditairement de cardinal strictement plus petit que \aleph_k , au sens où A , les éléments de A , les éléments des éléments de A , etc. sont tous de cardinal plus petit que \aleph_k ⁶.*

Considérons pour commencer la structure (H_0, \in) , c'est-à-dire le niveau des ensembles héréditairement finis. Notons **ZF^{fini}** le système **ZF** privé de l'axiome de l'infini.

³On se doute qu'ajouter simplement **HC** ou $\neg\text{HC}$ aux axiomes ne serait pas une très bonne solution !

⁴Comme l'existence d'un grand cardinal entraîne toujours la non-contradiction de **ZFC**, le second théorème d'incomplétude interdit qu'une telle existence puisse être démontrée à partir de **ZFC**, et la poser comme hypothèse constitue donc toujours un axiome propre.

⁵en fait, la collection de tous les ensembles *purs*, définis comme ceux pouvant être obtenus à partir de l'ensemble vide en itérant les opérations de passage à l'ensemble des parties, à la réunion, et aux éléments. On sait que de tels ensembles suffisent à représenter tous les objets mathématiques.

⁶On rappelle que $\aleph_0, \aleph_1, \dots$ est l'énumération croissante des cardinaux infinis, ceux-ci étant définis comme les ordinaux infinis qui ne sont en bijection avec aucun ordinal plus petit. Ainsi \aleph_0 (aussi noté ω), est le plus petit ordinal infini, donc aussi la limite supérieure des ordinaux finis, et \aleph_1 est le plus petit ordinal non dénombrable, donc la limite supérieure des ordinaux dénombrables. Alors \aleph_0 est le cardinal de \mathbb{N} , et, en notant 2^κ le cardinal de $\mathcal{P}(\kappa)$ comme dans le cas fini, 2^{\aleph_0} est celui de $\mathcal{P}(\mathbb{N})$, donc aussi de \mathbb{R} , de sorte que l'hypothèse du continu s'écrit $2^{\aleph_0} = \aleph_1$.

LEMME 2.2. — À partir des axiomes de $\mathbf{ZF}^{\text{fini}}$, on peut définir à l'intérieur de (H_0, \in) une copie de $(\mathbb{N}, +, \times)$; inversement, à partir des axiomes de Peano, on peut définir à l'intérieur de $(\mathbb{N}, +, \times)$ une copie de (H_0, \in) ⁷.

À un codage près, décrire (H_0, \in) équivaut donc à décrire $(\mathbb{N}, +, \times)$: le niveau <héréditairement fini> de la théorie des ensembles coïncide avec l'arithmétique. Une façon usuelle de décrire une structure S consiste à l'*axiomatiser*, c'est-à-dire à caractériser les énoncés satisfaits dans S comme ceux qui sont *prouvables* à partir d'un système d'axiomes suffisamment simple. Pour l'arithmétique, le système de Peano est bien connu, mais les théorèmes d'incomplétude de Gödel montrent que la description obtenue n'est pas complète : il existe des énoncés satisfaits dans $(\mathbb{N}, +, \times)$ mais non prouvables à partir des axiomes de Peano, et, de même, des énoncés satisfaits dans (H_0, \in) mais non prouvables à partir de $\mathbf{ZF}^{\text{fini}}$.

Se placer dans le cadre de la théorie des ensembles permet de démontrer davantage d'énoncés, donc de se rapprocher d'une description complète. Dans le cas d'un énoncé φ portant sur H_0 , cela signifie non plus chercher si φ est prouvable à partir de $\mathbf{ZF}^{\text{fini}}$, mais si l'énoncé $\langle (H_0, \in) \text{ satisfait } \varphi \rangle$ ⁸ est prouvable à partir de \mathbf{ZFC} . De même, dans le cas d'un énoncé φ portant sur \mathbb{N} , il s'agit, au lieu de chercher si φ est prouvable à partir des axiomes de Peano, de chercher si $\langle (\mathbb{N}, +, \times) \text{ satisfait } \varphi \rangle$ ⁹ est prouvable à partir de \mathbf{ZFC} .

Le théorème d'incomplétude s'applique derechef, et l'axiomatisation par \mathbf{ZFC} ne donne toujours pas une description complète. Pour autant, cette description est en pratique satisfaisante, en ce que la plupart des exemples d'énoncés vrais mais non prouvables sont des énoncés *ad hoc* plus ou moins directement issus de la logique¹⁰. De plus, et surtout, les manifestations de l'incomplétude de \mathbf{ZFC} au niveau de l'arithmétique diffèrent fondamentalement de ses manifestations à des niveaux ultérieurs, par exemple dans le cas de l'hypothèse du continu.

Expliquer cette différence de nature requiert d'introduire la notion de *forcing* et, d'abord, celle de *modèle de ZFC*. Comme \mathbf{ZFC} est une famille d'axiomes portant sur l'unique relation d'appartenance, on peut considérer de façon abstraite des structures (M, E) avec E relation binaire sur M telles que chacun des axiomes de \mathbf{ZFC} soit satisfait lorsque E est

⁷On obtient une copie $\underline{\mathbb{N}}$ de \mathbb{N} à l'intérieur de H_0 en définissant récursivement une copie \underline{i} de l'entier i par $\underline{0} = ?$ et $\underline{i+1} = \underline{i} \cup \{\underline{i}\}$: c'est la représentation de von Neuman des entiers par des ensembles. Il est alors facile de construire des copies $\underline{+}$ et $\underline{\times}$ de $+$ et \times , et de montrer à partir de $\mathbf{ZF}^{\text{fini}}$ que $(\underline{\mathbb{N}}, \underline{+}, \underline{\times})$ satisfait aux axiomes de Peano. À l'inverse, suivant Ackermann, on définit à l'intérieur de $(\mathbb{N}, +, \times)$ une relation $\underline{\in}$ en déclarant $p \underline{\in} q$ vraie si le p -ième chiffre du développement binaire de q est 1, et on montre à partir des axiomes de Peano que $(\mathbb{N}, \underline{\in})$ satisfait aux axiomes de $\mathbf{ZF}^{\text{fini}}$, et est isomorphe à (H_0, \in) .

⁸c'est-à-dire l'énoncé obtenu à partir de φ en ajoutant que toutes les variables prennent leurs valeurs dans l'ensemble définissable H_0 , cf. note 14

⁹ou, plus exactement et avec les notations de la note 7, $\langle (\underline{\mathbb{N}}, \underline{+}, \underline{\times}) \text{ satisfait } \varphi \rangle$

¹⁰Noter néanmoins les propriétés combinatoires isolées par H. Friedman [6], ou les résultats de Y. Matiyasevich sur l'existence d'équations diophantiennes dont la résolubilité est indémontrable [12].

prise comme valeur de l'appartenance : une telle structure est appelée *modèle* de **ZFC**. La validité des axiomes de **ZFC** s'exprime alors par le fait que la structure (V, \in) constituée des *vrais* ensembles et de la *vraie* appartenance est un modèle de **ZFC**¹¹.

C'est le cadre conceptuel fourni par la notion de modèle de **ZFC** qui permet d'établir les théorèmes 1.1 et 1.2 : pour montrer que $\neg\mathbf{HC}$ (*resp.* **HC**) n'est pas prouvable à partir de **ZFC**, il suffit de construire un modèle de **ZFC** satisfaisant **HC** (*resp.* $\neg\mathbf{HC}$), ce qu'on fait en partant d'un modèle (quelconque) M de **ZFC**, et en en construisant respectivement un sous-modèle L satisfaisant **HC** avec Gödel, et une extension $M[G]$ satisfaisant $\neg\mathbf{HC}$ avec Cohen¹². La méthode de Cohen, ou méthode du *forcing*, consiste à ajouter à M un ensemble G dont les propriétés sont définies (<forcées>) depuis l'intérieur de M par un ensemble ordonné \mathbf{P} , dit de *forcing*, qui décrit les éléments de $M[G]$ ¹³. Un modèle du type $M[G]$ est appelé extension *générique* de M .

L'existence du forcing introduit une variabilité essentielle dans la théorie des ensembles. Étant donné un modèle M , et un énoncé φ tel que ni φ , ni $\neg\varphi$ ne soient prouvables à partir de **ZFC**, il est fréquent qu'on puisse construire, à l'aide d'un premier ensemble de forcing \mathbf{P}_1 , une extension générique $M[G_1]$ dans laquelle φ est satisfait, et, à l'aide d'un second ensemble de forcing \mathbf{P}_2 , une autre extension générique $M[G_2]$ dans laquelle $\neg\varphi$ est satisfait, de sorte que privilégier φ ou $\neg\varphi$ semble difficile. Au demeurant, nous allons voir que cette situation ne peut pas se produire au niveau de l'arithmétique.

DÉFINITION 2.3. — *Soit H un ensemble définissable¹⁴. On dit que les propriétés de la structure (H, \in) sont invariantes par forcing si, quels que soient l'énoncé φ , le modèle M , et l'extension générique $M[G]$ de M , l'énoncé $\langle(H, \in) \text{ satisfait } \varphi\rangle$ est satisfait dans M si et seulement si il l'est dans $M[G]$ ¹⁵.*

PROPOSITION 2.4 (Shoenfield). — *Les propriétés de (H_0, \in) et de $(\mathbf{N}, +, \times)$ sont invariantes par forcing¹⁶.*

¹¹Cette description adopte un vocabulaire délibérément platonicien référant à un vrai monde de vrais ensembles : plus que d'une option philosophique, il s'agit d'une commodité de présentation, consistant à fixer un modèle de référence et à distinguer les modèles partageant la même relation d'appartenance.

¹²De même, pour montrer que les axiomes des groupes n'entraînent pas, disons, la commutativité, on pourrait partir d'un groupe G quelconque, et en construire respectivement un sous-groupe commutatif et une extension non commutative. La construction est plus délicate dans le second cas, car, rien n'excluant de partir avec $G = \{1\}$ ou $M = L$, le passage à une sous-structure ne saurait suffire.

¹³comme une extension de corps dont les éléments sont décrits par des polynômes du corps de base

¹⁴c'est-à-dire que H est *défini* comme l'ensemble des x vérifiant une certaine formule $\psi(x)$ du langage de la théorie des ensembles. Par exemple, chacun des ensembles H_k est définissable.

¹⁵Il y a un aspect subtil ici : à supposer que H soit défini par une formule $\psi(x)$, on ne demande pas que les ensembles définis par $\psi(x)$ dans M et dans $M[G]$, c'est-à-dire $\langle H \text{ calculé dans } M \rangle$ et $\langle H \text{ calculé dans } M[G] \rangle$, coïncident, on demande seulement que ces deux ensembles aient les mêmes propriétés.

¹⁶Le résultat ici est même plus fort : on a invariance non seulement par passage à une extension générique, mais aussi par passage à une extension quelconque, (M', E') étant appelé *extension* de (M, E) si M est inclus dans M' , E est la restriction de E' à M , et les ordinaux de (M, E) et (M', E') coïncident.

Les manifestations de l'incomplétude de **ZFC** au niveau de l'arithmétique ne sont donc pas liées à la variabilité due au forcing, et elle se réduisent à ce qu'on pourrait appeler une incomplétude résiduelle, dont on a dit qu'elle limite peu l'efficacité en pratique. Il est donc naturel de chercher à retrouver, par exemple pour les structures (H_k, \in) avec $k \geq 1$, la situation de (H_0, \in) et de l'arithmétique — si cela se peut. Ceci conduit à poser comme suit la question de la recherche de bons axiomes pour une structure (H, \in) :

PROBLÈME 2.5. — *Trouver un cadre axiomatique, **ZFC** ou **ZFC** complété d'axiome(s) compatible(s) avec l'existence de grands cardinaux, fournissant une description suffisamment complète de (H, \in) et en rendant les propriétés invariantes par forcing.*

Obtenir l'invariance des propriétés de (H, \in) par forcing, c'est neutraliser l'action du forcing au niveau de H , afin de limiter autant que faire se peut l'inévitable incomplétude de la description. L'invariance des propriétés par forcing est une contrainte forte, et il n'est pas clair *a priori* qu'elle puisse être réalisée au delà de H_0 ¹⁷. Par contre, l'éventuelle satisfaction de cette contrainte devrait apparaître comme un argument de poids en faveur du système axiomatique qui l'accomplit¹⁸ : ce sera notre point de vue ici.

Dans ce contexte, on peut maintenant proposer une réponse à la question : <Que peut signifier établir φ lorsque ni φ , ni sa négation $\neg\varphi$ ne sont prouvables à partir de **ZFC**?>, typiquement lorsqu'on peut, par forcing, réaliser tout aussi bien φ que $\neg\varphi$. Supposons que φ porte sur une certaine structure définissable (H, \in) . Si on accepte le cadre du problème 2.5, c'est-à-dire si on privilégie le critère d'invariance par forcing, il devrait apparaître raisonnable de considérer φ comme établi si on a montré deux choses :

- (i) qu'il existe bien *au moins une* solution au problème 2.5 pour (H, \in) , c'est-à-dire au moins un système axiomatique en satisfaisant toutes les requêtes¹⁹, et, d'autre part,
- (ii) que *toute* telle solution implique que φ soit vrai.

Autrement dit, il s'agit de considérer comme établi un énoncé φ qui est nécessairement vrai dans tout contexte cohérent neutralisant l'action du forcing jusqu'au niveau de φ ²⁰.

C'est en ces termes que le problème du continu, dont on verra qu'il s'exprime comme une propriété de H_2 , est abordé par Woodin.

¹⁷De fait, Woodin a montré qu'aucun résultat d'invariance par forcing n'est possible pour les propriétés de la structure (V_{\aleph_0+2}, \in) (cf. note 28), donc, essentiellement, pour tout fragment contenant $\mathcal{P}(\mathbb{R})$.

¹⁸Il n'existe cependant pas d'unanimité. Une objection est de considérer la variabilité due au forcing comme un flou dans notre perception des ensembles. De ce point de vue, réclamer l'invariance des propriétés par forcing, c'est restreindre l'observation à des fragments de l'univers qui échappent au flou. Mais rien ne dit que la solution du problème du continu, par exemple, doive se trouver dans ces fragments-là.

¹⁹Ceci, afin de montrer que l'approche n'est pas irréaliste.

²⁰En termes encore plus imagés : seul reste φ lorsque la température a été suffisamment abaissée pour que l'agitation thermique liée au forcing cesse de rendre φ et $\neg\varphi$ indistinguables...

3. LE SECOND NIVEAU : LE DÉNOMBRABLE

La situation simple de H_0 ne se retrouve pas dès que des ensembles infinis entrent en jeu : le système **ZFC** ne rend pas les propriétés de H_1 invariantes par forcing, et laisse ouvertes beaucoup d'entre elles parmi les plus naturelles. Mais on va voir qu'à condition d'amender convenablement **ZFC**, il existe une excellente solution au problème 2.5 pour (H_1, \in) .

De la même façon qu'on passe de (H_0, \in) à $(\mathbf{N}, +, \times)$, il est facile de passer de (H_1, \in) à $(\sum(\mathbf{N}), \mathbf{N}, +, \times, \in)$ ²¹. Déterminer les énoncés satisfaits dans $(\sum(\mathbf{N}), \mathbf{N}, +, \times, \in)$ revient à étudier les sous-ensembles de $\sum(\mathbf{N})$ qui y sont *définissables*, c'est-à-dire ont la forme

$$(3.1) \quad A = \{x \in \sum(\mathbf{N}) ; (\sum(\mathbf{N}), \mathbf{N}, +, \times, \in) \text{ satisfait } \varphi(x, \vec{a})\},$$

avec \vec{a} suite finie d'éléments de $\sum(\mathbf{N})$: typiquement, reconnaître si l'énoncé $\exists x \varphi(x, \vec{a})$ est satisfait équivaut à reconnaître si l'ensemble défini par (3.1) est non vide.

DÉFINITION 3.1 (Lusin). — *Soit X un espace Polonais. Un sous-ensemble de X^p est dit projectif s'il peut être obtenu à partir d'un borélien de X^{p+k} par un nombre fini de projections et de passages au complémentaire.*

Les sous-ensembles de $\sum(\mathbf{N})$ définissables dans $(\sum(\mathbf{N}), \mathbf{N}, +, \times, \in)$ sont exactement²² les sous-ensembles projectifs de l'espace de Cantor $\{0, 1\}^{\aleph}$. Comme les classes d'ensembles considérées incluent les boréliens, l'existence d'un isomorphisme borélien entre $\{0, 1\}^{\aleph}$ et la droite réelle permet de remplacer $\sum(\mathbf{N})$ par \mathbf{R} , et on conclut que l'étude de (H_1, \in) est essentiellement celle des sous-ensembles projectifs de \mathbf{R} .

Dans l'optique du problème 2.5, le premier point est de savoir si l'axiomatisation par **ZFC** fournit une description suffisamment complète de (H_1, \in) , donc des sous-ensembles projectifs de \mathbf{R} . Ce n'est pas le cas : si nous appelons PCA un ensemble (projectif) qui est projection de complémentaire de projection de borélien, alors, si **ZFC** est non contradictoire, ni l'assertion <Tous les sous-ensembles PCA de \mathbf{R} sont Lebesgue-mesurables>, ni sa négation, ne sont prouvables à partir de **ZFC**²³.

Il s'agit alors de chercher si l'adjonction d'un nouvel axiome permettrait d'obtenir une solution au problème 2.5. Une longue accumulation de résultats profonds mène à la conclusion que l'*axiome de détermination projective* est un tel axiome.

²¹c'est-à-dire à l'*arithmétique du second ordre*, où sont considérés, outre les entiers, les ensembles d'entiers et l'appartenance associée

²²Si φ est sans quantificateur, alors (3.1) définit un ouvert ; ajouter une quantification existentielle revient à effectuer une projection, tandis qu'ajouter une négation revient à passer au complémentaire.

²³Comme évoqué plus haut, Gödel démontre que $\neg \mathbf{HC}$ (de même que la négation de l'axiome du choix) n'est pas prouvable à partir de **ZF** en construisant un certain sous-modèle L de V . Or L est équipé d'un bon ordre canonique (impliquant l'axiome du choix) dont la restriction aux réels est un ensemble PCA, qui, par le théorème de Fubini, ne peut être mesurable. Donc L satisfait <il existe un PCA non mesurable>, et, par conséquent, il est impossible que **ZFC** prouve <tous les PCA sont mesurables>. Pour la négation, on utilise le forcing avec l'axiome de Martin \mathbf{MA} , dont il sera question dans la section 4.

DÉFINITION 3.2. — On dit qu'un sous-ensemble A de $[0, 1]$ est déterminé si l'énoncé infini suivant, où les ε_i valent 0 ou 1, est satisfait :

$$(\exists \varepsilon_1)(\forall \varepsilon_2)(\exists \varepsilon_3) \dots (\sum_i \varepsilon_i 2^{-i} \in A) \text{ ou } (\forall \varepsilon_1)(\exists \varepsilon_2)(\forall \varepsilon_3) \dots (\sum_i \varepsilon_i 2^{-i} \notin A)^{24}.$$

Tous les ouverts sont déterminés, et un théorème de Martin (1975) affirme qu'il en est de même de tous les boréliens. Ce résultat est le plus fort possible dans **ZFC**²⁵, et, par conséquent, poser comme hypothèse que les ensembles d'une famille au-delà des boréliens sont déterminés constitue un axiome (propre) par rapport à **ZFC**.

DÉFINITION 3.3. — On note **DP** (détermination projective) l'axiome : \langle Tout ensemble projectif est déterminé \rangle ²⁶.

La propriété de détermination est un paradigme permettant d'exprimer de nombreuses propriétés d'analyse, et il en résulte que l'axiome **DP** fournit une description très complète des propriétés des ensembles projectifs. Des résultats typiques sont les suivants, cf. [14] :

THÉORÈME 3.4 (Banach–Mazur, Mycielski–Swierczkowski, Moschovakis). — *Le système **ZFC** + **DP** prouve que tous les ensembles projectifs sont Lebesgue-mesurables, ont la propriété de Baire, et ont la propriété d'uniformisation*²⁷.

Les phénomènes d'incomplétude liés au théorème de Gödel restent inévitables, mais on peut affirmer sans tricher que l'axiomatisation de H_1 par **ZFC** + **DP** a la même efficacité pratique que celle de H_0 par **ZFC**.

Dans l'optique du problème 2.5, le point suivant est la recherche de conditions éventuelles rendant les propriétés de H_1 invariantes par forcing. C'est précisément l'étude de ce point qui a conduit à isoler la notion de *cardinal de Woodin*²⁸ et a mené en 1984, à partir du travail de Foreman, Magidor, et Shelah mentionné dans la section 4, au résultat suivant :

²⁴De façon équivalente, un des joueurs a une *stratégie gagnante* dans le jeu G_A où deux joueurs I et II construisent à tour de rôle une suite infinie $\varepsilon_1, \varepsilon_2, \dots$ de 0 et de 1, et où I (*resp.* II) est déclaré gagnant pour $\sum_i \varepsilon_i 2^{-i} \in A$ (*resp.* \notin).

²⁵Dans le modèle L , il existe un ensemble non déterminé qui est projection de borélien.

²⁶Notons **AD** l'axiome maximal \langle tout ensemble de réels est déterminé \rangle ; Woodin a démontré en 1987 que les systèmes **ZFC** + \langle il existe une infinité de cardinaux de Woodin \rangle (voir note 28) et **ZF** + **AD** sont équiconsistants; s'il existe une infinité de cardinaux de Woodin plus un cardinal mesurable au-dessus d'eux dans V , alors **AD** est satisfait dans le sous-modèle minimal $L(\mathbf{R})$ de V contenant \mathbf{R} .

²⁷C'est-à-dire : Si A est un sous-ensemble projectif de \mathbf{R}^2 , alors il existe une fonction $f : \mathbf{R} \rightarrow \mathbf{R}$ de graphe projectif choisissant, pour tout x tel qu'il en existe, un élément y vérifiant $(x, y) \in A$; la propriété d'uniformisation n'est qu'un corollaire d'une certaine propriété d'*échelle* qui est cruciale.

²⁸Comme de nombreux autres grands cardinaux, les cardinaux de Woodin se définissent par l'existence de plongements élémentaires, qui sont les homomorphismes entre modèles de **ZFC** préservant tout ce qui est définissable à partir de \in . Notons V_α l'ensemble des ensembles purs s'obtenant à partir de $\mathbf{?}$ en utilisant au plus α fois l'opération \mathcal{P} . Alors, un cardinal κ est *de Woodin* (dans V) si, pour toute fonction $f : \kappa \rightarrow \kappa$, il existe un sous-modèle M de V , un plongement élémentaire $j : V \rightarrow M$, et un ordinal $\alpha < \kappa$ tels que $\xi < \alpha$ entraîne $j(\xi) = \xi$ et $f(\xi) < \alpha$ et qu'on ait $j(\alpha) > \alpha$ et $V_{j(f)(\alpha)} \subseteq M$. Le principe est

THÉORÈME 3.5 (Woodin). — *Supposons qu'il existe une classe propre de cardinaux de Woodin*²⁹. *Alors les propriétés de (H_1, \in) sont invariantes par forcing.*

Ne reste alors qu'à établir la compatibilité de l'axiome **DP** avec l'existence de grands cardinaux — c'est-à-dire, en un sens, à prouver l'axiome **DP**. Woodin en a donné vers 1983 une démonstration à partir d'un axiome extrêmement fort, mais la mesure exacte de **DP** dans la hiérarchie des axiomes de grands cardinaux est venue en 1985, avec le remarquable résultat suivant, cf. [2] :

THÉORÈME 3.6 (Martin–Steel). — *Supposons qu'il existe une infinité de cardinaux de Woodin. Alors **DP** est vrai*³⁰.

Ainsi, le système formé par **ZFC** + **DP** fournit une bonne description de (H_1, \in) ³¹, et le système légèrement plus fort formé par **ZFC** plus l'existence d'une classe propre de cardinaux de Woodin (une sorte de version délocalisée de **DP**) constitue une solution complète au problème 2.5 pour H_1 : il permet de recouvrer pour H_1 , c'est-à-dire pour le niveau de l'infini dénombrable, le même type de complétude empirique que **ZFC** garantit pour H_0 , c'est-à-dire pour le niveau des ensembles finis et de l'arithmétique.

naturel : un cardinal est infini s'il possède une partie propre qui est en bijection avec lui, et il est <super-infini> s'il possède une partie propre qui non seulement est en bijection avec lui, mais lui est isomorphe, ce qu'exprime ici le plongement élémentaire. La force de l'hypothèse est proportionnelle au degré de ressemblance requis entre le modèle de départ V et le modèle d'arrivée M , ici l'hypothèse $V_{j(f)(\alpha)} \subseteq M$. Si cette dernière est affaiblie en $V_\kappa \subseteq M$, on obtient essentiellement la définition d'un cardinal *mesurable*, une notion plus faible que celle de cardinal de Woodin en termes de consistance.

²⁹c'est-à-dire : Pour chaque cardinal κ , il existe un cardinal de Woodin au-dessus de κ .

³⁰Il est apparu ensuite que cette implication est, au niveau de la force logique, quasiment une équivalence : pour tout entier k , le système **ZFC** + **DP** prouve la consistance (c'est-à-dire la non-contradiction) du système **ZFC** + <il existe k cardinaux de Woodin> ; encore d'autres liens sont apparus, par exemple que l'axiome de forcing **MM** (voir section 4) entraîne **DP** ; tout ceci illustre et explique l'ubiquité de **DP** et des cardinaux de Woodin dans la théorie récente.

³¹Un argument supplémentaire en faveur de **DP** est que non seulement cet axiome apporte des réponses sur H_1 , mais qu'en outre il apporte *les* réponses heuristiquement satisfaisantes : ainsi, l'uniformisation permet d'éviter tout recours à l'axiome du choix dans l'étude des ensembles projectifs ; de même, la mesurabilité interdit l'existence de décompositions paradoxales de la sphère en pièces projectives. À l'opposé, l'axiomatisation par **ZFC** + **v=L** (la <structure fine> de Jensen) exprimant que l'univers coïncide avec le modèle minimal L de Gödel, fournit aussi une description assez complète (au demeurant incompatible avec l'existence de grands cardinaux et ne résistant pas au forcing) mais dont les réponses sont moins satisfaisantes que celles fournies par **DP** : par exemple, on a vu que **ZFC** + **v=L** prouve l'existence d'ensembles projectifs non mesurables.

Un autre point distinguant **DP** de **v=L** est que tout modèle de **ZFC** inclut L comme sous-modèle, et, donc, une théorie de L est toujours présente comme sous-théorie. Adopter **DP** ne rejette donc pas **v=L**, présente comme sous-théorie, alors qu'adopter **v=L** restreindrait l'étude, comme restreindre l'étude des groupes à celle des groupes commutatifs. C'est cet argument qui conduit à requérir que tout axiome pour V , donc pour les vrais ensembles, soit compatible avec les grands cardinaux.

4. LE TROISIÈME NIVEAU : LA CARDINALITÉ \aleph_1

L'étape suivante est celle de la structure (H_2, \in) , c'est-à-dire celle de la cardinalité \aleph_1 . Au même sens que ci-dessus, H_2 est le niveau de $\sum(\aleph_1)$ ³², ou encore celui de l'ensemble des suites de réels de longueur \aleph_1 . Un rôle technique important est joué par les sous-ensembles *stationnaires* de \aleph_1 , qui n'ont pas d'équivalent au niveau de \aleph_0 , et qui proviennent de l'existence de points-limites dans la topologie de l'ordre sur l'ordinal \aleph_1 .

DÉFINITION 4.1. — *Un sous-ensemble de \aleph_1 est dit stationnaire s'il rencontre tout sous-ensemble non borné de \aleph_1 fermé pour la topologie de l'ordre ; on note \mathcal{I}_{NS} l'ensemble des parties non stationnaires de \aleph_1 .*

Le niveau de H_2 est le premier où l'axiome du choix commence à se manifester de façon essentielle, et, notons-le tout de suite, celui où le problème du continu se pose³³ :

LEMME 4.2. — Il existe un énoncé φ_{HC} tel que $\langle(H_2, \in)$ satisfait φ_{HC} équivaille à **HC**.

Le succès rencontré pour H_1 conduit à chercher un axiome³⁴ jouant pour H_2 le rôle joué pour H_1 par **DP** et par l'existence d'une classe propre de cardinaux de Woodin, c'est-à-dire résolvant le problème 2.5. La mauvaise nouvelle, annoncée par Levy et Solovay dès 1967, est qu'aucun axiome de grand cardinal ne peut convenir :

PROPOSITION 4.3. — *Aucun axiome de grand cardinal ne peut rendre les propriétés de (H_2, \in) invariantes par forcing.*

Tout vient de la définissabilité du forcing à partir d'un ensemble du modèle de base. D'après le lemme 4.2, il suffit de montrer qu'un axiome de grand cardinal **A** ne peut pas imposer la valeur de **HC**. Or, pour briser **HC**, il suffit d'ajouter \aleph_2 sous-ensembles à \aleph_1 , ce qui peut se faire avec un ensemble de forcing de cardinal \aleph_2 ; par construction, un tel *petit* forcing préserve les grands cardinaux. Donc, partant d'un modèle satisfaisant **A** + **HC**, on peut toujours en construire une extension générique satisfaisant **A** + \neg **HC**, et **A** ne peut rendre **HC** invariante par forcing.

Une éventuelle axiomatisation de (H_2, \in) neutralisant l'action du forcing n'est donc pas à chercher parmi les axiomes de grands cardinaux. Depuis les années 1980, il est apparu que des candidats naturels se trouvent dans la famille des *axiomes de forcing*³⁵. Ces axiomes sont des formes fortes du théorème de catégorie de Baire qui affirme que,

³²c'est-à-dire celui des ensembles d'ordinaux dénombrables, la construction des ordinaux étant faite de sorte que chaque ordinal coïncide avec l'ensemble des ordinaux plus petits : par exemple, \aleph_1 est l'ensemble des ordinaux dénombrables.

³³Ceci n'est pas évident, puisqu'*a priori* **HC** met en jeu tout l'ensemble $\mathcal{P}(\mathbb{R})$, lequel n'appartient à H_2 que si **HC** est vraie ; le lemme indique qu'on peut toujours *coder* l'hypothèse du continu dans (H_2, \in) .

³⁴Une fois pour toutes, nous parlons ici d'un axiome, mais il pourrait tout autant s'agir d'une famille finie d'axiomes, ou d'une famille infinie pourvu qu'elle soit suffisamment explicite, typiquement récursive.

³⁵D'autres candidats pourraient être les axiomes de *grand cardinal générique* de [4].

si X est un espace localement compact, alors toute intersection dénombrable d'ouverts denses de X est dense. La possibilité d'étendre le résultat aux intersections de cardinal \aleph_1 dépend de la cardinalité de X ³⁶, et elle exige que certaines contraintes sur X , et plus précisément sur l'algèbre des ouverts réguliers de X , soient satisfaites. Ces contraintes sont motivées par la théorie du forcing, d'où le nom d'«axiomes de forcing».

Le premier axiome de forcing a été introduit en 1970 pour résoudre le fameux problème de Souslin sur la caractérisation de la droite réelle. C'est l'*axiome de Martin* **MA** : «Si X est un espace localement compact où toute famille d'ouverts deux à deux disjoints est au plus dénombrable, alors toute intersection de \aleph_1 ouverts denses de X est dense». Cet axiome exprime une forme faible d'invariance des propriétés entre V et les extensions génériques associées à un ensemble de forcing suffisamment petit.

La théorie du *forcing itéré* a conduit à l'introduction de toute une hiérarchie d'extensions de **MA**, cf. [15] (et son millier de pages!). Dans [5], Foreman, Magidor et Shelah ont identifié la classe maximale d'espaces localement compacts pour laquelle une forme forte du théorème de Baire n'est pas *a priori* contradictoire, à savoir les espaces X pour lesquels l'algèbre de Boole des ouverts réguliers de X *préserve la stationnarité*³⁷. La forme la plus forte possible de l'axiome de Martin est donc :

DÉFINITION 4.4 (Foreman–Magidor–Shelah). — *L'axiome de Martin maximum **MM** est l'assertion : «Si X est un espace localement compact dont l'algèbre des ouverts réguliers préserve la stationnarité, alors toute intersection de \aleph_1 ouverts denses de X est dense».*

Comme avec **DP**, la question se pose de la compatibilité de **MM** avec l'existence de grands cardinaux. La réponse est positive [5], et repose sur le forcing itéré de Shelah [15] et sur un argument de Baumgartner reliant celui-ci aux cardinaux supercompacts³⁸ :

THÉORÈME 4.5 (Foreman, Magidor, Shelah). — *Supposons qu'il existe un cardinal supercompact. Alors l'axiome **MM** est satisfait dans une extension générique de V .*

Pour l'étude de H_2 , il est naturel de considérer une variante faible de l'axiome **MM** appelée **MMB** (*Martin maximum borné*)³⁹, introduite par Goldstern et Shelah [8]. Appelons *bornée* toute formule ne contenant que des quantifications $\forall y \in z$ et $\exists y \in z$. L'intérêt de **MMB** pour H_2 apparaît dans une reformulation due à Bagaria [1], à savoir que **MMB** est équivalent à l'assertion «Tout énoncé $\exists x \psi(x, a)$ avec ψ bornée et a dans H_2 satisfait dans

³⁶Si **HC** est vraie, alors l'intersection des \aleph_1 ouverts denses $\mathbf{R} \setminus \{a\}$ pour a dans \mathbf{R} est vide.

³⁷On dit qu'un ensemble ordonné **P** *préserve la stationnarité* si tout sous-ensemble stationnaire de \aleph_1 dans V reste stationnaire dans l'extension générique associée à **P**.

³⁸Un cardinal κ est dit *supercompact* si, pour tout cardinal $\lambda > \kappa$, il existe une classe M et un plongement élémentaire $j : V \rightarrow M$ vérifiant $j(\kappa) > \lambda$ et tels que toute suite d'éléments de M de longueur λ (dans V) soit dans M . L'axiome «il existe un cardinal supercompact» est plus fort que l'axiome «il existe une infinité de cardinaux de Woodin», et donc que **DP**.

³⁹L'énoncé est le même que celui de **MM**, à ceci près qu'on se restreint aux intersections d'ouverts denses qui sont unions d'au plus \aleph_1 ouverts réguliers.

une extension générique de V préservant la stationnarité est déjà satisfait dans H_2 ⁴⁰. Ainsi l'axiome **MMB** prouve toute propriété de H_2 pouvant être exprimée par un énoncé du type $\forall \dots \exists \dots \psi$ avec ψ bornée⁴¹ et ne pouvant être mise en défaut par un forcing préservant la stationnarité. Ces propriétés sont donc invariantes par forcing préservant la stationnarité, et la question suivante est donc naturelle :

*L'axiome **MMB** — ou une variante de celui-ci — résout-il le problème 2.5 pour H_2 ?*

La question reste ouverte, car il manque une complétude non restreinte aux énoncés de type $\forall \exists$ et une invariance par forcing non conditionnelle.

C'est de ce point que partent les travaux de Woodin. L'idée est d'obtenir d'abord l'invariance par forcing, en partant d'une version du théorème 3.5 affirmant que, s'il existe une classe propre de cardinaux de Woodin, alors les propriétés de la structure $(L(\mathbf{R}), \in)$ sont invariantes par forcing. À partir de là, Woodin cherche essentiellement à réaliser l'axiome **MMB** dans une extension générique convenable du modèle $L(\mathbf{R})$. La construction met en jeu un certain forcing \mathbf{P}_{max} d'un type nouveau et est très sophistiquée, les éléments de l'ensemble \mathbf{P}_{max} étant eux-mêmes des modèles de **ZFC**. L'argument aboutit à l'extension de l'invariance par forcing des propriétés de $L(\mathbf{R})$ à celles du modèle construit. Ce dernier est $L(\sum(\aleph_1))$, et il inclut donc H_2 par construction. Il s'ensuit que les propriétés de H_2 sont capturées. Le résultat final met en jeu un nouvel axiome, noté ici **MMW** pour <Martin maximum de Woodin>⁴², qui est la variante de (la reformulation par Bagaria de) l'axiome **MMB** dans laquelle les sous-ensembles stationnaires de \aleph_1 et un sous-ensemble de \mathbf{R} appartenant à $L(\mathbf{R})$ peuvent être pris comme paramètres :

THÉORÈME 4.6 (Woodin). — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors **ZFC** + **MMW** fournit une axiomatisation empiriquement complète⁴³ de la structure (H_2, \in) et rend les propriétés de celle-ci invariantes par forcing.*

Le théorème 4.6 ne constitue pas la fin de l'histoire pour H_2 à la façon dont **DP** le fait pour H_1 , car il y manque au moins la compatibilité de l'axiome **MMW** avec l'existence de grands cardinaux, c'est-à-dire l'équivalent du théorème de Martin–Steel. Le théorème 4.5 garantit cette compatibilité dans le cas de **MMB**, mais la question reste ouverte pour **MMW**. Pour autant, on verra dans la suite que le théorème 4.6 semble proche d'une solution : s'il n'existe pas à ce jour de démonstration d'une compatibilité complète, à savoir qu'aucun axiome de grand cardinal ne peut réfuter **MMW**, du moins Woodin montre qu'aucun axiome de grand cardinal possédant un modèle canonique ne peut réfuter **MMW**. Nous y reviendrons dans la section 6.

⁴⁰Ce résultat est à rapprocher du théorème de Levy–Shoenfield (dont résulte la proposition 2.4) affirmant : <Tout énoncé $\exists x \psi(x, a)$ avec ψ bornée et a dans H_2 satisfait dans V est déjà satisfait dans H_2 >.

⁴¹Dans la suite, un tel énoncé (à une seule alternance de quantificateurs) sera dit *de type $\forall \exists$* .

⁴²La forme originale, notée (*) par Woodin, est <**AD** est satisfait dans $L(\mathbf{R})$ et $L(\mathcal{P}(\aleph_1))$ est extension \mathbf{P}_{max} -générique de $L(\mathbf{R})$ > ; une reformulation plus aisément intelligible sera donnée dans la section 5.

⁴³au sens où elle prouve tout énoncé qui ne peut être réfuté par passage à une extension générique

5. LA Ω -LOGIQUE

Au cours des dernières années, Woodin a proposé un nouveau cadre conceptuel donnant des résultats précédents une formulation plus simple, et ouvrant de nombreuses perspectives. L'idée est d'utiliser une notion de prouvabilité intégrant directement l'invariance par forcing.

Jusqu'à présent, nous avons cherché à caractériser les énoncés satisfaits dans une structure (H, \in) comme ceux qui peuvent être prouvés à partir d'axiomes convenables, au moyen de la notion usuelle de preuve (logique du premier ordre). En utilisant une nouvelle relation de prouvabilité plus subtile⁴⁴, on peut espérer décrire de façon plus simple des objets qui ne le sont pas, et mettre à jour des phénomènes qui, sinon, resteraient cachés.

La Ω -logique de Woodin est fondée sur la notion d'ensemble *universellement Baire* [3].

DÉFINITION 5.1 (Feng–Magidor–Woodin). — *Un sous-ensemble B de \mathbf{R}^p est dit universellement Baire si, pour toute fonction continue $f : K \rightarrow \mathbf{R}^p$ avec K compact, $f^{-1}(B)$ a la propriété de Baire dans K ⁴⁵.*

Les boréliens sont universellement Baire ; s'il existe une classe propre de cardinaux de Woodin, il en est de même de tous les ensembles projectifs.

L'idée de Woodin est d'utiliser les ensembles universellement Baire comme témoins pour une nouvelle notion de prouvabilité. Si B est un sous-ensemble universellement Baire de \mathbf{R} , il peut s'écrire comme projection des branches infinies d'un arbre \tilde{B} sur $\lambda_B \times \aleph_0$, où λ_B est un ordinal convenable, et \tilde{B} joue le rôle d'un *code* pour B . Lorsqu'on passe de V à une extension générique $V[G]$, l'arbre \tilde{B} est préservé, mais l'ensemble de ses branches dans $V[G]$, qu'on notera B_G , peut inclure B strictement.

DÉFINITION 5.2. — *Soit (M, \in) un modèle transitif⁴⁶ de **ZFC**, et B un sous-ensemble universellement Baire de \mathbf{R} . On dit que (M, \in) est B -clos si, pour toute extension générique $V[G]$ de V , l'ensemble $B_G \cap M[G]$ appartient à $M[G]$.*

L'intuition doit être celle d'une propriété de clôture, à savoir que M contient les témoins nécessaires à établir le caractère universellement Baire de B . Si B est un borélien, tout modèle transitif dénombrable de **ZFC** est B -clos. Par contre, plus l'ensemble B est compliqué, et plus la condition d'être B -clos est exigeante.

DÉFINITION 5.3. — *Supposons qu'il existe une classe propre de cardinaux de Woodin⁴⁷. On dit qu'un sous-ensemble universellement Baire B de \mathbf{R} est une Ω -preuve pour un*

⁴⁴Attention ! Il ne s'agit d'utiliser une logique alternative qu'au niveau des *énoncés*, ce qui ne change rien aux *démonstrations* de ces énoncés : les théorèmes sont de *vrais* théorèmes...

⁴⁵c'est-à-dire qu'il existe un ouvert U tel que la différence symétrique entre $f^{-1}(B)$ et U soit maigre

⁴⁶Ceci signifie que M est inclus dans V , a la même appartenance, et que $x \in y \in M$ entraîne $x \in M$.

⁴⁷Ce contexte de grands cardinaux n'est pas indispensable, mais il permet une formulation plus simple.

énoncé φ si φ est satisfait dans tout modèle transitif dénombrable de **ZFC** qui est B -clos. On dit qu'un énoncé φ est Ω -prouvable s'il admet au moins une Ω -preuve.

Une Ω -preuve n'est pas une preuve au sens usuel, mais, comme une preuve, elle va être utilisée comme certifiant que l'énoncé considéré a une certaine propriété.

Tout énoncé prouvable est Ω -prouvable : si φ est prouvable (en logique usuelle) à partir de **ZFC**, alors φ est satisfait dans tout modèle de **ZFC**, donc en particulier dans tout modèle transitif dénombrable, et φ admet donc comme Ω -preuve n'importe quel ensemble universellement Baire (par exemple l'ensemble vide). Mais il existe des énoncés Ω -prouvables dont les seules Ω -preuves sont plus compliquées que les boréliens et qui ne sont pas prouvables au sens usuel : la Ω -logique étend strictement la logique usuelle.

Les énoncés Ω -prouvables sont-ils vrais ? D'après le théorème de complétude de Gödel, si un énoncé φ est Ω -prouvable, mais non prouvable, il existe au moins un modèle (M, E) de **ZFC** dans lequel φ n'est pas satisfait. Néanmoins, on va voir que φ doit être satisfait dans tous les modèles suffisamment proches du modèle V des vrais ensembles — donc, en un sens, dans tous les modèles qui nous intéressent.

Pour α ordinal, on note V_α l'ensemble de tous les ensembles purs pouvant être obtenus à partir de \emptyset en utilisant au plus α fois le passage à l'ensemble des parties. Les structures (V_α, \in) peuvent être vues comme des approximations de (V, \in) ; en général, (V_α, \in) n'est pas un modèle de **ZFC**, mais c'est le cas dès que α est un cardinal *inaccessible*⁴⁸.

DÉFINITION 5.4. — *On dit qu'un énoncé φ est Ω -valide si φ est satisfait dans tout modèle de **ZFC** du type $\langle V_\alpha, \in \rangle$ calculé dans une extension générique (quelconque) de V .*

PROPOSITION 5.5. — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors tout énoncé Ω -prouvable est Ω -valide.*

Toute notion de prouvabilité mène naturellement à une notion d'axiomatisation :

DÉFINITION 5.6. — *Supposons H définissable. On dit que \mathbf{A} est un Ω -axiome pour la structure (H, \in) si, pour tout φ , un et un seul des deux énoncés $\mathbf{A} \Rightarrow \langle (H, \in) \text{ satisfait } \varphi \rangle$, $\mathbf{A} \Rightarrow \langle (H, \in) \text{ satisfait } \neg\varphi \rangle$ est Ω -prouvable.*

Ainsi, un Ω -axiome pour (H, \in) $\langle \Omega$ -décide \rangle chaque propriété de H . Ce qui rend à la fois intéressante et naturelle la recherche d'un Ω -axiome est le point suivant. Si \mathbf{A} est un Ω -axiome pour (H, \in) , alors il peut certes exister des énoncés φ tels que $\mathbf{A} \Rightarrow \langle (H, \in) \text{ satisfait } \varphi \rangle$ soit Ω -prouvable mais non prouvable, mais, dans ce cas, on est du moins assuré par la proposition 5.5 que φ ne peut pas être réfuté par passage à une extension générique à partir d'un modèle de **ZFC** + \mathbf{A} . On n'a donc pas nécessairement

⁴⁸Un cardinal κ est dit *inaccessible* si κ est non dénombrable, $\lambda < \kappa$ implique $2^\lambda < \kappa$, et si la conjonction de $\lambda < \kappa$ et de $(\forall \alpha < \lambda)(\lambda_\alpha < \kappa)$ implique $\sup\{\lambda_\alpha; \alpha < \lambda\} < \kappa$ — autrement dit si κ ne peut pas être atteint à partir d'objets plus petits par passage à l'ensemble des parties ou à la limite. Les cardinaux *inaccessibles* sont les plus petits des grands cardinaux.

une description complète de H , mais on retrouve le même type de complétude empirique et libérée du forcing qu'avec **ZFC** et l'arithmétique, comme réclamé dans le problème 2.5.

Le résultat suivant montre que l'idée de $\langle \Omega$ -décider \rangle tout énoncé est raisonnable, puisque, dans le cas de H_1 , aucun axiome additionnel n'est nécessaire :

PROPOSITION 5.7. — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors, pour tout φ , un et un seul des énoncés $\langle (H_1, \in) \text{ satisfait } \varphi \rangle$, $\langle (H_1, \in) \text{ satisfait } \neg\varphi \rangle$ est Ω -prouvable⁴⁹.*

Passons à la structure (H_2, \in) , qui est le niveau critique. Le contexte de la Ω -logique permet de donner de l'axiome **MMW** une formulation plus facilement intelligible. Si on suppose l'existence d'une classe propre de cardinaux de Woodin, alors **MMW** équivaut à :

Pour tout A inclus dans \mathbf{R} appartenant à $L(\mathbf{R})$, tout énoncé de type $\forall\exists$ portant sur $(H_2, \mathcal{I}_{NS}, A, \in)$ et dont la négation n'est pas Ω -prouvable est satisfait.

Ceci montre que **MMW** est un principe de maximalité pour $(H_2, \mathcal{I}_{NS}, \in)$ analogue à la propriété de clôture algébrique⁵⁰ : un corps K est algébriquement clos si tout système non contradictoire d'équations algébriques à paramètres dans K y a une solution, c'est-à-dire précisément si toute propriété $\forall\exists$ de la structure $(K, +, \times)$ compatible avec les axiomes des corps y est satisfaite⁵¹. Ainsi, dire que l'axiome **MMW** est vrai est analogue à affirmer que H_2 est, en un certain sens, algébriquement clos.

À partir du théorème 4.6, Woodin montre le résultat suivant, qui établit la pertinence de la Ω -logique au niveau de H_2 :

THÉORÈME 5.8 (Woodin). — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors **MMW** est un Ω -axiome pour (H_2, \in) .*

La question de la compatibilité de l'axiome **MMW** avec l'existence de grands cardinaux a été laissée ouverte dans la section 4, et on la retrouve ici sous une forme équivalente. L'hypothèse que **A** est un Ω -axiome a deux composantes : d'une part le fait que, pour chaque φ , l'une au moins des implications $\mathbf{A} \Rightarrow \varphi$, $\mathbf{A} \Rightarrow \neg\varphi$ est Ω -prouvable, et, d'autre part, le fait que ces deux possibilités s'excluent, ce qui revient à la non- Ω -prouvabilité de $\neg\mathbf{A}$. Dans le cas de l'axiome **MMW**, ce dernier point provient de la construction, grâce au forcing \mathbf{P}_{max} , d'un modèle satisfaisant **MMW**. Ce que ne garantit pas ce résultat, c'est la satisfaction de **MMW** dans un modèle du type $\langle (V_\alpha, \in) \rangle$ calculé dans une extension

⁴⁹Le résultat peut se formuler en disant que $0 = 0$ est un Ω -axiome pour H_1 ; noter qu'il implique que \mathbf{DP} est Ω -prouvable, puisque, sous les hypothèses, \mathbf{DP} doit être satisfaite.

⁵⁰L'analogie est encore plus pertinente lorsqu'on restreint l'hypothèse à $\langle A \text{ projectif} \rangle$; on obtient alors un axiome \mathbf{MMW}_0 plus faible que **MMW**, mais qui a les mêmes propriétés vis-à-vis de l'axiomatisation de H_2 . Une formulation semblable serait possible pour l'axiome **MMB** dont **MMW** est une variante.

⁵¹Dans ce cas, un énoncé de type $\forall\exists$ ne met en jeu que des combinaisons booléennes d'équations puisqu'il n'y a pas de relation dans la structure considérée.

générique de V , qui, elle, fournirait la compatibilité avec l'existence de grands cardinaux. En termes de Ω -logique, ceci revient à dire qu'il manque une démonstration de ce que la négation de **MMW** n'est pas Ω -valide : le problème est de passer de la non- Ω -prouvabilité à la non- Ω -validité, soit, de façon équivalente, de la Ω -validité à la Ω -prouvabilité.

Il s'agit d'une problématique usuelle. La Ω -logique a une notion de preuve (existence d'un ensemble universellement Baire certifiant les propriétés d'un énoncé) et une notion de validité (satisfaction d'un énoncé dans toutes les approximations des extensions génériques de l'univers). La proposition 5.5 est un résultat de cohérence affirmant que les énoncés prouvables sont valides. Comme pour toute logique se pose la question de l'implication réciproque, c'est-à-dire de la complétude : tous les énoncés valides sont-ils prouvables ?

CONJECTURE 5.9 (Ω -conjecture, Woodin, 1999). — *Tout énoncé Ω -valide est Ω -prouvable*⁵².

Essentiellement, la Ω -conjecture affirme que tous les énoncés non réfutables par passage à une extension générique ont une <preuve> dans la famille des ensembles universellement Baire. Cette conjecture admet des formes équivalentes sur lesquelles nous reviendrons.

Appelons *problème 2.5** la variante du problème 2.5 dans laquelle on laisse de côté la contrainte (non précisément définie) <description suffisamment complète>.

PROPOSITION 5.10. — *Si la Ω -conjecture est vraie, alors **ZFC** + **A** est une solution au problème 2.5* pour (H, ϵ) si et seulement si **A** est un Ω -axiome pour (H, ϵ) .*

DÉMONSTRATION (esquisse). — Supposons que **A** est un Ω -axiome pour (H, ϵ) . On a vu ci-dessus que, par construction de la Ω -logique, les propriétés de H sont nécessairement invariantes par forcing. Reste la question de la compatibilité de **A** avec l'existence de grands cardinaux. Dans le contexte de la Ω -logique, il s'agit de montrer que $\neg\mathbf{A}$ n'est pas Ω -valide. Or l'hypothèse que **A** est un Ω -axiome garantit que $\neg\mathbf{A}$ n'est pas Ω -prouvable. Si la Ω -conjecture est vraie, non- Ω -prouvabilité entraîne non- Ω -validité.

Inversement, supposons que **ZFC**+**A** rend les propriétés de (H, ϵ) invariantes par forcing. Alors, pour tout énoncé φ , un et un seul des deux énoncés $\mathbf{A} \Rightarrow \langle (H, \epsilon) \text{ satisfait } \varphi \rangle$, ou $\mathbf{A} \Rightarrow \langle (H, \epsilon) \text{ satisfait } \neg\varphi \rangle$ est Ω -valide. Si la Ω -conjecture est vraie, cela implique que l'un au moins de ces énoncés est Ω -prouvable. De plus, la compatibilité de **A** avec l'existence de grands cardinaux entraîne que $\neg\mathbf{A}$ n'est pas Ω -valide, donc *a fortiori* pas Ω -prouvable, et **A** est par conséquent un Ω -axiome pour (H, ϵ) . \square

Du théorème 5.8, qui affirme que **MMW** est un Ω -axiome pour (H_2, ϵ) , et du théorème 4.6, qui garantit une forme de complétude, on déduit :

COROLLAIRE 5.11. — *Si la Ω -conjecture est vraie, alors **ZFC** + **MMW** est une solution au problème 2.5 pour (H_2, ϵ) .*

⁵²Cette formulation n'est adéquate que pour les énoncés de type $\forall\exists$; la formulation dans le cas général est légèrement plus compliquée.

6. Ω -LOGIQUE ET GRANDS CARDINAUX

La Ω -logique est liée aux grands cardinaux : en un sens, c'est la logique des grands cardinaux — ou, tout au moins, des grands cardinaux admettant des modèles canoniques d'un certain type. La description de ce lien va permettre de donner un sens précis à l'affirmation suivant laquelle les résultats de Woodin parviennent près d'une démonstration de la Ω -conjecture. D'autre part, on va voir que la Ω -logique fournit une explication conceptuelle élégante à la constatation empirique que les axiomes de grand cardinaux s'organisent en une hiérarchie.

Il est facile de vérifier que tous les axiomes de grands cardinaux considérés à ce jour entrent dans le cadre abstrait suivant⁵³ :

DÉFINITION 6.1. — *On appelle axiome de grand cardinal tout énoncé $\exists\kappa\psi(\kappa)$ avec ψ de type $\exists\forall$ tel que, si $\psi(\kappa)$ est satisfait dans V , alors κ est un cardinal inaccessible, et, de plus, $\psi(\kappa)$ reste satisfait dans toute extension générique de V associée à un ensemble de forcing de cardinal moindre que κ . On dit alors que $\exists\kappa\psi(\kappa)$ est accompli si, pour tout ensemble X , il existe un modèle transitif (M, \in) de **ZFC** et un ordinal κ de M tels que X appartienne à $V_\kappa \cap M$ et que (M, \in) satisfasse $\psi(\kappa)$.*

L'accomplissement de $\exists\kappa\psi(\kappa)$ signifie qu'il existe beaucoup de modèles de **ZFC** contenant au moins un cardinal ayant la propriété ψ . S'il existe une classe propre de cardinaux inaccessibles κ vérifiant $\psi(\kappa)$, alors $\exists\kappa\psi(\kappa)$ est accompli : avec les notations de la définition, il suffit de prendre $M = V_\lambda$ avec λ inaccessible assez grand.

Le résultat suivant montre que les énoncés Ω -prouvables sont ceux qui sont prouvables (en logique usuelle) à partir d'un axiome de grand cardinal accessible à la Ω -logique :

PROPOSITION 6.2. — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors un énoncé φ de type $\forall\exists$ est Ω -prouvable si, et seulement si, il existe un axiome de grand cardinal **A** tel que $\langle \mathbf{A}$ est accompli \rangle soit Ω -prouvable et que φ soit prouvable (en logique usuelle) à partir de **ZFC** + $\langle \mathbf{A}$ est accompli \rangle .*

COROLLAIRE 6.3. — *La Ω -conjecture équivaut à la condition : pour chaque axiome de grand cardinal **A** accompli dans V , l'assertion $\langle \mathbf{A}$ est accompli \rangle est Ω -prouvable.*

On peut maintenant préciser dans quelle mesure la Ω -conjecture est approchée dans les travaux de Woodin. D'après le corollaire 6.3, le problème est de déterminer les axiomes de grand cardinal **A** tels que l'énoncé $\langle \mathbf{A}$ est accompli \rangle soit Ω -prouvable. Or, il existe un programme de modèles canoniques qui, *grosso modo*, consiste à construire pour chaque axiome de grand cardinal **A** un modèle minimal où **A** soit satisfait, suivant le schéma du modèle de Gödel L ⁵⁴. Ce programme, qui repose sur une méthode dite de *comparaison*,

⁵³On ne prétend pas que tout énoncé du type suivant doive correspondre à une idée de grand cardinal.

⁵⁴Celui-ci correspond au cas de **ZFC**, c'est-à-dire au cas où aucun axiome de grand cardinal n'est ajouté.

atteint à l'heure actuelle le niveau de l'axiome <Il existe une infinité de cardinaux de Woodin> [13], mais pas (encore) celui de l'axiome <Il existe un cardinal supercompact>. Le résultat suivant repose sur une analyse fine d'une notion générale de modèle canonique :

THÉORÈME 6.4 (Woodin). — *Pour tout axiome de grand cardinal \mathbf{A} pour lequel un modèle canonique fondé sur la méthode de comparaison peut exister, l'énoncé < \mathbf{A} est accompli> est Ω -prouvable⁵⁵.*

Comme le théorème 6.4 est essentiellement une équivalence, l'enjeu de la Ω -conjecture est la possibilité de pousser la méthode de comparaison à tout grand cardinal — on voit donc que cette conjecture pourrait être réfutée en montrant qu'un certain *très* grand cardinal restera définitivement inaccessible à toute notion de modèle canonique.

Passons au second point de cette section, à savoir l'explication conceptuelle apportée par la Ω -logique à la hiérarchie des grands cardinaux. Comme on l'a dit, on constate que tous les axiomes de grands cardinaux considérés à ce jour s'organisent en une hiérarchie linéaire : étant donnés deux tels axiomes, il apparaît toujours que l'un implique l'autre, ou, au moins, que la *consistance* de l'un (c'est-à-dire sa non-contradiction) implique la consistance de l'autre. On obtient ainsi une hiérarchie bien ordonnée liée à la consistance relative et calibrant la force logique des axiomes : par exemple, la consistance (de l'existence) d'un cardinal supercompact entraîne celle d'une infinité de cardinaux de Woodin, qui elle-même entraîne celle d'un cardinal de Woodin ; cette dernière entraîne la consistance (de l'existence) d'un cardinal mesurable, qui elle-même implique celle d'un cardinal inaccessible (et même de beaucoup).

Le point de départ est l'existence d'une échelle de complexité sur les ensembles universellement Baire, déduite de la notion de *réductibilité de Wadge*.

DÉFINITION 6.5. — *Pour $B, B' \subseteq \mathbb{K}$ ⁵⁶, on dit que B est réductible (resp. fortement réductible) à B' si on a $B = f^{-1}(B')$ avec $f : \mathbb{K} \rightarrow \mathbb{K}$ continue (resp. 1/2-lipschitzienne).*

PROPOSITION 6.6. — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors, quels que soient les sous-ensembles universellement Baire B, B' de \mathbb{K} , ou bien B est réductible à B' , ou bien B' est fortement réductible à $\mathbb{K} \setminus B$.*

Comme aucun sous-ensemble de \mathbb{K} ne peut être fortement réductible à son complémentaire, on déduit un préordre sur les ensembles universellement Baire en déclarant $B \prec B'$ vrai si à la fois B et $\mathbb{K} \setminus B$ sont fortement réductibles à B' . On montre alors que \prec n'a pas de chaîne infinie descendante, ce qui permet d'attacher à chaque ensemble universellement Baire B un ordinal qu'on appellera sa *complexité*.

THÉORÈME 6.7 (Woodin). — *Pour \mathbf{A} axiome de grand cardinal, notons $\rho(\mathbf{A})$ la complexité minimale d'une Ω -preuve de l'énoncé < \mathbf{A} est accompli>. Alors, pour les niveaux où elles*

⁵⁵Il en résulte que la Ω -conjecture est vraie dans tout modèle canonique...

⁵⁶Pour simplifier la formulation, on quitte \mathbb{R} pour retourner à l'espace de Cantor $\mathbb{K} = \{0, 1\}^{\mathbb{N}}$.

sont définies⁵⁷, la hiérarchie de consistance des axiomes de grand cardinaux coïncide avec la hiérarchie définie par ρ .

Ce résultat esthétique et profond est un argument fort en faveur de la Ω -logique. Si la Ω -conjecture est vraie, la hiérarchie définie par ρ recouvre tous les axiomes de grands cardinaux ; si elle est fautive, cette hiérarchie n'est que le début d'une hiérarchie plus étendue dont nous ne savons encore rien.

7. LES RÉSULTATS SUR L'HYPOTHÈSE DU CONTINU

Un des aspects les plus spectaculaires des développements récents est la perspective nouvelle qu'ils ouvrent pour le problème du continu.

Avant de présenter cette avancée — et sans prétendre retracer l'historique des résultats sur le problème du continu depuis vingt ans — on mentionnera un résultat de Woodin (1984) soulignant le rôle critique de l'hypothèse du continu :

PROPOSITION 7.1. — *Supposons qu'il existe une classe propre de cardinaux de Woodin mesurables. Alors deux extensions génériques de V satisfaisant **HC** satisfont nécessairement les mêmes énoncés existentiels⁵⁸ à paramètre \mathbf{R} .*

En d'autres termes, dès que deux extensions génériques sont d'accord sur **HC**, elles sont aussi d'accord sur toutes les propriétés de même complexité syntaxique que **HC**.

Par ailleurs, il a été noté depuis longtemps qu'il existe une dissymétrie entre **HC** et $\neg\mathbf{HC}$ et qu'au moins certaines variantes de $\neg\mathbf{HC}$ peuvent être prouvées alors que ce n'est pas le cas pour les variantes correspondantes de **HC**. En outre, plusieurs résultats remarquables (et généralement de démonstration difficile) ont établi que, dans des contextes divers où elle n'est *a priori* pas clairement déterminée par les hypothèses, la valeur du continu, c'est-à-dire de 2^{\aleph_0} , se trouve être \aleph_2 . Par exemple, Foreman, Magidor et Shelah ont montré dans [5] que **MM** entraîne $2^{\aleph_0} = \aleph_2$, et Woodin a montré dans [18] qu'il en est de même de l'hypothèse que l'idéal \mathcal{I}_{NS} a une certaine propriété combinatoire dite de \aleph_2 -saturation. Dernièrement, Todorcevic a montré dans [17], par un argument combinatoire élégant et direct, que **MMB** entraîne une version effective forte de l'égalité $2^{\aleph_0} = \aleph_2$, à savoir la possibilité de *définir* un bon ordre de longueur \aleph_2 sur \mathbf{R} en utilisant comme seul paramètre une unique suite de réels de longueur \aleph_1 (donc un élément de H_2). Collectivement, tous ces résultats peuvent être vus comme une indication heuristique en défaveur de **HC**.

Venons-en au résultat récent de Woodin. Par le lemme 4.2, toute description suffisamment complète de H_2 doit inclure une solution de l'hypothèse du continu. Par exemple, il n'est pas très difficile de voir que, tout comme **MM** ou **MMB**, l'axiome **MMW** entraîne

⁵⁷c'est-à-dire pour les axiomes de grands cardinaux pour lesquels existent des modèles canoniques

⁵⁸c'est-à-dire dont les seules quantifications non bornées sont de type \exists

que **HC** est fausse et qu'en l'occurrence on a $2^{\aleph_0} = \aleph_2$. Mais ce résultat ne dit rien sur la situation de **HC** vis-à-vis d'autres axiomatisations possibles de (H_2, \in) . Établi en 2000, le dernier des théorèmes sur lequel on souhaite insister est le suivant :

THÉORÈME 7.2 (Woodin). — *Supposons qu'il existe une classe propre de cardinaux de Woodin. Alors tout Ω -axiome pour H_2 dont la négation n'est pas Ω -valide implique que l'hypothèse du continu est fausse.*

La démonstration du théorème 7.2 — qui est un tour de force technique — repose sur une analyse des ensembles Ω -récurifs, qui jouent pour la Ω -logique le rôle des ensembles récurifs de la logique classique. Un sous-ensemble T de \mathbb{N} est dit *récurif* s'il existe un algorithme (représentable à l'aide d'une machine de Turing) permettant de reconnaître ses éléments⁵⁹. Parmi de nombreuses caractérisations équivalentes, les ensembles récurifs sont ceux qui peuvent être définis dans (H_0, \in) à la fois par une formule existentielle et par la négation d'une formule existentielle. On note $L(B, \mathbb{R})$ le modèle de **ZFC** construit comme le modèle L de Gödel, mais en partant des ensembles B et \mathbb{R} .

DÉFINITION 7.3. — *Un sous-ensemble T de \mathbb{N} est dit Ω -récurif s'il existe un sous-ensemble universellement Baire B de \mathbb{R} tel que T puisse être défini dans $(L(B, \mathbb{R}), \in, \{\mathbb{R}\})$ à la fois par une formule existentielle et par la négation d'une formule existentielle.*

Comme H_0 est définissable dans $L(\mathbb{R})$, tout ensemble récurif est Ω -récurif, mais l'existence d'ensembles universellement Baire (beaucoup) plus compliqués que les boréliens entraîne que les ensembles Ω -récurifs peuvent être (beaucoup) plus compliqués que les ensembles récurifs.

Le point crucial consiste à étudier la définissabilité éventuelle des ensembles Ω -récurifs dans (H_2, \in) — ou, plus généralement, dans l'ensemble des ensembles héréditairement de cardinal au plus celui de \mathbb{R} . Ceci requiert de nombreux développements mettant en jeu à la fois des outils de théorie descriptive des ensembles (étude des sous-ensembles de \mathbb{R}) et de la théorie des grands cardinaux et de la détermination. Une composante essentielle est l'adaptation aux axiomes de détermination des méthodes de construction de modèles canoniques élaborées par Mitchell et Steel [13] pour les axiomes de grands cardinaux, en particulier cardinaux de Woodin. Pour ce faire, Woodin introduit une famille complètement nouvelle de modèles canoniques notés $\text{HOD}^{L(B, \mathbb{R})}$ et indexés par les ensembles universellement Baire. L'aboutissement de cette analyse techniquement très sophistiquée est

PROPOSITION 7.4. — *Supposons qu'il existe une classe propre de cardinaux de Woodin, et que T est un sous-ensemble Ω -récurif de \mathbb{N} . Alors*

⁵⁹Le caractère effectif des règles de la logique usuelle entraîne que l'ensemble des (numéros des) énoncés prouvables à partir de **ZFC** est la projection d'un ensemble récurif, alors que l'ensemble des (numéros des) énoncés satisfaits dans (H_0, \in) ne l'est pas, ce qui entraîne que l'inclusion du premier dans le second est stricte, d'où le premier théorème d'incomplétude de Gödel.

- (i) ou bien T est définissable dans (H_2, \in) ,
(ii) ou bien il existe une surjection définissable de \mathbf{R} sur \aleph_2 .

Le théorème 7.2 résulte de la proposition 7.4 par un argument de diagonalisation. En effet, si \mathbf{A} est un Ω -axiome pour (H_2, \in) , alors l'ensemble des numéros des énoncés φ tels que $\mathbf{A} \Rightarrow \langle (H_2, \in) \text{ satisfait } \varphi \rangle$ soit Ω -prouvable est un ensemble Ω -récursif (c'est facile). Or un résultat classique de Tarski affirme que l'ensemble des énoncés satisfaits dans une structure S ne peut jamais être définissable dans S . En particulier donc, l'ensemble des (numéros des) énoncés satisfaits dans (H_2, \in) ne peut être définissable dans (H_2, \in) , et la seule possibilité au regard de la proposition 7.4 est le cas (ii), c'est-à-dire la fausseté (en un sens effectif fort) de l'hypothèse du continu.

COROLLAIRE 7.5. — *Si la Ω -conjecture est vraie, alors toute solution au problème 2.5 pour H_2 implique que l'hypothèse du continu soit fausse*⁶⁰.

On obtient ainsi la version plus précise suivante de la conjecture de l'introduction :

THÉORÈME 7.6. — *Si la Ω -conjecture est vraie, alors toute théorie des ensembles obtenue en ajoutant à **ZFC** un axiome compatible avec l'existence de grands cardinaux et rendant les propriétés de (H_2, \in) invariantes par forcing implique que l'hypothèse du continu soit fausse.*

Nous avons vu que, toujours si la Ω -conjecture est vraie, l'axiomatisation par **ZFC** + **MMW** est une solution au problème 2.5 pour H_2 . Par conséquent, on se trouve alors, vis-à-vis de H_2 et de **HC**⁶¹, exactement dans la situation envisagée à la fin de la section 2 : le problème 2.5 a une solution, et toute solution implique que **HC** soit fausse. On parvient ainsi à la conclusion suivante :

COROLLAIRE 7.7. — *Si la Ω -conjecture est vraie, alors la négation de l'hypothèse du continu est établie, au sens décrit à la fin de la section 2.*

Le lecteur sceptique ou pressé pourra retenir que la solution du problème du continu est reportée sur celle d'un nouveau problème tout aussi ouvert, et d'énoncé de surcroît obscur. Ce serait une courte vue, en particulier parce que la nature de la Ω -conjecture est très différente de celle de **HC** et qu'on peut raisonnablement en escompter une démonstration⁶²

⁶⁰Signalons également pour terminer une variante du résultat sur **HC**. Énumérons toutes les formules à une variable libre, et définissons \mathcal{Q}^Ω comme l'ensemble des couples (n, r) dans $\mathbf{N} \times \mathbf{R}$ tels que $\varphi_n(r)$ soit Ω -prouvable. Woodin montre que, si l'ensemble \mathcal{Q}^Ω n'est pas universellement Baire, alors, essentiellement, toute solution au problème 2.5 pour H_2 implique que l'hypothèse du continu est fausse : autrement dit, l'hypothèse de pure analyse $\langle \mathcal{Q}^\Omega \text{ n'est pas universellement Baire} \rangle$ permet d'obtenir la même conclusion que la Ω -conjecture.

⁶¹ou plutôt de l'énoncé φ_{HC} qui code **HC** dans H_2

⁶²En particulier, Woodin établit un lien entre les énoncés Ω -valides et les ensembles universellement Baire, qui n'est pas (encore) la Ω -prouvabilité, mais s'en rapproche. Dans [21], Woodin propose un programme susceptible de mener à une démonstration fondée sur de nouveaux modèles canoniques.

ou une réfutation dans le futur — et, de là, une solution d’au moins un des aspects du problème du continu. Un point remarquable est que la résolution puisse venir de domaines très divers de la théorie des ensembles, ce qui témoigne à la fois du caractère central de la conjecture et de l’unité profonde de la théorie des ensembles.

8. CONCLUSION

Deux conclusions devraient se dégager — et attester que l’histoire de la théorie des ensembles ne s’est pas arrêtée avec le théorème de Cohen.

La première est que l’axiome de détermination projective **DP** constitue, lorsqu’ajouté au système **ZFC**, la bonne axiomatisation de H_1 , c’est-à-dire de l’analyse, à la façon dont **ZFC** constitue la bonne axiomatisation de H_0 , c’est-à-dire de l’arithmétique. L’axiome **DP** mène à une théorie heuristiquement complète et satisfaisante pour H_1 , c’est-à-dire pour le domaine des ensembles dénombrables. Ce succès constitue un argument de poids en faveur de la *vérité* de l’axiome **DP**. Il peut paraître étonnant d’identifier efficacité et vérité⁶³. Mais, faute d’évidence intuitive *a priori*⁶⁴, il est difficile d’imaginer d’autre critère de vérité que l’évidence empirique *a posteriori* née de l’efficacité. Que le lecteur réfléchisse à l’axiome affirmant l’existence d’ensembles infinis : son efficacité opératoire est telle que nul ne songe à le remettre en cause et à renoncer, par exemple, aux nombres réels. Pourtant, cet axiome ne possède aucune justification théorique intrinsèque, non plus qu’aucune évidence intuitive, sinon l’intériorisation d’une longue familiarité. La situation avec la détermination projective est similaire, et la familiarité acquise par les théoriciens des ensembles donne aujourd’hui à cette notion d’infini forte qu’est **DP** l’évidence intuitive qu’une familiarité semblable a donnée jadis à la notion d’infini simple⁶⁵.

La seconde conclusion est qu’il n’existe pour le moment pas de solution possédant le même degré d’évidence pour le niveau de H_2 , c’est-à-dire celui des sous-ensembles de \aleph_1 , mais qu’il existe au moins une solution globale à ce niveau, à savoir celle développée par Woodin à partir de l’axiome **MMW** et de la Ω -logique.

⁶³*A priori*, il serait étrange, ayant à étudier un corps K inconnu, de proclamer : <Les corps algébriquement clos ont une théorie satisfaisante, donc je vais supposer K algébriquement clos>; en fait, le problème ne se pose pas exactement dans les termes d’étudier un corps K inconnu et possiblement quelconque, mais d’étudier <le monde des corps>, auquel cas supposer qu’on vit dans un grand tout algébriquement clos est une hypothèse plus que raisonnable.

⁶⁴Les expériences de pensée (<thought experiments>) parfois invoquées [10] ne vont pas loin...

⁶⁵cf. Gödel [7] : <There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole discipline, and furnishing such powerful methods for solving given problems (and even solving them, as far as possible, in a constructivistic way) that quite irrespective of their intrinsic necessity they would have to be assumed at least in the same sense as any established physical theory.>

Même si la Ω -conjecture est établie un jour, la discussion sur la légitimité de l'invariance par forcing comme critère de base ne sera pas close⁶⁶, et il n'est pas certain que la Ω -logique soit le seul cadre raisonnable. C'est pourquoi il serait imprudent d'affirmer que la solution du problème du continu proposée par Woodin est la seule possible.

Par contre, et même sans le poids plus décisif qu'apporteraient des idées de démonstrations, les résultats survolés ici devraient rendre irréfutable la possibilité d'une véritable théorie conceptuelle de l'infini non dénombrable, tenant debout d'elle-même, possédant une logique interne et une intuition propres. Aucun argument comparable ne peut être avancé par les opposants à une telle théorie, en particulier par ceux qui considèrent le problème du continu comme essentiellement irrésoluble⁶⁷.

Enfin et dans tous les cas, j'espère en avoir dit assez pour faire soupçonner que les travaux de Woodin constituent un remarquable morceau de mathématiques.

RÉFÉRENCES

- [1] J. Bagaria, *Bounded forcing axioms as principles of generic absoluteness*, Arch. Math. Logic **69–6** (2000) 393–401.
- [2] P. Dehornoy, *La détermination projective (d'après Martin, Steel et Woodin)*, Séminaire Bourbaki, Exp. n° 710, Astérisque **177–178** (1989) 261–276.
- [3] M. Feng, M. Magidor, & H. Woodin, *Universally Baire sets of reals*, in : H. Judah, W. Just and H. Woodin (Eds), *Set Theory of the Continuum*, MSRI Publ. 26, 203–242, Springer, 1992.
- [4] M. Foreman, *Generic large cardinals : new axioms for mathematics ?*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998).
- [5] M. Foreman, M. Magidor, & S. Shelah, *Martin's maximum, saturated ideals, and nonregular ultrafilters*, Ann. of Math. **127–1** (1988) 1–47.
- [6] H. Friedman, *On the necessary use of abstract set theory*, Adv. in Math. **41** (1981) 209–280.
- [7] K. Gödel, *What is Cantor's Continuum Problem ?*, Amer. Math. Monthly **54** (1947) 515–545.

⁶⁶On a signalé notamment que l'invariance par forcing des propriétés ne peut être espérée beaucoup au-delà de H_2 , cf. note 17, donc elle ne peut pas constituer un critère universel ; voir aussi note 18.

⁶⁷cf. Woodin [20] : *<There is a tendency to claim that the Continuum Hypothesis is inherently vague and that this is simply the end of the story. But any legitimate claim that CH is inherently vague must have a mathematical basis, at the very least a theorem or a collection of theorems. My own view is that the independence of CH from ZFC, and from ZFC together with large cardinal axioms, does not provide this basis. I would hope this is the minimum metamathematical assessment of the solution to CH that I have presented. Instead, for me, the independence results for CH simply show that CH is a difficult problem.>*

- [8] M. Goldstern & S. Shelah, *The bounded proper forcing axiom*, J. Symbolic Logic **60–1** (1995) 58–73.
- [9] A. Kanamori, *The higher infinite*, Springer, Berlin, 1994.
- [10] Yu. Manin, *Georg Cantor and his heritage*, arXiv:math.AG/0209244 (2002).
- [11] D.A. Martin & J.R. Steel, *A proof of projective determinacy*, J. Amer. Math. Soc. **2–1** (1989) 71–125.
- [12] Y. Matijasevich & J. Robinson, *Reduction of an arbitrary Diophantine equation in one in 13 unknowns*, Acta Arith. **27** (1975) 521–553.
- [13] W. Mitchell & J. Steel, *Fine structure and iteration trees*, Springer, Berlin, 1994.
- [14] Y. Moschovakis, *Descriptive set theory*, North-Holland, Amsterdam, 1980.
- [15] S. Shelah, *Proper and improper forcing*, 2nd ed., Perspectives in Math. Logic, Springer, Berlin, 1998.
- [16] J. Steel, *Mathematics need new axioms*, Bull. Symb. Logic **6–4** (2000) 422–433.
- [17] S. Todorcevic, *Generic absoluteness and the continuum*, Math. Res. Letters **9** (2002) 465–471.
- [18] W. Hugh Woodin, *The Axiom of Determinacy, forcing axioms, and the nonstationary ideal*, Walter de Gruyter & co., Berlin, 1999.
- [19] W. Hugh Woodin, *The Continuum Hypothesis, I & II*, Notices Amer. Math. Soc. **48–6** (2001) 567–576, & **8–7** (2001) 681–690.
- [20] W. Hugh Woodin, *The Continuum Hypothesis*, Proceedings Logic Colloquium 2000, Paris; à paraître.
- [21] W. Hugh Woodin, *The Continuum Hypothesis and the Ω -Conjecture*, Coxeter Lectures, Fields Institute, Toronto (novembre 2002); Notes disponibles à l'adresse http://av.fields.utoronto.ca/slides/02-03/coxeter_lectures/woodin/

Patrick DEHORNOY

Laboratoire de Mathématiques Nicolas Oresme

CNRS UMR 6139

Université de Caen

B.P. 5186

F-14032 Caen

E-mail : dehornoy@math.unicaen.fr

GROUPES ALÉATOIRES
[d'après Misha GROMOV, ...]

par **Étienne GHYS**

INTRODUCTION

La *théorie combinatoire des groupes* traite pour l'essentiel des groupes de *présentation finie*, c'est-à-dire des groupes fondamentaux des polyèdres finis. Les méthodes employées, d'abord combinatoires et topologiques [CM82], ont pris par la suite un aspect métrique, en particulier sous l'impulsion de M. GROMOV. Aujourd'hui, on préfère parfois la terminologie « *théorie géométrique des groupes* » [Ha00]. Une place considérable est faite à des exemples remarquables qui sont analysés en détail : groupes libres, groupes fondamentaux de surfaces, groupes arithmétiques, etc. Par ailleurs, quelques classes de groupes sont mises en évidence, comme celles des groupes nilpotents, polycycliques, résolubles, moyennables etc., mais à l'évidence il ne s'agit que de classes très particulières qui n'illustrent pas le « comportement typique » d'un groupe de présentation finie. Dans une série d'articles étalés sur une vingtaine d'années, M. GROMOV propose une vision globale des propriétés des groupes « génériques » ou « aléatoires », dans un sens que nous préciserons plus loin. Cet exposé fait le point sur la question. Comme souvent dans ce séminaire, il n'est pas possible en quelques pages d'entrer dans les détails de preuves longues et difficiles et il nous faudra malheureusement nous contenter d'un survol superficiel.

Le rôle joué par les objets génériques par rapport aux exemples spécifiques dépend du domaine des mathématiques considéré. Dans la théorie des systèmes dynamiques par exemple, l'étude des dynamiques génériques est absolument fondamentale alors que la géométrie algébrique accorde peut-être moins d'importance aux variétés algébriques « génériques ». Les méthodes présentées dans cet exposé permettent une première approche aléatoire en théorie géométrique des groupes. L'avenir dira si cette approche est féconde et si une véritable « théorie des groupes aléatoires » est destinée à se développer (à l'instar de la théorie des graphes aléatoires ?). Quoi qu'il en soit, nous verrons que ces méthodes permettent d'ores et déjà de montrer l'existence de groupes aux propriétés surprenantes.

Je remercie THOMAS DELZANT, DAMIEN GABORIAU, YANN OLLIVIER, LIOR SILBERMAN, ALAIN VALETTE et ANDRZEJ ŻUK pour leur aide, et MISHA GROMOV pour ses belles mathématiques.

1. QUELQUES ÉNONCÉS

Pour que notre « marche aléatoire parmi les groupes aléatoires » [Gr02] ne soit pas trop désordonnée, je voudrais présenter d'abord quelques énoncés que nous rencontrerons dans cet exposé et qui serviront de repères.

Soit Γ un groupe engendré par une partie finie S symétrique (*i.e.* $S = S^{-1}$). Pour chaque γ de Γ , on note $|\gamma|_S$ sa *norme*, longueur minimale d'un mot en les éléments de S qui représente γ . Si γ_1 et γ_2 sont deux éléments de Γ , on note $d_S(\gamma_1, \gamma_2) = |\gamma_1^{-1}\gamma_2|_S$. Ceci munit Γ de la *métrie des mots*, invariante par translations à gauche. Il est souvent utile de plonger Γ dans son *graphe de CAYLEY*, dont les sommets sont les éléments du groupe et les arêtes connectent les éléments à distance 1. On prolonge naturellement la distance d_S au graphe (ou plus précisément à sa réalisation géométrique), de manière à ce que chaque arête soit isométrique à l'intervalle $[0, 1]$. Un *segment géodésique* est un plongement isométrique d'un intervalle $[0, n]$ dans le graphe de CAYLEY ; on confond souvent un tel segment avec son image. Le groupe Γ est *hyperbolique* s'il existe une constante $\delta_S \geq 0$ telle que pour tout triplet d'éléments γ_i ($i = 1, 2, 3$) et tout triplet de segments géodésiques c_i les connectant deux à deux, tout point de chacun des c_i est à distance inférieure à δ_S de la réunion des deux autres (finesse des triangles géodésiques). Cette propriété ne dépend pas du choix de la partie génératrice S (même si la valeur de δ_S en dépend). Les groupes hyperboliques, introduits par M. GROMOV dans [Gr81-84-87-93], jouissent de nombreuses propriétés remarquables et ont fait l'objet de nombreux travaux (voir par exemple [Gh90, GH90, CDP90, Al91] pour les fondements de la théorie) : on peut dire aujourd'hui qu'ils sont bien compris. L'un des thèmes essentiels de cet exposé est qu'en un certain sens la majorité des groupes de présentation finie sont hyperboliques.

Choisissons deux entiers $g \geq 2$ et $r \geq 1$ et considérons les présentations de groupes à g générateurs et r relateurs de la forme $\Gamma = \langle a_1, a_2, \dots, a_g \mid m_1, m_2, \dots, m_r \rangle$ où les m_j sont des mots cycliquement réduits en les lettres $a_i^{\pm 1}$ (un mot est *réduit* s'il ne contient pas deux lettres consécutives inverses et *cycliquement réduit* si de plus la première et la dernière lettre ne sont pas inverses). Si l'on fixe g et les longueurs l_j des relateurs m_j ($j = 1, \dots, r$), il n'y a qu'un nombre fini $N(g; l_1, \dots, l_r)$ de telles présentations. Notons $N_{hyp}(g; l_1, \dots, l_r)$ le nombre de celles qui définissent un groupe hyperbolique *non élémentaire* (c'est-à-dire infini et ne contenant pas de sous-groupe infini cyclique d'indice fini). Le théorème suivant a été énoncé par M. GROMOV dans [Gr87] puis démontré indépendamment par C. CHAMPETIER [Ch91-95] et Y. OL'SHANSKIĬ [Ols92a].

THÉORÈME A. – *La probabilité $N_{hyp}(g; l_1, \dots, l_r)/N(g; l_1, \dots, l_r)$ pour qu'une présentation de groupe à g générateurs et r relateurs définisse un groupe hyperbolique non élémentaire tend vers 1 lorsque les longueurs l_1, \dots, l_r des relateurs tendent vers l'infini.*

Nous verrons que la difficulté principale dans la preuve de ce théorème est dans le cas où les l_j ont des ordres de grandeur différents. Une autre approche a été proposée par la suite

par M. GROMOV dans [Gr93]. Il s'agit encore de fixer le nombre de générateurs mais de *faire tendre le nombre de relateurs vers l'infini, en les gardant tous de la même longueur l* (tendant également vers l'infini). Plus précisément, choisissons encore un entier $g \geq 2$ et une constante $c \geq 1$ et considérons l'ensemble $S(l-c, l+c)$ des mots réduits en les lettres $a_i^{\pm 1}$ ($i = 1, \dots, g$) dont les longueurs sont comprises entre $l-c$ et $l+c$. Chaque partie $R \subset S(l-c, l+c)$ peut être considérée comme un ensemble de relateurs et définit donc un groupe, quotient du groupe libre par le sous-groupe normal engendré par R . Fixons un réel $d \in [0, 1]$ et une constante $c' > 1$ et considérons l'ensemble des parties $R \subset S(l-c, l+c)$, dites de *densité d* , dont le cardinal est compris entre $c'^{-1}|S(l-c, l+c)|^d$ et $c'|S(l-c, l+c)|^d$ (nous notons $|X|$ le cardinal d'un ensemble fini X). La preuve du théorème suivant a été esquissée par M. GROMOV dans [Gr93] puis précisée par Y. OLLIVIER [Oll03a-c].

THÉORÈME B. – *Si $d > 1/2$, la probabilité pour qu'une partie $R \subset S(l-c, l+c)$ de densité d définisse le groupe trivial tend vers 1 lorsque la longueur l tend vers l'infini.*

– *Si $d < 1/2$, la probabilité pour qu'une partie $R \subset S(l-c, l+c)$ de densité d définisse un groupe hyperbolique non élémentaire tend vers 1 lorsque la longueur l tend vers l'infini.*

Ces théorèmes ont d'importantes généralisations dans lesquelles on remplace le groupe libre engendré par les a_i par un groupe hyperbolique non élémentaire quelconque. Il s'agit alors de considérer le quotient de ce groupe par le sous-groupe normalement engendré par un nombre fini d'éléments de grandes longueurs. En itérant le procédé on construit une suite de groupes hyperboliques et un passage à la limite donne des exemples intéressants de groupes de *type fini*. Une manière agréable de présenter ce genre de limite consiste à introduire une topologie sur l'ensemble $\mathcal{G}r_g$ des sous-groupes normaux du groupe libre F_g à g générateurs ou, ce qui revient au même, sur l'ensemble des groupes équipés d'une famille génératrice à g éléments. Dans cette topologie, deux sous-groupes normaux sont proches si leurs intersections avec une grande boule de F_g coïncident. Cette topologie (introduite par R. GRIGORCHUK [Gri85]) fait de $\mathcal{G}r_g$ un espace métrisable compact. Soit $\mathcal{H}yp_g \subset \mathcal{G}r_g$ la partie définie par les groupes hyperboliques non élémentaires (dénombrable car ceux-ci sont de présentation finie) et $\overline{\mathcal{H}yp}_g$ son adhérence dans $\mathcal{G}r_g$. Enfin, on note $\overline{\mathcal{H}yp}_g^{st}$ l'adhérence des groupes hyperboliques non élémentaires sans torsion. Le théorème suivant est exprimé dans le vocabulaire de C. CHAMPETIER [Ch91-00] mais on trouverait des énoncés du même genre dans l'article de Y. OLSHANSKIÏ [Ols92b].

THÉORÈME C. – *Il existe un G_δ dense dans l'adhérence des groupes hyperboliques non élémentaires $\overline{\mathcal{H}yp}_g$ formé de groupes infinis dont tous les éléments sont d'ordre fini.*

Il existe un G_δ dense dans $\overline{\mathcal{H}yp}_g^{st}$ formé de groupes infinis Γ qui :

- *possèdent la propriété (T) de KAZHDAN (voir [HV89]) et sont donc non moyennables,*
- *ne contiennent pas de sous-groupe non abélien libre,*
- *n'ont aucun quotient fini non trivial,*
- *possèdent une partie génératrice à deux éléments.*

Enfin, nous donnerons une idée générale de la démonstration par M. GROMOV du théorème suivant [Gr02].

THÉORÈME D. – *Il existe un groupe de présentation finie Γ qui ne se plonge pas de manière uniforme dans un espace de HILBERT \mathcal{H} , autrement dit pour lequel il n'existe aucun plongement $i : \Gamma \rightarrow \mathcal{H}$ vérifiant $F(d_S(\gamma_1, \gamma_2)) \leq \|i(\gamma_1) - i(\gamma_2)\| \leq d_S(\gamma_1, \gamma_2)$ pour tous γ_1, γ_2 dans Γ , avec $F : \mathbb{N} \rightarrow \mathbb{N}$ tendant vers l'infini à l'infini.*

L'intérêt de ce concept de plongement provient du fait qu'un théorème de YU affirme que tous les groupes qui se plongent uniformément dans un espace de HILBERT vérifient la conjecture de BAUM-CONNES « grossière » (coarse) (et donc celle de NOVIKOV) [Yu00]. Ces groupes sont aussi caractérisés par l'existence d'une action moyennable sur un espace compact [HR00] ; c'est le cas pour tous les groupes hyperboliques [Se92]. On pourra consulter à ce sujet l'exposé de G. SKANDALIS dans ce séminaire [Sk00]. Le théorème D permet la construction de contre-exemples à des versions généralisées de la conjecture de BAUM-CONNES [HLS02].

Dans cet exposé, la lettre l désignera toujours un entier destiné à tendre vers l'infini. Un événement aléatoire dépendra de l et se réalisera avec une probabilité dépendant également de l . Lorsque cette probabilité tend vers 1 quand l tend vers l'infini, nous dirons que l'événement se réalise très probablement.

2. MODÈLE À DENSITÉ

Densités

Commençons par donner une définition simple qui permet d'obtenir des estimations intuitives sur la combinatoire d'ensembles finis dont la taille tend vers l'infini. Soit X un ensemble fini de référence, dont le cardinal tendra par la suite vers l'infini. Si A est un ensemble fini non vide, nous dirons (suivant [Gr93]) que la *densité* de A (sous-entendu par rapport à X) est $dens(A) = \log |A| / \log |X|$. Lorsque A est une partie de X , sa *codensité* est définie par $codens(A) = 1 - dens(A)$. Si X est un espace vectoriel de dimension finie sur un corps fini et A est un sous-espace vectoriel, alors la densité est bien sûr le rapport des dimensions $\dim(A) / \dim(X)$. La codimension de l'intersection de deux sous-espaces en position générale est la somme de leurs codimensions, sauf si cette somme est supérieure à la dimension ambiante, auquel cas l'intersection est triviale. Il est remarquable que cette propriété s'étende aux parties d'un ensemble fini, sans aucune structure additionnelle, lorsque le cardinal tend vers l'infini. Plus précisément, fixons deux nombres cod_1 et cod_2 dans l'intervalle $[0, 1]$ et un $\epsilon > 0$. Considérons l'ensemble (fini) des couples de parties A_1, A_2 de X dont les codensités sont respectivement dans les intervalles $[cod_1 - \epsilon, cod_1 + \epsilon]$ et $[cod_2 - \epsilon, cod_2 + \epsilon]$. Parmi ces couples de parties (A_1, A_2) on peut compter la proportion de ceux qui sont tels que la codensité de l'intersection vérifie $|codens(A_1 \cap A_2) - (cod_1 + cod_2)| \leq 3\epsilon$. Dans cette dernière inégalité, nous convenons

d'attribuer n'importe quelle codensité supérieure ou égale à 1 à l'ensemble vide. Il se trouve que lorsque le cardinal de X tend vers l'infini, cette proportion tend (rapidement) vers 1. En particulier, si $\text{cod}_1 + \text{cod}_2 > 1$ et si ϵ est assez petit, la probabilité pour que $A_1 \cap A_2$ soit vide tend vers 1. La preuve est bien sûr facile mais nous en retiendrons qu'il est souvent possible d'estimer le nombre de solutions d'« équations aléatoires » dans un ensemble fini de grande taille en « comptant les équations ». Voici un autre exemple : la non injectivité d'une application $f : A \rightarrow X$ revient à l'existence de solutions non triviales à l'équation $f(x) = f(y)$ à deux inconnues dans A et à valeurs dans X ; si d désigne la densité de A , on s'attend donc à ce qu'une application aléatoire f soit injective si $2d - 1 < 0$ et non injective si $2d - 1 > 0$. Formellement, cela signifie que si les cardinaux de deux ensembles finis A_l et X_l tendent vers l'infini avec une limite $d = \lim_l \log |A_l| / \log |X_l|$, et si $2d - 1 < 0$ (resp. $2d - 1 > 0$) alors une application $f : A_l \rightarrow X_l$ est *très probablement* injective (resp. non injective). On reconnaît le *principe des tiroirs probabiliste* (aussi connu comme le paradoxe des dates d'anniversaire) : si on met $l^{1/2+\epsilon}$ objets au hasard dans l tiroirs, très probablement l'un des tiroirs contient au moins deux objets.

Dans le théorème B, nous utilisons un concept de « partie R de $S(l - c, l + c)$ de densité d » qui ne coïncide pas exactement avec celui que nous venons de définir mais il est clair que la densité de R dans notre nouveau sens tend vers d quand l tend vers l'infini et ce changement de terminologie n'a aucune importance.

Diagrammes

Avant d'esquisser la preuve du théorème B, il nous faut faire quelques rappels. Soit Γ un groupe possédant une présentation de la forme $\langle a_1, \dots, a_g \mid m_1, \dots, m_r \rangle$. Par définition, cela signifie que Γ est le quotient du groupe libre F_g de base les a_i par le sous-groupe normal engendré par les m_j . Ainsi, un élément γ de F_g définit l'élément neutre de Γ si et seulement si on peut l'écrire comme un produit de conjugués des $m_j^{\pm 1}$. On doit à VAN KAMPEN une représentation topologique d'une telle situation en termes de *diagrammes*. Il s'agit d'un graphe fini connexe \mathcal{D} plongé dans le plan. De plus, chaque arête orientée de \mathcal{D} est étiquetée par un mot réduit en les $a_i^{\pm 1}$ de telle sorte que l'étiquette est inversée lorsque l'on renverse l'orientation d'une arête. Sur le bord orienté de chaque face f , c'est-à-dire de chaque composante bornée du complémentaire du graphe, on lit un mot $m(f)$ en les $a_i^{\pm 1}$ (défini à permutation cyclique près). Sur le bord du diagramme, c'est-à-dire sur le bord de la composante non bornée, on lit également un mot m et il est clair que le diagramme permet d'exprimer m comme un produit dans F_g de conjugués des $m(f)$. Le résultat (élémentaire) de VAN KAMPEN consiste en la réciproque : si un mot réduit m en les $a_i^{\pm 1}$ définit l'élément neutre de Γ , il existe un diagramme dont les mots associés aux faces sont des $m_j^{\pm 1}$ et dont le mot lu sur le bord est m (voir par exemple [LS01]). Nous ne considérerons que des diagrammes *réduits*, c'est-à-dire tels que deux faces adjacentes ne portent pas des mots inverses lorsqu'on les lit dans le sens direct à

partir d'un point commun (dans ce cas on peut retirer ces deux faces et produire par recollement un diagramme possédant le même bord et moins de faces). La longueur du mot m s'appelle le *périmètre* du diagramme et se note $|\partial\mathcal{D}|$; le nombre de faces est l'*aire* et se note $|\mathcal{D}|$.

Théorème B

Une caractérisation importante des groupes hyperboliques est l'*inégalité isopérimétrique linéaire*. Un groupe $\Gamma = \langle a_1, \dots, a_g \mid m_1, \dots, m_r \rangle$ est hyperbolique si et seulement s'il existe une constante $C > 0$ telle que tout mot réduit m en les $a_i^{\pm 1}$, trivial dans Γ , est le bord d'un diagramme dont l'aire est bornée par C fois son périmètre (qui est la longueur de m). On trouvera une démonstration élémentaire de ce fait dans [Al91]. L'hyperbolicité d'un groupe possède en fait un caractère semi-local : c'est le principe local/global de CARTAN-HADAMARD-GROMOV qu'on peut exprimer de la manière suivante. Pour chaque $k > 0$, il existe des entiers K et $k' > 0$ tels que si Γ est un groupe de présentation finie dont les relateurs sont de longueur l et si tous les diagrammes de Γ d'aire inférieure à K vérifient $kl|\mathcal{D}| \leq |\partial\mathcal{D}|$ alors tous les diagrammes vérifient $k'l|\mathcal{D}| \leq |\partial\mathcal{D}|$ (voir [Gr87-93, Oll03a-c, Pa96]). Ainsi, une information sur un nombre fini de diagrammes peut permettre de conclure à l'hyperbolicité ; c'est la base des algorithmes de [Pa96].

Nous pouvons maintenant esquisser une preuve du théorème B. Commençons par la partie facile et supposons $d > 1/2$. À chaque élément de $S(l-c, l+c)$, c'est-à-dire à chaque mot réduit en les $a_i^{\pm 1}$ de longueur comprise entre $l-c$ et $l+c$, associons le mot obtenu en oubliant la première lettre. Ceci définit une application f de $S(l-c, l+c)$ vers $S(l-c-1, l+c-1)$ entre deux ensembles finis dont le rapport des cardinaux est borné (par $2g$), indépendamment de l . Si la densité d'une partie aléatoire $R \subset S(l-c, l+c)$ est supérieure à $1/2$, on peut affirmer que lorsque l tend vers l'infini, *très probablement*, la restriction de f à R n'est pas injective. Si deux relateurs de R ont la même image par f , leurs premières lettres sont identifiées dans le groupe défini par cette présentation. Il est facile d'en déduire que, *très probablement*, tous les générateurs $a_i^{\pm 1}$ sont identifiés dans le groupe quotient, de sorte que ce dernier a au plus deux éléments. Très probablement, les relateurs ne sont pas tous de longueur paire et le groupe est en fait trivial.

Supposons maintenant $d < 1/2$ et montrons que le groupe défini par R est très probablement hyperbolique non élémentaire. Pour cela, choisissons un graphe planaire connexe fini \mathcal{G} et cherchons à en étiqueter les arêtes de manière à obtenir un diagramme de VAN KAMPEN pour R , *i.e.* de façon à ce que les mots lus sur les bords des faces soient des (permutations cycliques d') éléments de R ou de leurs inverses. Imaginons pour simplifier que le bord $\partial\mathcal{G}$ soit une courbe simple. On peut considérer le problème comme une « équation » dans laquelle on cherche pour chaque face un élément de R avec des conditions de compatibilité le long des arêtes communes à deux faces. L'« inconnue »

est donc un élément de $R^{|\mathcal{G}|}$ et chaque arête commune à deux faces donne une contrainte. Ici, l'ensemble de référence X peut être choisi comme étant $S(l - c, l + c)$. Remarquons que la codensité dans X de l'ensemble des mots ayant l_0 lettres fixées est l_0/l (à $O(1/l)$ près, mais nous négligeons ce « détail »). On peut calculer la densité de l'ensemble des solutions pour un R aléatoire (*i.e.* toujours lorsque l tend vers l'infini) : on trouve bien sûr $d|\mathcal{G}| - \sum l_i/l$ où la somme est étendue sur toutes les longueurs l_i des frontières communes entre faces. Puisque chaque face a (environ) l arêtes dans son bord et que chaque arête intérieure est dans le bord de deux faces, on peut exprimer cette densité comme $d|\mathcal{G}| - (l|\mathcal{G}| - |\partial\mathcal{G}|)/2l = (d - \frac{1}{2})|\mathcal{G}| + |\partial\mathcal{G}|/2l$. Par conséquent, s'il y a « trop d'équations », c'est-à-dire si $|\partial\mathcal{G}| < l(1 - 2d)|\mathcal{G}|$, l'ensemble des solutions est très probablement vide. Cela signifie que si l'on considère l'ensemble fini de tous les graphes \mathcal{G} dont l'aire est inférieure à une constante K , très probablement, seuls ceux qui vérifient l'inégalité inverse $|\partial\mathcal{G}| \geq l(1 - 2d)|\mathcal{G}|$ apparaissent comme diagramme du groupe défini par R . Le principe de CARTAN-HADAMARD-GROMOV peut alors s'appliquer et, en choisissant K assez grand, on conclut que le groupe associé à R est très probablement hyperbolique.

Avant de montrer que le groupe est infini, rappelons une construction classique. Soit $\Gamma = \langle a_1, \dots, a_g \mid m_1, \dots, m_r \rangle$ un groupe de présentation finie. On recolle r disques topologiques le long de leurs bords sur un bouquet de g cercles, les applications de recollement étant données par les mots m_j . Le groupe fondamental du 2-complexe fini ainsi obtenu est bien sûr isomorphe à Γ et son revêtement universel est le 2-complexe de CAYLEY associé à la présentation, dont le 1-squelette est le graphe de CAYLEY (si les éléments de S ne sont pas d'ordre 2, ce que nous supposerons dans la suite).

L'argument précédent peut s'appliquer à des « diagrammes sphériques », c'est-à-dire à des graphes étiquetés tracés sur la sphère et tels que les mots lus sur *toutes* les faces soient des conjugués cycliques d'éléments de R ou de leurs inverses. Ces nouveaux diagrammes n'ayant pas de bord, le calcul précédent montre que très probablement un tel diagramme ne se rencontre pas dans le groupe défini par R . Ceci entraîne que le 2-complexe de CAYLEY est 2-connexe et donc contractile. Le groupe associé à R agit donc sur un complexe contractile de dimension 2 et *sa dimension cohomologique est donc égale à 2*. En particulier, ce groupe est *sans torsion* et *non élémentaire* car sa caractéristique d'EULER-POINCARÉ est égale à $1 - g + |R|$.

L'esquisse de preuve que nous venons de présenter n'est qu'une esquisse... De nombreux problèmes « techniques » se présentent dont le principal provient du fait que nous avons supposé (implicitement !) dans notre calcul de densité que les relateurs placés sur les faces du diagramme sont différents, de façon à pouvoir traiter les équations comme « indépendantes ». Pour obtenir une « vraie preuve », il faut des estimations beaucoup plus soigneuses qui sont parfois délicates. Y. OLLIVIER a rédigé une preuve complète dans [Oll03b].

Quotients aléatoires d'un groupe hyperbolique

Si la densité de R est strictement inférieure à $1/2$, le groupe Γ engendré par les a_i et soumis aux relations R est très probablement hyperbolique non élémentaire ; on aimerait itérer la construction précédente et ajouter de nouvelles relations R' dont les longueurs sont grandes par rapport à l . Nous allons décrire rapidement cette généralisation.

Soit Γ un groupe engendré par une partie finie symétrique S à $2g$ éléments. Considérons le nombre $f(2n)$ de mots réduits dans les éléments de S de longueur $2n$ qui représentent l'élément neutre de Γ . La « densité des mots triviaux », c'est-à-dire la limite $\eta = \lim_{n \rightarrow +\infty} \frac{1}{2n} \log_{2g-1} f(2n)$ existe : c'est la *cocroissance* du groupe Γ (relativement à S), reliée à la vitesse de fuite à l'infini de la marche aléatoire sur Γ . Lorsque Γ n'est pas librement engendré par S , on a toujours $\eta > 1/2$ (car si w est un mot réduit trivial dans Γ , tous les mots de la forme $\gamma w \gamma^{-1}$ le sont également de sorte que $f(2n + |w|) \geq (2g - 1)^n$). Pour cette raison, on convient de définir la cocroissance du groupe libre par rapport à une base comme étant $1/2$.

Soit Γ un groupe hyperbolique non élémentaire sans torsion engendré par (a_1, a_2, \dots, a_g) et soit $\eta \geq 1/2$ la cocroissance correspondante. Soit d une densité dans l'intervalle $[0, 1]$ et $c, c' \geq 1$ deux constantes auxiliaires. On note toujours $S(l - c, l + c)$ l'ensemble des mots réduits dont la longueur est comprise entre $l - c$ et $l + c$. Une partie $R \subset S(l - c, l + c)$ est de densité d si son cardinal est compris entre $c'^{-1}|S(l - c, l + c)|^d$ et $c'|S(l - c, l + c)|^d$. On note $\langle R \rangle$ le sous-groupe normal de Γ engendré par la projection de R dans Γ . Le théorème suivant est dû à Y. OLLIVIER [Oll03a-c].

THÉORÈME.— *Si $d > 1 - \eta$, la probabilité pour qu'une partie $R \subset S(l - c, l + c)$ de densité d définisse un groupe quotient $\Gamma/\langle R \rangle$ trivial tend vers 1 quand la longueur l tend vers l'infini.*

Si $d < 1 - \eta$, la probabilité pour qu'une partie $R \subset S(l - c, l + c)$ de densité d définisse un groupe quotient $\Gamma/\langle R \rangle$ hyperbolique non élémentaire tend vers 1 quand la longueur l tend vers l'infini.

Ce théorème possède plusieurs variantes : au lieu de considérer des parties R dans $S(l - c, l + c)$, c'est-à-dire dans le groupe libre, on peut prendre des parties dans Γ formées d'éléments de norme l dans Γ . On obtient des énoncés analogues.

L'esprit de la preuve de ces théorèmes est proche de celui que nous avons décrit lorsque Γ est un groupe libre mais les difficultés techniques sont assez considérables. On considère d'abord une présentation de Γ sous la forme $\langle a_1, a_2, \dots, a_g \mid m_1, m_2, \dots, m_r \rangle$, de sorte que le groupe quotient $\Gamma/\langle R \rangle$ possède une présentation dont les relateurs sont d'une part les m_j et d'autre part les éléments de R . Un diagramme pour $\Gamma/\langle R \rangle$ possède donc deux sortes de faces. Si l est grand, les premières sont « petites » par rapport aux secondes. Si l est grand par rapport à la constante d'hyperbolicité δ_S du groupe Γ , la géométrie de la « partie petite » d'un diagramme est proche de celle d'un arbre (les espaces δ -hyperboliques s'approchent bien par des arbres [Gr87, GH90]). On peut alors

établir qu'un diagramme est constitué de « grandes faces » entourées par des zones fines, dont l'épaisseur est de l'ordre de $\log l$. On estime le nombre de manières d'étiqueter un diagramme avec des générateurs et les « équations de compatibilité » que nous avons rencontrées le long des arêtes communes à plusieurs faces doivent maintenant être remplacées par des « quasi-compatibilités » dues au fait que les grandes faces ne sont pas exactement jointives. C'est à ce niveau que la cocroissance intervient de manière naturelle puisqu'elle permet bien sûr d'estimer le nombre de mots qui sont petits quand on les projette dans Γ . Pour les (nombreux) détails de cette preuve, voir [Oll03a-c].

3. PETITE SIMPLIFICATION

La théorie des *groupes à petite simplification* tire son origine d'un article de 1949 par V.A. TARTAKOVSKII qui cherchait à généraliser la solution géométrique au problème des mots donnée par M. DEHN dans le cas du groupe fondamental d'une surface compacte. Initialement de nature très combinatoire, cette théorie a pris un aspect plus topologique grâce à R.C. LYNDON, P.E. SCHUPP et M. GREENDLINGER [LS01]. Plus récemment, M. GROMOV lui donne un aspect purement géométrique qui permet de vastes généralisations [Gr01a-01b-01c] que nous allons commencer à survoler dans cette partie.

Théorie classique

Considérons deux mots m_1 et m_2 en les lettres $a_1^{\pm 1}, \dots, a_g^{\pm 1}$ que nous supposons cycliquement réduits. Une *pièce* entre m_1 et m_2 est un mot p qui apparaît comme un segment initial dans l'écriture de conjugués cycliques m'_1 et m'_2 de $m_1^{\pm 1}$ et $m_2^{\pm 1}$ tels que m'_1 est *distinct* de m'_2 . Par exemple, le mot $\mathbf{a}_1\mathbf{a}_2$ est une pièce entre $a_2a_2\mathbf{a}_1\mathbf{a}_2a_1a_1a_2$ et $a_1a_1\mathbf{a}_2^{-1}\mathbf{a}_1^{-1}a_2$, entre $\mathbf{a}_1\mathbf{a}_2a_3\mathbf{a}_1\mathbf{a}_2$ et lui-même, mais pas entre $a_1a_2a_1a_2$ et lui-même. Pour $0 < \lambda < 1$, on dit qu'un ensemble de mots $\{m_1, \dots, m_r\}$ vérifie la condition de petite simplification $C'(\lambda)$ si toute pièce entre deux mots m_{j_1} et m_{j_2} est de longueur strictement inférieure à λ fois la plus petite des longueurs de m_{j_1} et de m_{j_2} . Dans ce cas, on dit que le groupe $\Gamma = \langle a_1, a_2, \dots, a_g \mid m_1, m_2, \dots, m_r \rangle$ vérifie la condition de petite simplification $C'(\lambda)$ (même s'il s'agit en fait d'une propriété de la présentation).

La remarque importante est que, dans un diagramme de VAN KAMPEN réduit relatif à une présentation, les composantes connexes de l'intersection de deux faces adjacentes sont (étiquetées par) des pièces. Si l'on suppose une condition $C'(1/6)$, il en résulte que chaque face intérieure au diagramme est entourée par au moins 7 de ces pièces. Un argument facile de caractéristique d'EULER permet alors de montrer que l'intersection d'au moins une des faces du diagramme avec le bord du diagramme a une composante connexe de longueur strictement supérieure à la moitié du périmètre de cette face. Si un mot w réduit en les $a_i^{\pm 1}$ est trivial dans Γ , il contient donc un sous-mot w' qui est le début d'un conjugué cyclique \overline{m} de l'un des $m_j^{\pm 1}$, de longueur strictement supérieure à la moitié de celle de \overline{m} . Remplaçant alors w' dans w par $\overline{m}^{-1}w'$, on obtient après réduction un nouveau mot w_1 , plus court que w , qui est encore trivial dans Γ et l'itération du procédé mène au

mot trivial : c'est l'*algorithme de DEHN* pour les groupes $C'(1/6)$, qui fut la motivation initiale pour l'introduction de ce concept. Une autre conséquence facile est l'inégalité isopérimétrique linéaire pour ces groupes : *les groupes à petite simplification $C'(1/6)$ sont donc hyperboliques*. Il s'agit en fait de groupes hyperboliques assez particuliers et la forme des triangles dans le graphe de CAYLEY est très facile à analyser [GH90, Ch94]. Dans [Gr01a], M. GROMOV montre que beaucoup de groupes $C'(1/6)$ agissent en fait par isométries sur un 2-complexe simplement connexe muni d'une métrique à courbure négative (*i.e.* vérifiant une condition $CAT(-k^2)$, voir [BH99]) mais cette construction ne s'étend malheureusement pas aux cas plus généraux que nous allons rencontrer dans la suite de ce texte.

Étant donnée une partie R de $S(l-c, l+c)$ de densité $d \in [0, 1]$ comme au paragraphe précédent, il est facile d'estimer la taille probable des plus grandes pièces entre les éléments de R : si $2d < \lambda$, le groupe défini par les relations R vérifie très probablement $C'(\lambda)$. On remarque donc que ceci donne une autre preuve du théorème B lorsque $d < 1/12$ mais que pour $1/12 \leq d < 1/2$ les groupes hyperboliques que nous avons construits grâce à ce théorème ne sont pas redevables de la théorie des groupes $C'(1/6)$.

Théorie relative : le théorème A

La théorie classique de la petite simplification est parfaitement adaptée pour étudier des groupes $\langle a_1, a_2, \dots, a_g \mid m_1, m_2, \dots, m_r \rangle$ dont on fixe le nombre de générateurs g et de relateurs r , lorsque les longueurs des relateurs tendent vers l'infini, à condition que ces longueurs gardent le même ordre de grandeur. Pour chaque $0 < \lambda < 1$, la condition $C'(\lambda)$ est alors très probablement satisfaite et le groupe est hyperbolique non élémentaire. Ce cas correspond d'ailleurs à celui où la densité d est nulle dans le théorème B. Un cas très analogue est celui où l'on choisit les r relateurs aléatoirement dans la boule de rayon l car dans cette boule, une grande proportion des éléments ont en fait une norme très proche de l (par la croissance exponentielle du groupe libre). Dans ce dernier modèle élémentaire, on peut cependant obtenir des informations algébriques plus précises sur les groupes aléatoires obtenus [Ar97-98, AO96, KS02].

Une difficulté sérieuse apparaît lorsque les m_j sont tous très longs mais de longueurs très différentes. Si par exemple, la longueur de m_2 est de l'ordre de l'exponentielle de celle de m_1 , il est probable que m_1 soit un sous-mot de m_2 et aucune condition de petite simplification n'est satisfaite. C'est la raison pour laquelle il est nécessaire d'introduire un concept de *petite simplification relativement à un groupe hyperbolique*. Suivant des indications succinctes de M. GROMOV ([Gr87, 5.5.F]), cette théorie relative a été mise au point indépendamment par C. CHAMPETIER, T. DELZANT et Y. OLSHANSKIÏ [Ch91-94, De96, Ols92a-92b].

Soit Γ un groupe hyperbolique non élémentaire engendré par une partie finie symétrique S , et $\delta_S \geq 0$ la constante d'hyperbolicité correspondante. Un mot m de longueur $|m|$ en les éléments de S est *géodésique* ou *minimal* si l'élément γ de Γ qu'il définit est de norme

$|\gamma|_S = |m|$, et *minimal dans sa classe de conjugaison* si de plus tous les conjugués de γ ont une norme supérieure ou égale à $|m|$. Si m_1 et m_2 sont minimaux dans leur classe de conjugaison, une *a-pièce* est la donnée de deux sous-mots p_1 et p_2 de conjugués cycliques de m_1 et m_2 qui sont *a-proches* dans le sens suivant : il existe deux mots u, v de longueur inférieure à a tels que $p_1 = up_2v$ dans Γ . Il faut cependant exclure de la définition un cas trivial : si m'_1 et m'_2 désignent les conjugués cycliques qui commencent par p_1 et p_2 , on demande que les éléments m'_1 et um'_2u^{-1} de Γ ne soient pas égaux. Il est important de remarquer que la valeur de a n'est pas très importante pour des éléments de grandes longueurs : un argument élémentaire montre que si m_1 et m_2 partagent une *a-pièce* de longueur t , alors ils partagent une $2\delta_S$ -pièce de longueur $t - 2a$, ce qui justifie de se limiter aux $2\delta_S$ -pièces.

Soit $0 < \lambda < 1$ et R un ensemble de mots minimaux dans leur classe de conjugaison. On dit que R vérifie une *condition de petite simplification* $C'(\lambda)$ relativement à (Γ, S) si d'une part tous les éléments de R sont assez grands (de longueur supérieure à $cst \cdot \delta_S$ pour une constante universelle cst dont la valeur précise n'a que peu d'intérêt) et si, d'autre part, pour tout couple de mots m_1, m_2 , toute $2\delta_S$ -pièce entre m_1 et $m_2^{\pm 1}$ est de longueur inférieure à $\lambda \inf(|m_1|, |m_2|)$. Nous énonçons le résultat principal de cette théorie relative sous la forme donnée par T. DELZANT [De96].

THÉORÈME – *Si $\lambda < 1/8$ et si R vérifie la condition de petite simplification $C'(\lambda)$, alors le quotient $\Gamma/\langle R \rangle$ de Γ par le sous-groupe normal engendré par (la projection dans Γ de) R est hyperbolique (de constante d'hyperbolicité inférieure à $cst \cdot \sup_{m \in R} |m|$). De plus, la boule de rayon $\frac{1}{4} \inf_{m \in R} |m| - cst \cdot \delta_S$ dans Γ s'injecte dans le quotient $\Gamma/\langle R \rangle$.*

La démonstration de ce théorème est fondée sur une propriété importante des espaces métriques δ -hyperboliques (X, d) [Gr87] : pour toute partie finie $E \subset X$, il existe un arbre métrique fini (T, d') et une application $i : E \rightarrow T$ telle que pour tout (x, y) , on a $d(x, y) - 2\delta \log_2(|E| - 2) \leq d'(i(x), i(y)) \leq d(x, y)$.

La *démonstration du théorème A* de généricité des groupes hyperboliques parmi les groupes dont on fixe le nombre g de générateurs et le nombre r de relateurs est un corollaire facile de cette théorie relative. Supposons par exemple, $r = 2$ et $l_1 \leq l_2$. Si $l_2 \leq l_1^3$ et l_1 tend vers l'infini, la probabilité pour que le groupe $F_g/\langle m_1, m_2 \rangle$ soit à petite simplification $C'(1/100)$ tend vers 1. Si par contre l_1 tend vers l'infini et $l_2 \geq l_1^3$, on procède en deux étapes. Tout d'abord, la probabilité pour que le groupe $F_g/\langle m_1 \rangle$ soit $C'(1/100)$ tend vers 1. Puis on montre (assez facilement) que la probabilité pour que (la réduction géodésique de) $\{m_2\}$ soit à petite simplification $C'(1/100)$ relativement à $F_g/\langle m_1 \rangle$ tend également vers 1, de sorte que $F_g/\langle m_1, m_2 \rangle = (F_g/\langle m_1 \rangle)/\langle m_2 \rangle$ est également hyperbolique. Bien entendu, les groupes hyperboliques ainsi construits sont assez particuliers : C. CHAMPETIER montre aussi (pour $r = 2$) que lorsque les longueurs l_i tendent vers l'infini, les groupes obtenus sont très probablement *sans torsion et de dimension cohomologique 2*.

4. ESPACE DES GROUPES DE TYPE FINI

L'ensemble des classes d'isomorphisme de groupes de type fini (*i.e.* engendrés par une partie finie) est non dénombrable (voir [Ne37] pour une jolie preuve). On appelle *groupe marqué* à g générateurs la donnée d'un sous-groupe normal N du groupe libre à g générateurs F_g , ou d'une surjection $F_g \rightarrow F_g/N$ de F_g sur un de ses quotients. De manière équivalente, il s'agit d'un groupe dont on a fixé une famille génératrice (a_1, \dots, a_g) . L'ensemble $\mathcal{G}r_g$ de ces groupes marqués est naturellement muni d'une topologie métrisable qui en fait un compact totalement discontinu : deux groupes marqués sont proches si les noyaux des surjections correspondantes ont la même intersection avec une grande boule dans F_g . Deux groupes marqués proches ont des graphes de CAYLEY qui sont « les mêmes » dans une grande boule. Dans cette topologie, les groupes de présentation finie forment une partie dénombrable dense. L'étude de cet espace n'en est qu'à ses débuts : nous allons décrire rapidement quelques résultats de C. CHAMPETIER [Ch91-00] et M. GROMOV [Gr93].

Relation d'isomorphisme

Un groupe de type fini peut être engendré de nombreuses manières par une partie finie, de sorte qu'il existe une relation d'équivalence naturelle, l'isomorphisme \simeq , sur ces espaces $\mathcal{G}r_g$, dont l'ensemble quotient est l'ensemble des classes d'isomorphisme de groupes à g générateurs. C'est ce quotient qu'on aimerait comprendre.

Chaque $\mathcal{G}r_g$ se plonge dans $\mathcal{G}r_{g+1}$ comme un ouvert fermé (par composition à la source avec la surjection de F_{g+1} sur F_g qui envoie le dernier générateur sur l'identité). On peut donc considérer la réunion croissante $\mathcal{G}r$ des $\mathcal{G}r_g$: c'est l'espace localement compact des groupes marqués de type fini, ou encore l'espace des sous-groupes normaux de $F_\infty = F(a_1, \dots)$ qui contiennent les a_i avec i assez grand. Si un groupe est engendré par (a_1, a_2, \dots, a_g) , il est également engendré par $(id, a_1, a_2, \dots, a_g)$, par $(a_1^{-1}, a_2, \dots, a_g)$, par $(a_1 a_2, a_2, \dots, a_g)$ et par (a_2, a_1, \dots, a_g) . Chacun de ces changements de famille génératrice peut être vu comme un homéomorphisme de $\mathcal{G}r$ et un théorème classique de TIETZE peut s'interpréter en disant que les orbites du groupe T engendré par ces quatre homéomorphismes de $\mathcal{G}r$ sont précisément les classes d'isomorphisme [LS01]. Une fonction $i : \mathcal{G}r \rightarrow E$ définit un invariant de groupe (*i.e.* indépendant de la famille génératrice) si et seulement si elle est invariante par l'action de T . Un tel invariant est particulièrement intéressant si E est un espace borélien standard (isomorphe à un espace métrique séparable muni de la tribu de ses boréliens) et si i est une application borélienne. Par exemple, le rang d'un groupe (cardinal minimum d'une famille génératrice) où les nombres de BETTI sont de tels invariants. Il se trouve que l'action de T sur $\mathcal{G}r$ est trop compliquée pour que le quotient soit standard : voici un résultat de [Ch91-00] (voir aussi [SZ94]).

THÉORÈME.— *Il n'existe pas d'application borélienne $i : \mathcal{G}r \rightarrow E$ dans un espace borélien standard qui sépare les classes d'isomorphisme, i.e. telle que $i(x) = i(y)$ si et seulement si $x \simeq y$.*

La démonstration de ce théorème est une jolie application de la théorie de la petite simplification classique. Soit ϕ un automorphisme du groupe libre F_2 et w un élément de F_2 . Pour chaque suite $e : \mathbf{Z} \rightarrow \{0, 1\}$, on considère le groupe marqué Γ_e qui est le quotient de F_2 par le sous-groupe normal engendré par les $\phi^i(w)$ pour les i tels que $e(i) = 1$. Ceci définit une application (borélienne) de $\{0, 1\}^{\mathbf{Z}}$ dans $\mathcal{G}r$. Il est clair que si on définit les suites décalées $e_k(i) = e(i + k)$, tous les groupes marqués Γ_{e_k} sont dans la même classe d'isomorphisme. Si l'on choisit ϕ défini par $\phi(a_1) = a_1 a_2$ et $\phi(a_2) = a_1$, il n'est pas difficile de trouver des mots w assez compliqués pour s'assurer que la famille $\{\phi^i(w)\}$ ($i \in \mathbf{N}$) vérifie la condition de petite simplification $C'(1/6)$. Ceci permet de montrer que Γ_e et $\Gamma_{e'}$ ne sont isomorphes que si e et e' sont décalées l'une de l'autre. On obtient donc un plongement de $\{0, 1\}^{\mathbf{Z}}$ dans $\mathcal{G}r$ et la relation induite par \simeq sur $\{0, 1\}^{\mathbf{Z}}$ est celle définie par le décalage. Le théorème résulte du fait que les fonctions boréliennes invariantes par décalage ne séparent pas les orbites (il suffit par exemple d'utiliser l'ergodicité du décalage pour la mesure de BERNOULLI $(1/2, 1/2)$).

Dans [TV99], S. THOMAS et B. VELICKOVIC vont plus loin et montrent que la relation \simeq est *universelle*, c'est-à-dire que pour toute relation d'équivalence \equiv borélienne à classes dénombrables sur un borélien standard X , il existe une application borélienne $f : X \rightarrow \mathcal{G}r$ telle que $f(x) \simeq f(y)$ si et seulement si $x \equiv y$. Autrement dit, la relation d'isomorphisme \simeq est aussi compliquée que possible...

Puisque l'action de T sur $\mathcal{G}r$ n'a pas de bon quotient mesurable, on pourrait tenter une description « ergodique » de cet espace quotient. Malheureusement, on ne sait pas s'il existe une mesure de RADON sur $\mathcal{G}r$ qui soit quasi-invariante par T (*i.e.* telle que la collection des boréliens de mesure nulle soit invariante par T). La non existence d'une mesure *invariante* paraît probable mais elle ne semble pas établie. L'approche topologique paraît donc plus adaptée mais là encore notre compréhension n'est que très partielle. On trouve une discussion intéressante des problèmes de cette nature dans [Gr93].

Adhérence des groupes hyperboliques : le théorème C

Les groupes marqués de présentation finie forment évidemment une partie dénombrable dense dans $\mathcal{G}r$. Les propriétés des groupes hyperboliques, extrêmement stables grâce à la théorie de la petite simplification relative, permettent d'obtenir des informations sur leur adhérence dans l'espace des groupes de type fini. On note $\mathcal{H}yp$ et $\mathcal{H}yp^{st}$ les parties de $\mathcal{G}r$ définies par les groupes hyperboliques non élémentaires et hyperboliques non élémentaires sans torsion, et $\overline{\mathcal{H}yp}$, $\overline{\mathcal{H}yp}^{st}$ leurs adhérences dans $\mathcal{G}r$.

Étant donné un élément γ d'ordre infini d'un groupe hyperbolique marqué non élémentaire Γ , on montre sans (trop de) difficulté qu'il possède des puissances arbitrairement grandes γ^n telles que le singleton $\{\gamma^n\}$ vérifie une condition de petite simplification $C'(1/10)$ relativement à Γ . Ces groupes $\Gamma_n = \Gamma / \langle \gamma^n \rangle$ sont donc hyperboliques non élémentaires et convergent vers le groupe Γ . Par conséquent, pour chaque élément w de F_g , l'ensemble des groupes marqués à g générateurs tels que (la projection de) l'élément

w soit d'ordre fini, est un ouvert qui rencontre \mathcal{Hyp} sur une partie dense dans $\overline{\mathcal{Hyp}}$. Le théorème de BAIRE implique alors :

THÉORÈME.— *Il existe un G_δ dense dans l'adhérence des groupes hyperboliques non élémentaires $\overline{\mathcal{Hyp}}$ formé de groupes dont tous les éléments sont d'ordre fini.*

L'existence de groupes infinis de type fini dont tous les éléments sont de torsion est déjà remarquable. Dans [IO96], S. IVANOV et Y. OLSHANSKIÏ montrent en fait que tout groupe hyperbolique non élémentaire possède un quotient infini de torsion dont les ordres des éléments sont *bornés*. Une approche plus géométrique de ce résultat est en cours de rédaction par T. DELZANT et M. GROMOV.

Si Γ est un groupe quelconque, son *centre virtuel* $Z^{virt}(\Gamma)$ est le sous-groupe normal formé des éléments qui n'ont qu'un nombre fini de conjugués. Lorsque Γ est hyperbolique non élémentaire, $Z^{virt}(\Gamma)$ est un groupe fini et on s'assure que $\Gamma/Z^{virt}(\Gamma)$ est un groupe hyperbolique dont le centre virtuel est trivial. L'existence d'un centre virtuel non trivial est source de difficultés techniques. Notons en effet $C(\Gamma)$ le centralisateur dans Γ de $Z^{virt}(\Gamma)$ (d'indice fini dans Γ) et soit γ un élément de $\Gamma - C(\Gamma)$. Alors $\{\gamma\}$ ne peut pas vérifier de condition de petite simplification. En effet, il existe par définition un élément z du centre virtuel tel que $\gamma z \gamma^{-1} \neq z$ de sorte que si $a = \sup(|z|_S, |\gamma z \gamma^{-1}|_S)$, on peut dire que γ partage une a -pièce avec lui-même. Dans la construction précédente d'éléments de torsion par exemple, pour s'assurer que $\{\gamma^n\}$ est à petite simplification, il fallait choisir n tel que $\{\gamma^n\}$ soit dans $C(\Gamma)$. En général, il est agréable de supposer que Γ est à centralisateurs cycliques, *i.e.* le centralisateur de chaque élément non trivial est cyclique, fini ou infini. Cette propriété garantit la trivialité du centre virtuel ; elle est par exemple satisfaite si le groupe hyperbolique est sans torsion. On note $\overline{\mathcal{Hyp}}^{cc}$ l'adhérence des groupes hyperboliques non élémentaires dont les centralisateurs de tous les éléments non triviaux sont cycliques, finis ou infinis : il s'agit d'un ensemble de Cantor.

Nous pouvons maintenant esquisser la preuve du théorème C, donnée dans ses grandes lignes dans [Gr87] puis mise au point dans [Ols92a-92b] et [Ch91-00] sous des formes légèrement différentes. Nous suivons ici [Ch00] qui travaille directement dans l'espace \mathcal{Gr} . Soit Γ un groupe hyperbolique non élémentaire et sans torsion (resp. à centralisateurs cycliques) engendré par (a_1, \dots, a_g) , soient H_1, \dots, H_k des sous-groupes non élémentaires et l un entier (grand). La première étape consiste à construire une surjection π de Γ sur un groupe hyperbolique $\bar{\Gamma}$ non élémentaire sans torsion (resp. à centralisateurs cycliques) tel que d'une part chacun des H_i se surjecte sur $\bar{\Gamma}$ et, d'autre part, la restriction de π à la boule de rayon l dans Γ est injective. De plus, on peut choisir π de telle sorte que $\bar{\Gamma}$ soit engendré par deux éléments. Supposons pour simplifier que $k = 1$. On cherche des éléments h_i ($i = 1, \dots, g + 2$) qui appartiennent à H_1 et qui sont tels que $a_1^{-1}h_1, \dots, a_g^{-1}h_g, h_{g+1}, h_{g+2}$ soient de très grande norme et forment une famille vérifiant une condition de petite simplification relative. Si l'on dispose d'une telle famille, le quotient $\bar{\Gamma} = \Gamma / \langle a_1^{-1}h_1, \dots, a_g^{-1}h_g \rangle$

est hyperbolique. Par construction, H_1 se surjecte sur $\bar{\Gamma}$ car l'image de H_1 dans $\bar{\Gamma}$ contient tous les $\pi(a_i)$. Par ailleurs les projections de h_{g+1} et h_{g+2} dans $\bar{\Gamma}$ engendrent un groupe libre de sorte que $\bar{\Gamma}$ est non élémentaire. C'est la condition sur la torsion (ou sur les centralisateurs cycliques) qui permet de construire assez facilement de tels éléments h_i (on choisit des h_i de longueurs comparables et on montre que lorsque ces longueurs tendent vers l'infini la proportion des choix qui ne vérifient pas la condition cherchée tend vers 0) ; et on peut par ailleurs assurer que le quotient $\bar{\Gamma}$ est lui-même sans torsion ou à centralisateurs cycliques suivant le cas. Enfin, en appliquant la construction précédente au produit libre $\Gamma \star F_2$ muni des sous-groupes H_i et F_2 , on peut faire en sorte que $\bar{\Gamma}$ soit engendré par deux éléments.

Considérons maintenant k groupes hyperboliques marqués à g générateurs $\Gamma_1, \dots, \Gamma_k$, sans torsion (resp. à centralisateurs cycliques) et appliquons le fait précédent à leur produit libre. On obtient ainsi une surjection de $\Gamma_1 \star \dots \star \Gamma_k$ sur un groupe $\bar{\Gamma}$ hyperbolique non élémentaire sans torsion (resp. à centralisateurs cycliques) engendré par deux éléments. Cette surjection est injective sur une grande boule et surjective en restriction à chaque facteur. En termes de l'espace des groupes de type fini $\mathcal{G}r$, cela signifie que si l'on choisit des voisinages U_i des Γ_i dans $\mathcal{G}r$, il existe un $\bar{\Gamma}$ dont la classe d'isomorphisme rencontre chacun des U_i . Utilisant encore le théorème de BAIRE et le fait qu'il n'existe qu'un nombre dénombrable de classes d'isomorphisme de groupes hyperboliques, on obtient le fait que la relation \simeq est topologiquement transitive dans $\overline{\mathcal{H}yp}^{cc}$ et $\overline{\mathcal{H}yp}^{st}$:

THÉORÈME.— *Dans $\overline{\mathcal{H}yp}^{cc}$ (resp. $\overline{\mathcal{H}yp}^{st}$), il existe un G_δ dense formé de groupes marqués dont la classe d'isomorphisme est dense dans $\overline{\mathcal{H}yp}^{cc}$ (resp. $\overline{\mathcal{H}yp}^{st}$). Ces groupes peuvent être choisis de rang 2, c'est-à-dire engendrés par deux éléments.*

Il existe des groupes hyperboliques sans torsion ayant la propriété (T) de KAZHDAN (voir par exemple [HV89] et la fin de cet exposé). Puisqu'un quotient d'un groupe ayant cette propriété a également cette propriété, on obtient un ouvert non vide dans $\mathcal{G}r$ formé de tels groupes. Comme cet ouvert rencontre nécessairement le G_δ dense construit précédemment, on en déduit l'existence d'un ouvert dense dans $\overline{\mathcal{H}yp}^{cc}$ (resp. $\overline{\mathcal{H}yp}^{st}$) formé de groupes ayant la propriété (T). Les autres propriétés génériques mentionnées dans le théorème C se montrent de manière analogue.

5. MODÈLE À GRAPHE

Petite simplification sur un graphe fini

La théorie classique de la petite simplification montre ses faiblesses dans des cas très simples. Considérons par exemple trois mots réduits m_1, m_2, m_3 , de même longueur, en les lettres $a_1^{\pm 1}, a_2^{\pm 1}$, et le groupe dont une présentation est $\langle a_1, a_2 \mid m_1 = m_2 = m_3 \rangle$. On peut évidemment écrire cette présentation sous la forme habituelle $\langle a_1, a_2 \mid m_1 m_2^{-1}, m_3 m_2^{-1} \rangle$ mais il est clair que cette nouvelle présentation n'est pas à petite simplification car m_2

est une pièce dont la longueur est la moitié de celle des relateurs (au moins si $m_1 m_2^{-1}$ et $m_3 m_2^{-1}$ sont des mots réduits). Dans [Gr01a-01b-01c-02], on trouve une théorie bien plus générale qui peut par exemple s'adapter à la situation précédente lorsque les mots m_i n'ont pas de grandes pièces entre eux. Soit (X, A) un graphe fini constitué d'un ensemble X de sommets et d'un ensemble $A \subset X \times X$ d'arêtes. On suppose qu'il n'y a pas de boucles et que le graphe est symétrique : si $a = (x, x')$ est une arête, il en est de même pour $\bar{a} = (x', x)$. On munira toujours les sommets d'un graphe de la distance combinatoire, longueur minimale des chemins joignant deux sommets. Étant donnée une famille finie de lettres (a_1, \dots, a_g) , on peut étiqueter le graphe en écrivant sur chaque arête l'une des lettres $a_i^{\pm 1}$, en faisant en sorte que les étiquettes portées par deux arêtes a, \bar{a} sont inverses l'une de l'autre. Sur chaque chemin du graphe, on peut donc lire un mot en les $a_i^{\pm 1}$. Un tel graphe étiqueté définit naturellement un groupe, quotient du groupe libre de base les a_i par le sous-groupe normal engendré par les mots lus sur les cycles du graphe. Lorsque la réalisation géométrique du graphe est un bouquet de cercles, on retrouve la notion classique de présentation de groupe. Il est bien clair que les groupes définis de cette manière ne sont autres que les groupes de présentation finie mais la « présentation » en termes de graphes est souvent plus adaptée.

Nous allons définir une notion de pièce pour un tel graphe étiqueté (X, A, e) . Supposons d'abord que l'étiquetage soit *réduit*, c'est-à-dire que deux arêtes consécutives (et non inverses) ne portent pas des étiquettes inverses. Une *pièce* est un graphe étiqueté qui se plonge de deux manières différentes dans (X, A, e) en respectant les étiquettes. Si $0 < \lambda < 1$, on dit que le graphe étiqueté vérifie la *condition de petite simplification* $C'(\lambda)$ si le diamètre de toute pièce est strictement inférieur à λ fois le *tour de taille* du graphe (la longueur de son plus petit cycle). Une esquisse de preuve géométrique du théorème suivant a été donnée par M. GROMOV et Y. OLLIVIER en a donné une preuve combinatoire détaillée dans l'esprit de la démonstration classique [Oll03c].

THÉORÈME.— *Si un graphe étiqueté réduit (X, A, e) vérifie la condition de petite simplification $C'(1/6)$, alors le groupe Γ qu'il définit est un groupe hyperbolique non élémentaire. La constante δ_S d'hyperbolicité peut être choisie égale au diamètre de (X, A) . L'application naturelle de (X, A) dans le graphe de CAYLEY de Γ est un plongement isométrique.*

Partons d'un graphe fini (X, A) et choisissons aléatoirement un étiquetage e . Il est bien sûr probable que cet étiquetage n'est pas réduit mais on peut le réduire, c'est-à-dire construire un morphisme surjectif de graphes étiquetés $\pi : (X, A, e) \rightarrow (\bar{X}, \bar{A}, \bar{e})$ dont le but est réduit. On peut construire cette réduction par identifications successives de deux arêtes consécutives qui portent des étiquettes inverses, ou directement en identifiant deux sommets de (X, A, e) s'ils sont reliés par un chemin portant un mot trivial dans le groupe libre.

Nous considérons maintenant une suite de graphes finis connexes (X_l, A_l) (pour être précis, il n'est pas nécessaire que l décrive tous les entiers mais une suite infinie nous

suffira). Nous allons dégager des propriétés géométriques de cette famille de graphes qui garantiront que très probablement, les graphes étiquetés vérifient la condition $C'(1/6)$ (après réduction) :

- (1) La valence de tous les sommets est au moins 2 (pas de sommet terminal) et bornée par 3 (par exemple).
- (2) Le tour de taille de (X_l, A_l) est supérieur à l et le diamètre de (X_l, A_l) est inférieur à $100l$.
- (3) Le nombre de chemins plongés dans (X_l, A_l) de longueur inférieure à $l/2$ est inférieur à $cst \cdot \beta^{l/2}$ pour une certaine constante $\beta > 1$ que nous déterminerons plus loin.

Supposons les conditions (1-2) satisfaites. Le nombre de sommets est alors inférieur à $3(2^{100l} - 1)$ de sorte que la condition (3) est vérifiée avec la constante $\beta_0 = 2^{100} + 1/2$. Si l'on subdivise toutes les arêtes de (X_l, A_l) en N arêtes, on obtient des graphes $(X_{l/N}, A_{l/N})$ dont le tour de taille est supérieur à Nl , le diamètre inférieur à $100Nl$ et dont le nombre de chemins de longueur inférieure à $Nl/2$ est inférieur à $cst_N(\beta_0)^{l/2} = cst_N(\beta_0^{1/N})^{Nl/2}$ et les conditions (1-2-3) sont donc satisfaites pour les graphes $(X_{l/N}, A_{l/N})$ avec la constante $\beta = (\beta_0)^{1/N}$ (pour l multiple de N). *Si les deux premières conditions sont vérifiées et si $\beta > 1$ est donné il est donc facile de construire une suite de graphes vérifiant également la troisième condition, quitte à subdiviser suffisamment.*

THÉORÈME.— *Si les conditions géométriques (1-2-3) sont satisfaites pour un choix de $\beta > 1$ suffisamment proche de 1, un étiquetage aléatoire de (X_l, A_l) mène très probablement à un graphe étiqueté réduit qui vérifie la condition $C'(1/6)$.*

Nous allons indiquer les étapes principales de la preuve. Pour un entier $n \geq 1$, considérons les $(2g)^n$ mots, réduits ou non, en les lettres $a_i^{\pm 1}$. Si p_n désigne la probabilité qu'un tel mot représente l'élément trivial du groupe libre, on sait que $\theta_0 = \lim_{n \rightarrow \infty} (p_{2n})^{\frac{1}{2n}} < 1$ (avec $\theta_0 = \sqrt{2g-1}/g$ mais nous préférons garder la notation θ_0 pour une référence ultérieure). On a donc une estimation de la forme $p_n \leq cst \cdot \theta^n$ si $\theta_0 < \theta < 1$. Soit $0 < \alpha < 1$. La probabilité pour qu'un mot de longueur n représente un élément du groupe libre de norme inférieure à αn est inférieure à $cst \cdot (2g)^{\alpha n \theta^{(1+\alpha)n}} \leq cst \cdot (2g)^{\alpha n \theta^n}$. Nous choisissons α en fonction de g et θ de telle sorte que $(2g)^{\alpha \theta} < 1$; pour fixer les idées, nous supposons $\alpha = -\frac{1}{2} \log_{2g} \theta$ de sorte que $(2g)^{\alpha \theta} = \theta^{1/2}$.

Considérons maintenant un étiquetage aléatoire (X_l, A_l, e_l) et majorons la probabilité pour qu'un mot lu sur au moins un chemin plongé de longueur n comprise entre ϵl et $l/2$ représente un élément du groupe libre de norme inférieure à αn , pour un $\epsilon > 0$ que nous déterminerons plus loin. D'après la propriété (3), cette probabilité est inférieure à $cst \cdot \beta^{l/2} \theta^{\epsilon l/2}$. On constate donc que pour $\beta \theta^\epsilon < 1$ on peut affirmer que très probablement quand l tend vers l'infini, tous les chemins de longueur n comprise entre ϵl et $l/2$ lus sur

le graphe étiqueté (X_l, A_l, e_l) représentent des éléments du groupe libre dont la norme est au moins αn .

Si l'on choisit un point base dans (X_l, A_l) , la donnée d'un étiquetage définit une application f_l du revêtement universel de (X_l, A_l) (en ce point base) vers le graphe de CAYLEY de F_g : on associe à chaque chemin issu du point base la valeur dans F_g du mot lu sur le chemin. Rappelons qu'une suite de points $(x_i)_{i \in \mathbb{Z}}$ dans un espace métrique δ -hyperbolique forme une (L_0, L_1, α^{-1}) *quasi-géodésique locale* si $|i - j| \leq L_1$ entraîne que la distance entre x_i et x_j est comprise entre $\alpha|i - j| - L_0$ et $\alpha^{-1}|i - j| + L_0$. On sait que si $L_1 \geq cst \cdot \alpha^{-1}(\delta + L_0)$, ceci entraîne que pour tous les entiers i, j , la distance entre x_i et x_j est supérieure à $\frac{1}{2}(\alpha|i - j| - L_0)$, c'est-à-dire que (x_i) est une *quasi-géodésique globale* [Gr87, §7] (cst est une constante universelle, de l'ordre de 10^5). Les estimations de distance pour $\epsilon l \leq n \leq l/2$ montrent que la restriction de f_l à une géodésique est une $(\alpha \epsilon l, l/2, \alpha^{-1})$ -quasi-géodésique locale. Par conséquent, si $\epsilon \leq 1/(3cst)$ et si β est choisi comme précédemment, ces applications f_l , pour l assez grand, sont en fait des plongements quasi-isométriques globaux (remarquons en passant que nous n'utilisons pas le fait que le groupe est libre, c'est-à-dire que $\delta = 0$). Si x, x' sont deux sommets du revêtement universel de (X_l, A_l) à distance $s \geq \epsilon l$, la distance entre leurs images est comprise entre $\alpha s/2$ et s . Ceci entraîne alors une quasi-convexité de l'image de f_l : tout segment géodésique dans le graphe de CAYLEY de F_g qui joint deux points de l'image de f_l reste à une distance bornée de cette image.

Nous allons montrer maintenant que le graphe réduit ne présente pas de pièce de longueur supérieure à $\alpha l/12$ (on remarquera que le tour de taille du graphe réduit est supérieur à αl). Nous pouvons supposer que l'étiquetage est tel que la propriété précédente de quasi-isométrie est satisfaite puisque nous savons que c'est le cas avec une probabilité proche de 1. Une pièce de longueur s donne lieu à deux plongements différents d'un intervalle de longueur s dans le graphe réduit. Pour chaque chemin \bar{c} de longueur s , plongé dans le graphe réduit, il existe un chemin c de longueur inférieure à $2s/\alpha$ dans le graphe (X_l, A_l) dont la projection contient \bar{c} . Si $s \leq \alpha l/12$, le chemin c est de longueur inférieure à $l/6$ et on peut le supposer plongé (remarquons que les boules dans le graphe (X_l, A_l) de rayon inférieur à $l/2$ sont des arbres). Une pièce de longueur supérieure à $\alpha l/12$ dans le graphe réduit permet donc de construire deux chemins c_1, c_2 , plongés dans (X_l, A_l, e_l) , de longueur comprise entre $\alpha l/12$ et $l/6$, et sur lesquels on lit le même élément du groupe libre. Inversement, partons de deux chemins c_1, c_2 plongés dans (X_l, A_l) de longueur comprise entre $\alpha l/12$ et $l/6$ et calculons la probabilité pour qu'un étiquetage aléatoire les munissent de mots égaux dans le groupe libre. Si ces chemins sont disjoints, les étiquettes qu'on place sur chacune des arêtes de c_1 et c_2 sont indépendantes et la question revient donc à la suivante. Étant donnés deux entiers n_1 et n_2 compris entre $\alpha l/12$ et $l/6$ et deux mots aléatoires de longueurs n_1, n_2 , quelle est la probabilité qu'ils représentent le même élément du groupe libre ? Cela revient évidemment à la trivialité d'un élément de

longueur $n_1 + n_2$ de sorte que cette probabilité est bornée par $cst \cdot \theta^{2\alpha l/12}$. Nous laissons au lecteur le soin d'étudier le cas un peu plus délicat où les deux chemins c_1 et c_2 se rencontrent et d'établir une borne analogue. Puisque le nombre de paires de chemins de longueur inférieure à $l/2$ est inférieur à $cst \cdot \beta^l$, la probabilité d'existence d'une pièce de longueur supérieure à $\alpha l/12$ dans le graphe réduit est ainsi majorée par $cst \cdot \beta^l \theta^{2\alpha l/12}$ (auquel il convient d'ajouter la petite probabilité pour que la condition de quasi-isométrie ne soit pas satisfaite). Puisque nous pouvons choisir β aussi proche que nécessaire de 1 on peut faire en sorte que ceci tende vers 0 quand l tend vers l'infini. Pour un tel choix de la constante β , nous avons bien établi que très probablement un étiquetage de (X_l, A_l) vérifie la propriété de petite simplification $C'(1/6)$, après réduction.

Très petite simplification relative

On voudrait itérer la construction précédente. De la même manière qu'il a fallu étendre la théorie de la petite simplification classique en une théorie relative adaptée aux quotients des groupes hyperboliques, nous allons décrire ici une version relative du modèle à graphe, adaptée aux groupes hyperboliques. Soit Γ un groupe hyperbolique non élémentaire engendré par une partie finie symétrique S de cardinal $2g$ et notons δ_S la constante d'hyperbolicité associée à cette famille génératrice. Notons θ_Γ le *rayon spectral de la marche aléatoire* sur (Γ, S) ; c'est la limite de $\lim_{n \rightarrow \infty} (p_{2n})^{\frac{1}{2n}}$ où p_n désigne la probabilité pour qu'un mot de longueur n en les générateurs soit trivial dans Γ . D'après [Ke59], on sait que $\theta_\Gamma < 1$. On a donc une inégalité de la forme $p_n \leq cst \cdot \theta^n$ si $\theta_\Gamma < \theta < 1$. On supposera toujours que Γ est à centralisateurs cycliques (par exemple sans torsion). Si (X, A) est un graphe fini comme précédemment, on considère toujours des étiquetages e des arêtes par des éléments de S (et on suppose encore que deux arêtes opposées portent des étiquettes inverses).

Soit (X_l, A_l) une suite de graphes vérifiant les mêmes conditions géométriques (1-2-3) pour une certaine constante $\beta > 1$. Bien que le groupe Γ ne soit pas nécessairement libre, il n'est pas difficile de généraliser les arguments précédents lorsque Γ était le groupe libre à g générateurs.

THÉORÈME.— *Pour $g \geq 2$, $\theta < 1$, $\epsilon > 0$, il existe $\alpha < 1$ et $\beta > 1$ tels que si les conditions géométriques (1-2-3) sont satisfaites, un étiquetage aléatoire de (X_l, A_l) mène très probablement à un graphe qui vérifie la condition de quasi-isométrie $(QI)_{\epsilon, \alpha}$: tout chemin de longueur s supérieure à ϵl dans le revêtement universel du graphe (X_l, A_l) porte un mot dont la norme dans Γ est supérieure à $\alpha s/2$.*

Supposons cette propriété (QI) vérifiée et cherchons à définir le concept de a -pièce. Il s'agit de deux mots m_1, m_2 lus sur deux chemins de (X_l, A_l) pour lesquels il existe deux mots u_1, u_2 de norme inférieure à a dans Γ tels que $m_1 u_1$ et $u_2 m_2$ sont égaux dans Γ . On demande par ailleurs qu'il n'existe pas de chemin connectant le début du premier chemin au début du second et qui porte un mot égal à u_2 dans Γ . Pour la constante a , on peut

choisir $10\delta_S$ (mais ce n'est pas très important). On cherche bien sûr à éviter les grandes pièces, c'est-à-dire celles pour lesquelles les normes dans Γ de m_1 et m_2 sont supérieures à $\lambda\alpha l$ pour un certain $0 < \lambda < 1$. Si ces grandes pièces n'existent pas, on dit que la *propriété de petite simplification* $C'(\lambda)$ relative est satisfaite. Plus géométriquement, on considère la partie quasi-convexe T du graphe de CAYLEY de Γ image par f_l du revêtement universel de (X_l, A_l) et tous ses translatés à gauche $\gamma.T$. Une pièce correspond à l'intersection de deux voisinages convenables de deux translatés de T qui ne coïncident pas.

La démonstration du théorème suivant est une « simple » généralisation de celle que nous avons présentée dans le cas du groupe libre : la non-liberté du groupe Γ est compensée par son hyperbolicité.

THÉORÈME.— *Pour $g \geq 2$, $\theta < 1$, $\epsilon > 0$, $0 < \lambda < 1$, il existe $\alpha < 1$ et $\beta > 1$ tels que si les conditions géométriques (1-2-3) sont satisfaites, un étiquetage aléatoire de (X_l, A_l) mène très probablement à un graphe qui vérifie les conditions $(QI)_{\epsilon, \alpha}$ et $C'(\lambda)$.*

Le fait important qu'il faut retenir est que ces propriétés (QI) et $C'(\lambda)$ sont (très probablement) satisfaites pour une valeur de β qui ne dépend que du nombre (fixé) de générateurs g , de la valeur de λ (qui sera également fixée) et du rayon spectral θ_Γ .

Dans le cas du groupe libre, la condition $C'(1/6)$ est suffisante pour garantir l'hyperbolicité. Dans le cas général, une valeur universelle suffit.

THÉORÈME.— *Il existe $\lambda_0 > 0$ (très petit) ayant les propriétés suivantes. Soit (X_l, A_l) une suite de graphes vérifiant la condition (2), munis d'étiquetages e_l vérifiant la condition de quasi-isométrie $(QI)_{\epsilon, \alpha}$ (pour un certain $\alpha < 1$ et $\epsilon > 0$ suffisamment petit) et la condition de petite simplification relative $C'(\lambda_0)$. Si l est assez grand, le quotient Γ_1 de Γ par le sous-groupe normal engendré par les éléments de Γ lus sur les cycles, est un groupe hyperbolique non élémentaire. La boule de rayon $\alpha l/5$ dans Γ se projette injectivement dans le quotient Γ_1 . Le graphe (X_l, A_l) se plonge de manière quasi-isométrique dans le graphe de CAYLEY de Γ_1 : deux points à distance $s \geq \epsilon l$ dans le graphe sont envoyés sur deux points à distance supérieure à $\alpha s/c$ pour une constante universelle c .*

Les articles [Gr01b, Gr02] contiennent des théorèmes plus puissants (utiles dans le contexte du problème de Burnside) dont la lecture peut être précédée par celle de notes non publiées de T. DELZANT [De03]. On trouve une preuve combinatoire du théorème précédent dans [Oll03b].

Procédé limite

Partons d'un groupe hyperbolique non élémentaire Γ engendré par une partie finie symétrique S de cardinal $2g$. Supposons que Γ possède la propriété (T), que nous discuterons un peu plus loin mais dont nous retiendrons pour l'instant qu'elle entraîne que tous ses quotients infinis ont des rayons spectraux uniformément majorés par un certain

$\theta < 1$. Pour fixer les idées, on peut choisir pour Γ un réseau cocompact dans le groupe des isométries d'un espace hyperbolique quaternionique (voir [HV89]).

Choisissons une famille de graphes (X_l, A_l) vérifiant les propriétés (1-2-3). Nous savons que si β est inférieur à une valeur dépendant de θ et de g , lorsque l tend vers l'infini, le groupe Γ_1 défini par un étiquetage aléatoire est très probablement hyperbolique non élémentaire. Soit donc Γ_1 l'un de ces quotients hyperboliques non élémentaires, défini par un certain étiquetage d'un certain graphe (X_{l_1}, A_{l_1}) . La projection de Γ sur Γ_1 est injective sur la boule de rayon $\alpha l_1/5$. Il existe une application naturelle i_1 du graphe (X_{l_1}, A_{l_1}) vers le graphe de CAYLEY de Γ_1 et nous savons que cette application est une quasi-isométrie dans le sens indiqué plus haut.

Puisque Γ_1 est un groupe hyperbolique non élémentaire dont le rayon spectral θ_{Γ_1} est majoré par *le même* θ , on conclut que lorsque l tend vers l'infini, le quotient de Γ_1 par le sous-groupe normal engendré par les mots lus sur les cycles d'un étiquetage aléatoire est très probablement hyperbolique non élémentaire. On peut considérer l'un de ces quotients Γ_2 . Puisque les rayons spectraux de tous les quotients infinis sont bornés par θ , on peut construire par récurrence une suite de groupes hyperboliques non élémentaires Γ_n , chacun étant un quotient du précédent, associée à une suite d'indices l_n de croissance suffisamment rapide. Par construction la projection de chaque Γ_n sur tous ses quotients successifs Γ_p (avec $p > n$) est une injection isométrique dans la boule de rayon $\alpha l_n/5$. De plus chaque graphe (X_{l_n}, A_{l_n}) est muni d'applications $i_{n,p}$ dans les graphes de CAYLEY des groupes Γ_p pour $p > n$ et ces applications sont quasi-isométriques : si deux points de (X_{l_n}, A_{l_n}) sont à distance s supérieure à ϵl_n , leurs images par $i_{n,p}$ sont à distance comprise entre $\alpha s/c$ et s .

La suite de groupes Γ_n converge clairement dans l'espace des groupes marqués vers un groupe Γ_∞ . Il s'agit d'un groupe de type fini qui n'est pas de présentation finie mais le lecteur pourra se convaincre qu'on peut le choisir de présentation récursive : il suffit pour cela de construire la suite l_n de manière récursive en considérant à chaque étape le plus petit indice pour lequel il existe un étiquetage vérifiant toutes les conditions requises et en choisissant par exemple le premier étiquetage qui fait l'affaire dans un ordre alphabétique. Résumons le résultat obtenu :

THÉORÈME.— *Soit Γ un groupe hyperbolique non élémentaire et sans torsion possédant la propriété (T). Soit (X_l, A_l) une suite de graphes vérifiant les conditions (1-2). Alors il existe des groupes $\bar{\Gamma}$ vérifiant les propriétés suivantes :*

- $\bar{\Gamma}$ est un quotient de Γ , de présentation récursive.
- Il existe une suite d'indices l_n et une suite d'applications i_n des graphes (X_{l_n}, A_{l_n}) dans le graphe de CAYLEY de $\bar{\Gamma}$ qui sont des « quasi-plongements » dans le sens où deux points à distance $s \geq \epsilon l_n$ dans (X_{l_n}, A_{l_n}) sont envoyés par i_n sur deux points à distance supérieure à $\alpha s/c$.

6. PROPRIÉTÉ (T)

Graphes expandeurs : le théorème D

Considérons encore un graphe fini connexe (X, A) et supposons pour simplifier qu'il soit v -régulier dans le sens où la valence de tous les sommets est égale à $v \geq 3$. Le laplacien Δ correspondant agit sur l'espace $l^2(X)$: on définit $\Delta f(x)$ comme la différence entre $f(x)$ et la valeur moyenne de f sur les sommets voisins de x . C'est un opérateur symétrique dont la première valeur propre non nulle se note $\lambda_1(X, A)$. Pour toute fonction f sur X à valeurs dans \mathbf{R} ou dans un espace de HILBERT \mathcal{H} , de somme nulle, le produit scalaire $\langle \Delta f | f \rangle$ est supérieur à $\lambda_1(X, A) \|f\|^2$. Il en résulte facilement que la moyenne arithmétique des $\|f(x) - f(x')\|^2$ sur tous les $|X|^2$ couples de sommets est inférieure à $\lambda_1^{-1}(X, A)$ fois la moyenne des $\|f(x) - f(x')\|^2$ sur les $|A|$ couples de sommets reliés dans le graphe. Une suite de graphes (X_l, A_l) v -réguliers est appelée un *expandeur* si le diamètre de (X_l, A_l) tend vers l'infini et si la suite $\lambda_1(X_l, A_l)$ est minorée par un nombre strictement positif λ . On pourra consulter [Lu94] pour une discussion (combinatoire et géométrique) de ces familles de graphes dont l'existence même est loin d'être évidente !

Soient (X_l, A_l) un expandeur et $j_l : X_l \rightarrow \mathcal{H}$ une famille d'applications dans un espace de HILBERT qui sont 1-lipschitziennes, *i.e.* envoyant deux sommets connectés dans le graphe sur deux points à distance inférieure à 1. Nous allons montrer qu'il existe deux suites x_l, x'_l de sommets de X_l dont les distances tendent vers l'infini dans les graphes (X_l, A_l) mais dont les images par j_l restent à distance bornée. Quitte à translater j_l , on peut supposer que chaque j_l est de somme nulle. La condition $\lambda_1(X_l, A_l) \geq \lambda > 0$ implique alors que la moyenne des $\|j_l(x) - j_l(x')\|^2$ sur tous les couples de sommets est inférieure à λ^{-1} . Cette moyenne est aussi égale à 2 fois la moyenne de $\|j_l(x)\|^2$. En particulier, le cardinal de l'ensemble des sommets x tels que $\|j_l(x)\|^2 \leq \lambda^{-1}$ est supérieur à $|X_l|/2$. D'autre part, nous savons que les boules de rayon k dans (X_l, A_l) ont au plus $1 + \dots + v^k \leq v^{k+1}$ éléments de sorte que si $v^{k+1} \leq |X_l|/10$, on peut affirmer qu'au moins 90% des couples de sommets sont à distance supérieure à k . Ces deux estimations montrent qu'on peut trouver deux sommets x_l, x'_l qui sont à distance supérieure à $\log_v(|X_l|/10) - 1$ et dont les images par j_l sont à distance inférieure à $2\lambda^{-1/2}$. Notre affirmation est donc démontrée.

On dit qu'une application f d'un espace métrique (E_1, d_1) vers un autre espace métrique (E_2, d_2) est un *plongement uniforme* si on peut trouver deux fonctions $\phi, \psi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ tendant vers l'infini à l'infini telles que pour tous les x, x' de E_1 :

$$\phi(d_1(x, x')) \leq d_2(f(x), f(x')) \leq \psi(d_1(x, x')).$$

Par exemple, il est clair que si Γ_1 et Γ_2 sont deux groupes de type fini munis de la métrique des mots, tout homomorphisme injectif de Γ_1 dans Γ_2 est un plongement uniforme. Nous venons de voir que si un espace métrique contient une famille de graphes expandeurs plongés de manière isométrique, cet espace métrique ne peut pas se plonger de manière uniforme dans un espace de HILBERT.

Il se trouve qu'il existe des familles de graphes expandeurs (X_l, A_l) qui vérifient de plus les conditions géométriques (1-2) relatives au diamètre et au tour de taille. De tels exemples peuvent être construits par des méthodes arithmétiques [Lu94] ou encore par des méthodes de graphes aléatoires [Bo01]. On peut donc utiliser le procédé limite décrit précédemment pour produire un groupe de type fini et de présentation récursive Γ dont le graphe de CAYLEY contient une copie quasi-isométrique des graphes (X_l, A_l) dans le sens décrit plus haut. Notons que l'argument précédent de non plongement uniforme s'étend immédiatement aux images *quasi-isométriques* d'expandeurs (car la proportion des paires de sommets à distance inférieure à el dans (X_l, A_l) tend vers 0 quand l tend vers l'infini). En particulier, ce groupe de type fini Γ , muni de la métrique des mots, ne peut pas se plonger de manière uniforme dans un espace de HILBERT. Comme Γ est de présentation récursive, le théorème de HIGMAN entraîne qu'il s'injecte dans un groupe de *présentation finie* qui n'admet pas non plus de plongement uniforme. *Ceci établit le théorème D.*

Constantes de Kazhdan et de Poincaré

Considérons maintenant un ensemble *dénombrable* X . Une *marche aléatoire* sur X est une application qui associe à chaque point x de X une probabilité « de transition » $\mu(x \rightarrow)$ sur X , donnant une masse $\mu(x \rightarrow y)$ au point y , nulle pour tous les y sauf pour un nombre fini d'entre eux. Si μ_1 et μ_2 sont deux telles marches on définit leur produit par $\mu_1 \star \mu_2(x \rightarrow y) = \sum_z \mu_1(x \rightarrow z)\mu_2(z \rightarrow y)$. La puissance n -ième d'une marche μ se note μ^n ; elle décrit bien sûr les transitions en n pas. On dit que la marche μ est *symétrique* pour une mesure ν sur X si $\nu(x)\mu(x \rightarrow y) = \nu(y)\mu(y \rightarrow x)$ pour tous les couples de points (x, y) . Le cas classique est celui où X est un groupe Γ engendré par une partie finie symétrique S et pour lequel on définit $\mu(x \rightarrow)$ comme la mesure de probabilité équirépartie sur les sommets voisins de x dans le graphe de CAYLEY. Cette marche est alors symétrique pour la mesure ν donnant une masse 1 à chaque sommet : c'est la *marche aléatoire simple* sur Γ associée à S . Nous nous plaçons dans un cas un peu plus général : nous supposons que le groupe de type fini Γ opère librement sur X , que X/Γ est fini, et que la marche μ et la mesure ν sont invariantes par l'action. Pour éviter des cas dégénérés, nous supposons également que le graphe dont les sommets sont les points de X et les arêtes les (x, y) avec $\nu(x)\mu(x \rightarrow y) \neq 0$ est connexe.

Nous fixons par ailleurs une représentation linéaire isométrique ρ de Γ dans un espace de HILBERT \mathcal{H} .

Soit \mathcal{E} l'espace des applications $f : X \rightarrow \mathcal{H}$ qui sont équivariantes sous les actions de Γ sur X et \mathcal{H} . Évidemment, la fonction $\nu(x)||f(x)||^2$ est Γ -invariante de sorte qu'on peut la sommer sur l'ensemble fini X/Γ et ceci définit une structure hilbertienne sur \mathcal{E} , qu'on note $\langle f | f \rangle_\nu = ||f||_\nu^2$. De même, en sommant sur X^2/Γ la fonction invariante $\nu(x)\mu(x \rightarrow y)||F(x, y)||^2$ on définit une norme sur l'espace des fonctions Γ -équivariantes de X^2 vers \mathcal{H} , notée $\langle F | F \rangle_\mu = ||F||_\mu^2$.

Soit $M : \mathcal{E} \rightarrow \mathcal{E}$ l'opérateur de moyennisation : $Mf(x)$ est la moyenne de f pour la mesure $\mu(x \rightarrow)$. C'est un opérateur symétrique. Le laplacien est l'opérateur $Id - M$. Si f est dans \mathcal{E} , son énergie $E_\mu(f)$ est le carré de la norme de la fonction $df(x, x') = f(x) - f(x')$ définie sur X^2 autrement dit $\|df\|_\mu^2$. Plus généralement, pour tout entier $n \geq 1$, on peut considérer l'énergie $E_{\mu^n}(f) = \|df\|_{\mu^n}^2$.

L'existence d'un séminaire BOURBAKI récent consacré à la propriété (T) de KAZHDAN permet de présenter cette propriété sans motivation [Va02] ! Nous nous contentons de rappeler qu'on dit qu'une représentation linéaire isométrique ρ possède *presque des vecteurs fixes* si on peut trouver une suite de vecteurs v_n de norme 1 telle que pour tout γ de Γ , la suite $\rho(\gamma)(v_n) - v_n$ converge vers 0. Un groupe a la propriété (T) si toute représentation linéaire isométrique possédant presque des vecteurs fixes possède des vecteurs fixes non nuls.

M. GROMOV donne plusieurs interprétations équivalentes de la propriété (T) en termes des concepts que nous venons de définir, qui lui permettent surtout de définir des renforcements de cette propriété lorsque l'on remplace ρ par une action isométrique *non linéaire* sur des espaces métriques à courbure négative ou nulle assez généraux. Faute d'espace, nous nous contentons ici d'un énoncé très faible :

THÉORÈME.— *Fixons un groupe Γ agissant sur un ensemble X muni d'une marche aléatoire μ comme ci-dessus. Alors les propriétés suivantes sont équivalentes pour tout $n \geq 1$ et pour tout $k \geq 2$:*

- Γ possède la propriété (T).
- Il existe une constante $\kappa_n < 1$ telle que pour toute représentation linéaire isométrique ρ et toute application équivariante f , on a $E_\mu(M^n f) \leq \kappa_n E_\mu(f)$ (inégalité de KAZHDAN).
- Il existe une constante $\pi_k < k$ telle que pour toute représentation linéaire isométrique ρ et toute application équivariante f , on a $E_{\mu^k}(f) \leq \pi_k E_\mu(f)$ (inégalité de POINCARÉ).

Pour le montrer, on remarque d'abord que les points fixes de M correspondent exactement aux fonctions f qui sont constantes (donc à valeurs dans les points fixes de ρ). Il en résulte facilement que le groupe Γ a la propriété (T) si et seulement s'il existe $\sigma < 1$ tel que pour toute représentation ρ sans vecteur fixe non nul, le spectre de l'opérateur M correspondant est contenu dans $[0, \sigma]$. Par projection sur l'orthogonal de l'espace des vecteurs fixes de ρ , on se ramène au cas où ρ n'a pas de vecteur fixe non nul. On exprime ensuite les quantités qui sont en jeu en fonction de M . On trouve :

$$E_\mu(f) = \langle f | (I - M)f \rangle_\nu, \quad E_\mu(M^n f) = \langle f | (I - M)M^{2n} f \rangle_\nu, \quad E_{\mu^k}(f) = \langle f | (I - M^k)f \rangle_\nu.$$

Les trois propriétés énoncées dans le théorème sont maintenant clairement équivalentes, car $\sigma < 1$, $\sigma^{2n} < 1$, et $(1 - \sigma^k)/(1 - \sigma) < k$ sont équivalents.

Considérons le cas de la marche aléatoire simple sur un groupe Γ engendré par une partie finie symétrique S . Si $\bar{\Gamma}$ est un quotient de Γ , on dispose d'une représentation régulière naturelle de Γ sur l'espace $l^2(\bar{\Gamma})$, sans vecteur invariant non nul si $\bar{\Gamma}$ est infini.

Le rayon spectral de l'opérateur M correspondant est le rayon spectral de la marche aléatoire simple sur $\bar{\Gamma}$ et on retrouve bien le fait que nous avons déjà mentionné : les quotients infinis d'un groupe ayant la propriété (T) ont des marches aléatoires simples dont les rayons spectraux sont uniformément majorés par un réel $\theta < 1$.

Critères locaux pour la propriété (T)

Supposons maintenant que X soit l'ensemble des sommets d'un complexe simplicial connexe P de dimension 2 sur lequel Γ agit de manière libre avec un quotient fini P/Γ . Pour chaque x de X , on note $\tau(x)$ le nombre de triangles qui contiennent x et pour chaque couple (x, x') de sommets distincts, on note $\tau(x, x')$ le nombre de triangles qui les contiennent tous les deux. On suppose $\tau(x) \geq 1$ pour tout x , et on définit une marche aléatoire par $\mu_P(x \rightarrow y) = \tau(x, y)/2\tau(x)$ symétrique pour la mesure $\nu_P(x) = 2\tau(x)$.

Considérons le *link* d'un sommet x . Il s'agit du graphe fini dont l'ensemble des sommets X_x est l'ensemble des points de X à distance 1 de x dans le 1-squelette de P et dont l'ensemble des arêtes $A_x \subset X_x \times X_x$ contient les (y, y') tels que x, y, y' sont les trois sommets d'un triangle de P . On munit A_x de la mesure de probabilité uniforme μ_x donnant donc une masse $1/\tau(x)$ à chaque arête. On munit X_x de la mesure de probabilité $\nu_x = \mu_P(x \rightarrow \cdot)$ donnant donc à chaque sommet du link une masse proportionnelle à sa valence dans le link.

Supposons le link connexe et notons $\lambda_1(x) > 0$ la première valeur propre non nulle du laplacien de ce graphe fini. On a donc l'inégalité suivante pour toute fonction f de X_x vers un espace de HILBERT \mathcal{H} :

$$\sum_{(y, y') \in X_x \times X_x} \|f(y) - f(y')\|^2 \nu_x(y) \nu_x(y') \leq \lambda_1(x)^{-1} \sum_{(y, y') \in A_x} \|f(y) - f(y')\|^2 \mu_x(y, y').$$

Supposons maintenant que l'on dispose d'une minoration $\lambda_1(x) > \lambda > 0$ et soit $f : X \rightarrow \mathcal{H}$ une application Γ équivariante comme précédemment. On peut alors sommer les inégalités précédentes sur tous les sommets x modulo Γ , en les affectant du poids $\tau(x)$. Il est clair que l'inégalité obtenue peut aussi s'écrire :

$$\sum_{(y, y') \in X \times X/\Gamma} \|f(y) - f(y')\|^2 \nu_P(y) \mu_P^2(y, y') \leq \lambda^{-1} \sum_{(y, y') \in X \times X/\Gamma} \|f(y) - f(y')\|^2 \nu_P(y) \mu_P(y, y').$$

Autrement dit, nous avons une inégalité de POINCARÉ $E_{\mu_P^2}(f) \leq \pi_2 E_{\mu_P}(f)$ avec $\pi_2 = \lambda^{-1}$ de sorte que si $\lambda > 1/2$, on peut conclure que Γ a la propriété (T). Nous avons donc montré :

THÉORÈME. — *Soit P un complexe simplicial connexe de dimension 2 dont tous les links des sommets sont connexes et ont une première valeur propre non nulle strictement supérieure à $1/2$. Si un groupe Γ agit librement sur P avec un quotient fini, alors Γ a la propriété (T).*

Ce théorème a une histoire intéressante : il prend sa source dans des théorèmes de H. GARLAND d'annulation de certaines cohomologies de réseaux arithmétiques [Ga73]. La

preuve fut ensuite simplifiée par A. BOREL [Bo75]. Puis W. BALLMANN-J. SWIATKOWSKI, P. PANSU et A. ŻUK en déduisirent indépendamment une preuve de la propriété (T) pour certains groupes agissant sur certains immeubles de TITS euclidiens [BS97, Pa98, Żu96]. L'énoncé général précédent est dû à W. BALLMANN-J. SWIATKOWSKI et A. ŻUK, et la preuve particulièrement élégante que nous venons de présenter est due à M. GROMOV [Gr02].

On peut appliquer ce résultat au 2-complexe de CAYLEY associé à une présentation de groupes si tous les relateurs sont des mots de longueur 3. A. ŻUK obtient ainsi des critères simples et effectifs qui permettent de garantir la propriété (T) à partir de propriétés combinatoires d'une présentation. Toute présentation peut d'ailleurs être triangulée et ce critère est donc de portée générale [Żu03]. Nous renvoyons à l'exposé de A. VALETTE [Va02].

La connaissance assez précise des valeurs propres d'un graphe fini aléatoire (voir [Bo01]) permet à A. ŻUK d'appliquer ce critère pour une présentation triangulaire aléatoire. Considérons une paire de permutations σ_1, σ_2 de l'ensemble fini $\{a_1^{\pm 1}, \dots, a_l^{\pm 1}\}$ à $2l$ éléments et associons-leur les $2l$ relateurs triangulaires $a_i^{\pm 1} \sigma_1(a_i^{\pm 1}) \sigma_2(a_i^{\pm 1})$. Un ensemble R de couples de permutations permet donc de définir une présentation de groupe à l générateurs et $2l|R|$ relateurs. Fixons le cardinal de R (assez grand) et faisons tendre le nombre l de générateurs vers l'infini. A. ŻUK montre que les groupes ainsi définis ont très probablement la propriété (T) lorsque les R permutations sont choisies aléatoirement [Żu03].

Propriété (T) dans le modèle à graphe

Dans la construction limite, nous sommes partis d'un groupe hyperbolique possédant la propriété (T) de façon à s'assurer que tous ses quotients infinis ont également cette propriété et que leurs rayons spectraux ne s'approchent pas de 1. M. GROMOV montre dans [Gr02] que ce n'était pas nécessaire :

THÉORÈME.— *Soit (X_l, A_l) une suite de graphes finis vérifiant les propriétés (1-2-3) et qui est par ailleurs un expanseur. Alors, pour β suffisamment proche de 1, lorsque l tend vers l'infini, le groupe défini par un étiquetage aléatoire possède très probablement la propriété (T).*

Même si ce théorème n'est pas nécessaire pour la construction de groupes ne se plongeant pas uniformément dans un espace de HILBERT, il est cependant remarquable car il montre une fois de plus le caractère générique de la propriété (T) (dans de nombreux modèles de graphes finis aléatoires, les propriétés (1-2) sont génériques, de même que la propriété d'expanseur). Nous n'avons malheureusement pas la place pour indiquer les étapes principales de la preuve et nous allons nous contenter d'indications très générales. L. SILBERMAN a rédigé une preuve complète [Si03] en accompagnement de l'article de M. GROMOV.

L'idée consiste à appliquer l'argument de géométrie intégrale que nous venons d'expliquer. Nous savons que, très probablement, le graphe de CAYLEY $X_{\Gamma, S}$ du groupe Γ défini par un étiquetage contient une copie quasi-isométrique Y_l du graphe (X_l, A_l) . Cette copie peut être translatée par chaque élément de Γ de sorte que $X_{\Gamma, S}$ est « pavé » par ces diverses

copies. Si $f : X_{\Gamma,S} \rightarrow \mathcal{H}$ est une application équivariante, on peut la restreindre à chacun de ces pavés $\gamma.Y_l$ et utiliser le fait que la première valeur propre de (X_l, A_l) est minorée. Ceci donne une minoration de type POINCARÉ pour la moyenne des $\|f(x) - f(x')\|^2$ lorsque (x, x') décrit tous les sommets de $\gamma.Y_l$ à distance n en termes de la moyenne de cette même quantité lorsque (x, x') ne décrit que les arêtes, c'est-à-dire les sommets à distance 1. En sommant cette inégalité sur tous les pavés, on obtient une inégalité de type POINCARÉ entre deux énergies de f de la forme $E_{\mu_n}(f) < \pi_n E_{\mu}(f)$ pour certaines marches aléatoires μ, μ_n . Mais il n'est pas vrai que μ_n coïncide avec μ^n et l'inégalité ainsi obtenue ne garantit pas immédiatement la propriété (T). Puisque Γ est un quotient du groupe libre, on peut relever les fonctions f et ces inégalités sur le groupe libre à g générateurs. De même, les mesures μ et μ_n se relèvent au niveau du groupe libre. L'avantage est maintenant que ce sont deux mesures sur un espace fixe, indépendant de l'étiquetage, et on peut donc en considérer les moyennes $\tilde{\mu}$ et $\tilde{\mu}_n$ sur tous les étiquetages. L'étude de $\tilde{\mu}$ et $\tilde{\mu}_n$ est facile car ce sont des mesures radiales dans le groupe libre et M. GROMOV montre que le phénomène bien connu de concentration permet de passer d'estimations concernant la moyenne de mesures, à des estimations presque aussi bonnes pour la majorité des étiquetages. Ceci donne finalement une bonne inégalité de type POINCARÉ pour la majorité des étiquetages et entraîne le théorème.

REMARQUES FINALES

Beaucoup d'idées contenues dans [Gr02] n'ont pas été évoquées dans cet exposé. L'une des principales est probablement l'introduction de versions renforcées de la propriété (T) qui permettent de garantir l'existence de points fixes pour des actions isométriques sur des espaces métriques « réguliers » à courbure négative ou nulle. M. GROMOV montre que les groupes définis par les graphes (X_l, A_l) possèdent très probablement cette propriété, et il en résulte par exemple qu'ils ne peuvent pas se plonger dans un groupe linéaire.

Comme le lecteur l'aura constaté, les groupes aléatoires que nous avons considérés sont « lacunaires » dans le sens où les relations imposées entre les générateurs, bien que nombreuses, viennent en « paquets » de longueurs très différentes. Est-il possible de comprendre la structure d'un groupe aléatoire dont la répartition des relations est plus homogène ?

La théorie géométrique des groupes s'intéresse (à juste titre) à des groupes bien différents des groupes aléatoires, possédant souvent une géométrie plus riche. Il faudra peut-être chercher d'autres modèles probabilistes, se concentrant autour de la densité $1/2$, pour rendre compte d'une abondance de groupes ayant ces riches propriétés géométriques, comme par exemple une dimension cohomologique supérieure ou égale à 3. Faudra-t-il se résigner au fait que les groupes aléatoires ne sont après tout que des « quite simple two-dimensional creatures » [Gr87] ?



BIBLIOGRAPHIE

- [Al91] ALONSO J. M. ET AL.: Notes on word hyperbolic groups. *In Group theory from a geometrical viewpoint (Trieste, 1990)*, pages 3–63. World Sci. Publishing, River Edge, NJ, 1991.
- [Ar97] ARZHANTSEVA G.: Sur les groupes dont les sous-groupes ayant un nombre fixé de générateurs sont libres (en russe). *Fundam. Prikl. Mat.*, 3(3):675–683, 1997.
- [Ar98] ARZHANTSEVA G.: Generic properties of finitely presented groups and Howson’s theorem. *Comm. Algebra*, 26(11):3783–3792, 1998.
- [AO96] ARZHANTSEVA G. & OL’SHANSKIĬ A. YU.: Généricité de la classe des groupes dont les sous-groupes ayant moins de générateurs sont libres (en russe). *Mat. Zametki*, 59(4):489–496, 1996.
- [BS97] BALLMANN W. & SWIATKOWSKI J.: On L^2 -cohomology and property (T) for automorphism groups of polyhedral cell complexes. *Geom. Funct. Anal.*, 7(4):615–645, 1997.
- [Bo01] BOLLOBÁS B.: *Random graphs*, volume 73 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, deuxième édition, 2001.
- [Bo75] BOREL A.: Cohomologie de certains groupes discrets et laplacien p -adique (d’après H. Garland). *In Séminaire Bourbaki, 26e année (1973/1974), Exp. No. 437*, pages 12–35. Lecture Notes in Math., Vol. 431. Springer, Berlin, 1975.
- [BH99] BRIDSON M. R. & HAEFLIGER A.: *Metric spaces of non-positive curvature*, volume 319 de *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1999.
- [Ch91] CHAMPETIER C.: Propriétés génériques des groupes de présentation finie. Thèse de doctorat, Université de Lyon I, décembre 1991.
- [Ch93] CHAMPETIER C.: Croissance des groupes à petite simplification. *Bull. London Math. Soc.*, 25(5):438–444, 1993.
- [Ch94] CHAMPETIER C.: Petite simplification dans les groupes hyperboliques. *Ann. Fac. Sci. Toulouse Math. (6)*, 3(2):161–221, 1994.
- [Ch95] CHAMPETIER C.: Propriétés statistiques des groupes de présentation finie. *Adv. Math.*, 116(2):197–262, 1995.
- [Ch00] CHAMPETIER C.: L’espace des groupes de type fini. *Topology*, 39(4):657–680, 2000.
- [CM82] CHANDLER B. & MAGNUS W.: *The history of combinatorial group theory*, volume 9 de *Studies in the History of Mathematics and Physical Sciences*. Springer-Verlag, New York, 1982. A case study in the history of ideas.
- [CDP90] COORNAERT M., DELZANT T. & PAPADOPOULOS A.: *Géométrie et théorie des groupes*, volume 1441 de *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1990. Les groupes hyperboliques de Gromov.
- [De96] DELZANT T.: Sous-groupes distingués et quotients des groupes hyperboliques. *Duke Math. J.*, 83(3):661–682, 1996.
- [De03] DELZANT T.: Mesoscopic curvature and very small cancellation theory (after M. Gromov). Manuscrit, 2003.
- [Ga73] GARLAND H.: p -adic curvature and the cohomology of discrete subgroups of p -adic groups. *Ann. of Math. (2)*, 97:375–423, 1973.
- [Gh90] GHYS É.: Les groupes hyperboliques. *In Séminaire Bourbaki, 42e année (1989/1990), Exp. No. 722, Astérisque*, (189-190): 203–238, 1990.
- [GH90] GHYS É. & DE LA HARPE P.: Sur les groupes hyperboliques, d’après M. Gromov. *Progress in Mathematics*, vol 83, Birkhäuser, Boston 1990.

- [Gri85] GRIGORCHUK R.: Degrees of growth of finitely generated groups and the theory of invariant means. *Mathematics of the USSR Izvestiya*, 25(2):259–300, 1985.
- [Gr81] GROMOV M.: Hyperbolic manifolds, groups and actions. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978)*, volume 97 de *Ann. of Math. Stud.*, pages 183–213, Princeton, N.J., 1981. Princeton Univ. Press.
- [Gr84] GROMOV M.: Infinite groups as geometric objects. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Varsovie, 1983)*, pages 385–392, Varsovie, 1984. PWN.
- [Gr87] GROMOV M.: Hyperbolic groups. In *Essays in group theory*, volume 8 de *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, New York, 1987.
- [Gr93] GROMOV M.: Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, volume 182 de *London Math. Soc. Lecture Note Ser.*, pages 1–295. Cambridge Univ. Press, Cambridge, 1993.
- [Gr99] GROMOV M.: *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 de *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999. Version développée issue de la version française de 1981, avec des appendices de M. Katz, P. Pansu et S. Semmes.
- [Gr00] GROMOV M.: Spaces and questions. *Geom. Funct. Anal.* (Special Volume, Part I):118–161, 2000. GAFA 2000 (Tel Aviv, 1999).
- [Gr01a] GROMOV M.: CAT(κ)-spaces: construction and concentration. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 280(Geom. i Topol. 7):100–140, 299–300, 2001.
- [Gr01b] GROMOV M.: Mesoscopic curvature and hyperbolicity. In *Global differential geometry: the mathematical legacy of Alfred Gray (Bilbao, 2000)*, volume 288 de *Contemp. Math.*, pages 58–69. Amer. Math. Soc., Providence, RI, 2001.
- [Gr01c] GROMOV M.: Small cancellation, unfolded hyperbolicity, and transversal measures. In *Essays on geometry and related topics, Vol. 1, 2*, volume 38 de *Monogr. Enseign. Math.*, pages 371–399. Enseignement Math., Genève, 2001.
- [Gr02] GROMOV M.: Random walk in random groups. Prépublication IHÉS M/02/03 Janvier 2002. À paraître dans *Geom. Funct. Anal.*
- [Ha00] DE LA HARPE P.: *Topics in geometric group theory*. Chicago Lectures in Mathematics Series, 2000.
- [HV89] DE LA HARPE P. & VALETTE A.: La propriété (T) de Kazhdan pour les groupes localement compacts (avec un appendice de Marc Burger). *Astérisque*, (175):158, 1989.
- [HR00] HIGSON N. & ROE J.: Amenable group actions and the Novikov conjecture. *J. Reine Angew. Math.*, 519:143–153, 2000.
- [HLS02] HIGSON N., LAFFORGUE V. & SKANDALIS G.: Counterexamples to the Baum-Connes conjecture. *Geom. Funct. Anal.*, 12:330–354, 2002.
- [IO96] IVANOV S. & OL'SHANSKIĬ A. YU.: Hyperbolic groups and their quotients of bounded exponents. *Trans. Amer. Math. Soc.*, 348(6):2091–2138, 1996.
- [KS02] KAPOVICH I. & SCHUPP P.: Genericity, the Arzhantseva-Ol'shanskii method and the isomorphism problem for one-relator groups. Prépublication, Octobre 2002. ArXiv:math.GR/0210307
- [Ke59] KESTEN H.: Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.
- [Lu94] LUBOTZKY A.: *Discrete groups, expanding graphs and invariant measures*, volume 125 de *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1994. Avec un appendice de Jonathan D. Rogawski.
- [LS01] LYNDON R.C. & SCHUPP P.E.: *Combinatorial group theory*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Nouveau tirage de l'édition de 1977.

- [Ne37] NEUMANN B.H.: Some remarks on infinite groups. *J. London Mat. Soc.*, 12:120-127, 1937.
- [Oll03a] OLLIVIER Y.: Sharp phase transition theorems for hyperbolicity of random groups. Prépublication, janvier 2003. ArXiv:math.GR/0301187
- [Oll03b] OLLIVIER Y.: On a small cancellation theorem of Gromov. Manuscrit, janvier 2003.
- [Oll03c] OLLIVIER Y.: Critical densities for random quotients of hyperbolic groups. À paraître dans *C. R. Acad. Sci. Paris Sér. I Math.*.
- [Ols84] OL'SHANSKIĬ A. YU.: On the geometric method in the combinatorial group theory. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Varsovie, 1983)*, pages 415–423, Varsovie, 1984. PWN.
- [Ols92a] OL'SHANSKIĬ A. YU.: Almost every group is hyperbolic. *Internat. J. Algebra Comput.*, 2(1):1–17, 1992.
- [Ols92b] OL'SHANSKIĬ A. YU.: Periodic factor groups of hyperbolic groups. *Mathematics of the USSR Sbornik*, (72):519–541, 1992.
- [Pa98] PANSU P.: Formules de Matsushima, de Garland et propriété (T) pour des groupes agissant sur des espaces symétriques ou des immeubles. *Bull. Soc. Math. France*, 126(1):107–139, 1998.
- [Pa96] PAPASOGLU P.: An algorithm detecting hyperbolicity. In *Geometric and computational perspectives on infinite groups (Minneapolis, MN and New Brunswick, NJ, 1994)*, volume 25 de *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 193–200. Amer. Math. Soc., Providence, RI, 1996.
- [Se92] SELA Z.: Uniform embeddings of hyperbolic groups in Hilbert spaces. *Israel J. Math.*, 80(1-2):171–181, 1992.
- [Si03] SILBERMAN L.: Random walk in random groups (after Gromov). À paraître dans *Geom. Funct. Anal.*
- [Sk00] SKANDALIS G.: Progrès récents sur la conjecture de Baum-Connes. Contribution de Vincent Lafforgue. In *Séminaire Bourbaki, 52e année (1999/2000), Exp. No. 829, Astérisque*, (276): 105–135, 2002.
- [SZ94] STUCK G. & ZIMMER R.J.: Stabilizers for ergodic actions of higher rank semisimple groups. *Annals Math.*, vol.139 (1994), pp. 723-747.
- [TV99] THOMAS S. & VELICKOVIC B.: On the complexity of the isomorphism relation for finitely generated groups. *J. Algebra*, 217(1):352–373, 1999.
- [Va02] VALETTE A.: Nouvelles approches de la propriété (T) de Kazhdan. In *Séminaire Bourbaki, 55e année (2002/2003), Exp. No. 913*.
- [Yu00] YU G.: The coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert space. *Invent. Math.*, 139(1):201–240, 2000.
- [Žu96] ŽUK A.: La propriété (T) de Kazhdan pour les groupes agissant sur les polyèdres. *C. R. Acad. Sci. Paris Sér. I Math.*, 323(5):453–458, 1996.
- [Žu03] ŽUK A.: Property (T) and Kazhdan constants for discrete groups. À paraître dans *Geom. Funct. Anal.*

Étienne GHYS

École Normale Supérieure de Lyon

UMPA – UMR CNRS 5669

46 allée d'Italie – F-69364 LYON Cedex 7

E-mail : Etienne.GHYS@umpa.ens-lyon.fr

LA PRIMALITÉ EN TEMPS POLYNOMIAL
[d'après Adleman, Huang ; Agrawal, Kayal, Saxena]

par François MORAIN

1. INTRODUCTION

Le théorème fondamental de l'arithmétique affirme que tout entier positif s'écrit comme un produit de puissances de nombres premiers distincts de manière unique, à l'ordre près des facteurs. On sait depuis Euclide qu'il existe une infinité de nombres premiers, même si la démonstration de ce résultat ne fournit pas de grands nombres premiers de manière constructive. Pendant des millénaires, trouver les facteurs premiers d'un entier esthétique a été une motivation pour le développement de méthodes de décomposition en facteurs premiers. Les nombres $F_n = 2^{2^n} + 1$ (nombres de Fermat) ou $M_m = 2^m - 1$ (nombres de Mersenne) ont ainsi servi de pierres de touche. Nul doute que la conjecture de Fermat sur la primalité de tous les F_n a stimulé bien des recherches, à l'époque d'Euler ou bien encore aujourd'hui. On renvoie le lecteur intéressé à [74] pour un aperçu de la période précédant l'arrivée des ordinateurs. Les calculs étant faits à la main (à l'exception de plusieurs tentatives de mécanisation), les spécialistes de l'époque savaient bien ce qui était *faisable* ou *infaisable*. L'exploit réalisé par Lucas en prouvant la primalité de M_{127} (39 chiffres décimaux) est resté inégalé jusqu'à l'arrivée des ordinateurs.

Ceux-ci ont apporté avec eux une puissance de calcul prodigieuse, et on peut imaginer sans trop de difficulté la joie des pionniers de la théorie algorithmique des nombres qui ont trouvé les premiers nombres premiers de Mersenne depuis Lucas ! Au fil des ans, la théorie de la complexité est née et a grandi, son but étant de classer les problèmes faciles et les problèmes difficiles. Pour simplifier, un problème est facile quand on lui connaît un algorithme de résolution dont le temps de calcul est polynomial en la taille de l'entrée. Une différenciation importante s'est faite entre algorithmes déterministes et algorithmes randomisés, ces derniers pouvant tirer à pile ou face pour prendre des décisions. La théorie de la complexité s'est ramifiée au cours des ans, et il serait présomptueux de la résumer en quelques lignes. Nous avons toutefois rassemblé dans la section 2 de cet article les

notions essentielles pour notre propos, ce qui nous permettra de signaler au fil de l'article les interprétations qu'on peut donner des théorèmes dans le langage de la complexité.

S'il fallait encore une motivation pour étudier les algorithmes de primalité, nul doute que nous ferions appel aux plus gros consommateurs de nombres premiers, à savoir les cryptographes, qui utilisent désormais quotidiennement des nombres premiers dans leurs cryptosystèmes [53], à commencer par le fameux RSA [4].

L'un des buts de cet article est de présenter l'algorithme récent d'Agrawal, Kayal et Saxena, qui montre que la primalité (ou plutôt le problème de décision `estPremier?`) est décidable en temps polynomial déterministe. Avant d'arriver à cela, il ne nous a pas paru inutile de revenir sur les vingt dernières années (grosso modo depuis l'article fondateur de la primalité moderne [47]), marquées entre autres par l'utilisation des courbes algébriques en primalité, fournissant le premier algorithme randomisé de primalité fonctionnant en temps polynomial. Ces considérations expliquent le titre choisi pour l'article.

Nous n'oublierons pas non plus les aspects pratiques des tests de primalité. Les progrès là encore ont été foudroyants dans les vingt dernières années, en liaison avec les avancées théoriques.

Les ouvrages traitant de nombres premiers et de factorisation sont nombreux. Outre l'incontournable [65], nous ferons référence principalement à [24] et [29].

Notations : dans ce qui suit, p et q dénoteront des nombres premiers, et N l'entier dont la primalité doit être étudiée.

2. BRÈVE INTRODUCTION À LA THÉORIE DE LA COMPLEXITÉ

Même si nous avons délibérément choisi de ne pas regarder la primalité à travers le prisme de la théorie de la complexité, il convient de donner quelques pistes pour comprendre les différentes règles du jeu (d'informatique théorique) qui interviennent. Le livre [58] est une bonne référence pour ce qui suit. On trouvera dans [32] une section sur la primalité et la complexité.

La façon la plus simple de mesurer la complexité d'un algorithme est par son temps de calcul. Le paramètre d'entrée est généralement la taille des données fournies à l'algorithme. On cherche alors une fonction de cette taille, notée n , qui donne le résultat. À titre d'exemple, l'addition de deux polynômes de degré $n - 1$ à coefficients dans $K = \mathbf{F}_2$ est n additions dans K , soit $f_K(n) = n$.

Deux algorithmes différents réalisant la même opération seront comparés à l'aide de leur fonction de coût respective. Une fois cet ordre de grandeur des calculs établis, la détermination des constantes sera primordiale dans la mesure de l'efficacité pratique des algorithmes.

Par exemple, l'algorithme classique de multiplication de deux polynômes de degré n prendra $O(n^2)$ multiplications d'éléments de K . Si le corps K le permet, on peut utiliser

la transformée de Fourier rapide (FFT), et on obtient alors un temps de calcul $O(n \log n)$ (voir [41, 32, 29]).

Le but est maintenant, pour chaque problème de calcul donné, de déterminer quelle est la meilleure fonction de coût possible pour la résolution du problème. Cela permet de répartir les problèmes dans différentes *classes de complexité*. Par exemple, la classe $\mathcal{T}(n^2)$ désigne la classe des problèmes pour lesquels il existe un algorithme de résolution en temps au plus $O(n^2)$ pour une entrée de taille n . Les problèmes de la classe $\mathbf{P} = \cup_{k \geq 1} \mathcal{T}(n^k)$ sont ceux dont le temps de calcul est au plus polynomial en n : c'est la classe des problèmes faciles par excellence. Ceux de $\mathbf{EXP} = \cup_{k \geq 1} \mathcal{T}(2^{n^k})$ nécessitent un temps de calcul exponentiel en n . Une autre classe permet de simplifier certains énoncés. Si $f(n) \rightarrow \infty$, on note $\tilde{O}(f(n)) = \cup_{k \geq 1} O(f(n) \log(f(n))^k)$. Par exemple, les algorithmes de multiplication à base de FFT décrits ci-dessus montrent que la multiplication est dans $\tilde{O}(n)$.

Pour le moment, nous avons sous-entendu que les algorithmes que nous utilisons étaient *déterministes*, comme dans le cas de l'addition ou de multiplication de polynômes. Cette contrainte est très forte. Considérons un problème de base de la théorie des nombres, qui est celui de la recherche d'un non résidu quadratique dans $(\mathbf{Z}/p\mathbf{Z})^*$ pour p premier. Sans l'hypothèse de Riemann, on ne sait pas trouver un tel résidu en temps polynomial déterministe, alors qu'on sait qu'un élément sur 2 n'est pas un carré et qu'il suffit donc de tirer au sort quelques valeurs pour en trouver une bonne. Un algorithme tirant des nombres au hasard a ainsi une probabilité de succès plus grande que $1/2$.

Les problèmes considérés ci-dessus sont des problèmes de *calcul*. Un autre type de problème très important en complexité est celui des problèmes de *décision*. Celui qui nous intéressera dans la suite est bien sûr **estPremier?**. Quand on ne dispose pas d'algorithmes déterministes pour répondre à la question, on peut chercher des algorithmes randomisés, qui peuvent s'aider de tirages à pile ou face. La classe des algorithmes randomisés qui nous intéresse au premier chef est celle dite de Monte Carlo polynomiale, notée **RP** (pour *random polynomial*). Un tel algorithme permet de répondre en temps polynomial à la question

« x appartient-il¹ à l'ensemble E ? »

Si $x \notin E$, alors l'algorithme répond toujours que x n'est pas dans E . Si $x \in E$, alors l'algorithme le reconnaît avec probabilité supérieure à $1/2$. Itérant alors l'algorithme en faisant k choix indépendants aboutit à rendre la probabilité de succès de l'algorithme itéré aussi proche qu'on veut de 1.

¹On parle d'appartenance à un *langage* en théorie de la complexité.

Les algorithmes de primalité sont répartis en deux catégories : les *tests de composition* répondent à la question **estComposé?** (on en verra un exemple avec le test de Solovay-Strassen). Les *tests de primalité* répondent quant à eux à la question **estPremier?** (c'est le cas de l'algorithme AKS).

La classe RP fournit souvent des algorithmes raisonnables pour résoudre un problème donné. En primalité, il se trouve que les tests de composition sont généralement très rapides en pratique, contrairement aux tests de primalité qui sont souvent plus lourds à mettre en œuvre. Ces tests peuvent souvent fournir un *certificat*, c'est-à-dire des éléments qui permettent de convaincre un observateur extérieur de la justesse du calcul. Nous verrons à la section qui suit des exemples de ces concepts.

Il existe d'autres classes de complexité, la plus intéressante étant celle appelée ZPP (pour *zero-probability polynomial*). Elle contient les problèmes de décision pour lesquels existe un algorithme de type *Las Vegas*. Un tel algorithme répond « oui », « non », ou « je ne sais pas ». La probabilité qu'il réponde « je ne sais pas » peut être rendue aussi petite que possible. Cette classe est l'intersection de RP et de co-RP (co-RP est la classe des problèmes pour laquelle on peut décider de la non appartenance en temps polynomial randomisé). La classe ZPP contient les problèmes qui sont « moralement résolus » par des algorithmes randomisés.

3. FERMAT, LUCAS, LEHMER

3.1. Tests de composition

Le plus simple de ces tests est fondé sur le petit théorème de Fermat : si a est premier avec N et $a^{N-1} \not\equiv 1 \pmod{N}$, alors N n'est pas premier. Cela nous fournit une preuve de composition, qui ne consiste pas en un facteur de N . On peut fabriquer un algorithme de composition de la façon suivante :

fonction estComposéAvecFermat(N)

1. Choisir $a \neq 0$ au hasard dans $\mathbf{Z}/N\mathbf{Z}$.
2. Calculer $g = \text{pgcd}(a, N)$; **si** $g > 1$, **alors** retourner (oui, g est un facteur de N).
3. **si** $a^{N-1} \not\equiv 1 \pmod{N}$, **alors** retourner (oui, a) **sinon** retourner je ne sais pas.

PROPOSITION 3.1. — La probabilité d'échec est $P(N)/(N-1)$ où $P(N) = \prod_i \text{pgcd}(p_i - 1, N - 1)$ si $\prod_i p_i^{\alpha_i}$ est la décomposition de N en facteurs premiers avec $p_i \neq p_j$.

Démonstration. — Il est plus facile d'estimer la probabilité de succès de l'algorithme. La probabilité de trouver un diviseur de N à l'étape 2 est donnée par :

$$\left(1 - \frac{\varphi(N)}{N-1}\right)$$

où $\varphi(N) = \text{Card}((\mathbf{Z}/N\mathbf{Z})^*)$ est la fonction indicatrice d'Euler.

On laisse au lecteur le soin de démontrer que le nombre de $a \in (\mathbf{Z}/N\mathbf{Z})^*$ satisfaisant

$$a^{N-1} \equiv 1 \pmod{N}$$

est $P(N)$. La probabilité de trouver un a prouvant la composition de N à l'étape 3 est donc :

$$\frac{\varphi(N)}{N-1} \left(1 - \frac{P(N)}{\varphi(N)} \right).$$

En additionnant les deux termes, on trouve le résultat. \square

Que conclure de cette proposition ? Tout d'abord, si N est premier, alors l'algorithme échoue toujours à prouver la composition de N , ce qui est rassurant (en effet $P(N) = N - 1$). Ensuite, la fonction $P(N)$ gouverne le succès de l'algorithme. Cette quantité est non triviale, quand N est *nombre pseudopremier en base a* , c'est-à-dire qu'il satisfait $a^{N-1} \equiv 1 \pmod{N}$ pour a fixé. Il peut même arriver que $P(N) = \varphi(N)$, c'est le cas quand N est un *nombre de Carmichael*, dont on sait qu'il existe une infinité [7]. La probabilité d'échec de l'algorithme est donc proche de 1 pour ces nombres, ce qui fait que cet algorithme ne suffit pas à prouver que `estPremier?` est dans RP.

Pour faire mieux, on fait appel au théorème d'Euler, qui nous dit que si N est un nombre premier et a un entier premier avec N , alors

$$(1) \quad a^{(N-1)/2} \equiv \left(\frac{a}{N} \right) \pmod{N},$$

où $\left(\frac{a}{N} \right)$ désigne le symbole de Legendre. Le nombre a étant fixé, un nombre composé N vérifiant (1) (où $\left(\frac{a}{N} \right)$ désigne cette fois le symbole de Jacobi) est appelé *nombre pseudo-premier d'Euler en base a* (noté ppE- a). On pose

$$A_N = \left\{ a \in (\mathbf{Z}/N\mathbf{Z})^*, a^{(N-1)/2} \equiv \left(\frac{a}{N} \right) \pmod{N} \right\}.$$

Si N est premier, alors $A_N = (\mathbf{Z}/N\mathbf{Z})^*$. Par contre, Lehmer [43] a montré que, si N est composé, A_N est un sous-groupe strict de $(\mathbf{Z}/N\mathbf{Z})^*$.

Le test de composition de Solovay et Strassen [70] est le suivant :

fonction `estComposé(N)`

1. Choisir a au hasard dans $\mathbf{Z}/N\mathbf{Z} \setminus \{0\}$.
2. Calculer $g = \text{pgcd}(a, N)$; **si** $g > 1$, **alors** retourner (**oui**, g est un facteur de N).
3. **Si** (1) n'est pas satisfaite **alors** retourner (**oui**, N n'est pas ppE- a) **sinon** retourner **je ne sais pas**.

Cet algorithme est bien dans RP, puisque la probabilité de choisir un bon a est au moins 1/2 (quand N est composé) par le théorème de Lehmer. Quand il réussit, l'algorithme renvoie un *témoin de composition* a .

Miller est allé plus loin [55] (on consultera avec profit [46, 48]). Si l'hypothèse de Riemann pour les fonctions L de Dirichlet pour les caractères réels est vraie, alors le plus

petit témoin est plus petit que $c(\log N)^2$. Bach [10] a montré que $c = 2$ était suffisant. L'algorithme de primalité correspondant est alors :

fonction ESTPREMIERAVECMILLER(N)

1. **pour** $a = 2$ à $2(\log N)^2$ **faire**

i) calculer $g = \text{pgcd}(a, N)$; **si** $g > 1$ **alors** retourner (**non**, g est un facteur de N).

ii) **si** l'équation (1) n'est pas satisfaite **alors** retourner (**non**, N n'est pas ppE- a) ;

2. retourner oui.

Quelle est la complexité de cet algorithme ? Calculer a^e dans un groupe prend $O(\log e)$ opérations de groupe. Multiplier deux entiers de taille $\log N$ prend $O((\log N)^2)$ ou encore $\tilde{O}(\log N)$, la division euclidienne a le même coût. Cela conduit à un temps de calcul $O((\log N)^5)$ ou $\tilde{O}((\log N)^4)$.

Tant que ces hypothèses de Riemann ne sont pas prouvées, cet algorithme ne peut être utilisé dans la pratique et il faut se tourner vers d'autres méthodes.

D'autres tests de composition ont été proposés, comme par exemple l'algorithme préféré des cryptographes, celui d'Artjuhov-Miller-Rabin [8, 55, 63], dont la probabilité de succès est $\geq 3/4$. Nous renvoyons à [29] pour une plus longue liste.

3.2. Preuve de primalité

Les seules méthodes connues et utilisées pour prouver la primalité des entiers jusqu'à la fin des années 1970 étaient basées sur le théorème de Fermat, et des généralisations obtenues par Lucas et Lehmer. Nous ne nous intéresserons pas ici au côté pratique des algorithmes, ni aux avancées plus récentes, préférant renvoyer à la littérature [19, 20, 29].

THÉORÈME 3.2. — N est premier si et seulement si $(\mathbf{Z}/N\mathbf{Z})^*$ est cyclique d'ordre $N - 1$.

PROPOSITION 3.3. — Supposons connus les facteurs premiers p_i de $N - 1$. Le nombre $g \in (\mathbf{Z}/N\mathbf{Z})^*$ est d'ordre $N - 1$ si et seulement si $g^{N-1} \equiv 1 \pmod{N}$ et $g^{(N-1)/p_i} \not\equiv 1 \pmod{N}$ pour tout i .

On utilise généralement ce théorème de concert avec un résultat de Pocklington

THÉORÈME 3.4. — Soient s tel que $s \mid N - 1$ et a tel que $a^{N-1} \equiv 1 \pmod{N}$, et pour tout q premier divisant s , $\text{pgcd}(a^{\frac{N-1}{q}} - 1, N) = 1$. Alors tout diviseur premier p de N vérifie $p \equiv 1 \pmod{s}$.

Le corollaire est le plus utile :

COROLLAIRE 3.5. — Avec les hypothèses du théorème précédent, si $s > \sqrt{N}$, alors N est premier.

L'utilisation pratique de ces résultats nécessite la factorisation de $N - 1$ qui est difficile à déterminer, le meilleur algorithme de factorisation connu à ce jour, le crible algébrique [45, 29] ayant une complexité sous-exponentielle. Même si cette factorisation est connue, il reste à trouver g , ce qu'on ne sait pas faire actuellement en temps polynomial déterministe, à moins d'admettre l'hypothèse de Riemann. Pour l'anecdote, l'algorithme randomisé de Shor [69] qui factorise en temps polynomial dans le modèle quantique ne s'attaque qu'au problème de la factorisation [22].

Malgré tout, ces tests permettent de traiter les nombres tels que $N - 1$ soit facilement factorisable, c'est-à-dire de la forme $N_1 \prod p_i^{\beta_i}$, où les p_i sont de petits nombres premiers, et N_1 probablement premier au sens de la partie 2.1. On construit ainsi une suite décroissante d'entiers $(N_i)_{1 \leq i \leq k}$ avec $N_0 = N$, et N_{i+1} un facteur probablement premier de $N_i - 1$. La primalité de N_k étant prouvée, il ne reste plus qu'à remonter pour prouver de proche en proche celles de tous les N_i , en terminant par N_0 . Une telle suite est appelée DOWNRUN par Selfridge.

La proposition 3.3 est utilisée en complexité. En effet, la donnée de g et des facteurs de $N - 1$ accompagnés de leurs certificats respectifs permet de montrer que le problème `estPremier?` appartient à la classe NP (voir [62]).

On peut généraliser cette idée de base au cas où on chercherait à construire un corps fini de degré plus élevé, en utilisant la factorisation de $N + 1$, ou celle de $\Phi_k(N)$ pour le k -ième polynôme cyclotomique. Cela donne par exemple l'algorithme de primalité de Lucas-Lehmer qui est très rapide pour les nombres de Mersenne. Ce n'est pas un hasard si ces nombres font régulièrement la une des journaux. Ainsi, le plus grand nombre premier connu, à l'instant où je tape ces lignes², est $M_{13466917}$, qui a 4053946 chiffres décimaux³.

Tous ces tests apparaissent maintenant comme des cas particuliers d'un théorème de Lenstra, énoncé pour la première fois dans [47] et développé dans [50].

THÉORÈME 3.6. — *Soit $s > 0$. Soit \mathbf{A} un anneau contenant $\mathbf{Z}/N\mathbf{Z}$ comme sous-anneau. Supposons qu'il existe $\alpha \in \mathbf{A}$ tel que :*

$$\begin{aligned} \alpha^s &= 1, \\ \alpha^{s/q} - 1 &\in \mathbf{A}^* \text{ pour tout } q \text{ premier } \mid s, \\ \Psi_\alpha(X) = \prod_{i=0}^{t-1} (X - \alpha^{N^i}) &\in \mathbf{Z}/N\mathbf{Z}[X], \text{ pour un certain } t > 0. \end{aligned}$$

Alors : $\forall r \mid N, \exists i, r = N^i \bmod s$.

Le cas le plus classique d'application est celui où \mathbf{A} est une extension d'anneau, construite à l'aide d'un polynôme à coefficients dans $\mathbf{Z}/N\mathbf{Z}$. Si N est premier, \mathbf{A} n'est autre que le corps fini \mathbf{F}_{N^t} . En particulier, le cas $t = 1$ nous redonne le théorème 3.3.

²24 février 2003

³<http://www.utm.edu/research/primelargest.html>

4. VINGT ANS DE PRIMALITÉ

Par opposition à la période précédente, les vingt dernières années ont vu l'écllosion et le développement de deux tests de primalité praticables pour des nombres entiers *généraux* (par opposition aux nombres pour lesquels $N - 1$ ou $N + 1$ ont des factorisations faciles). Le premier algorithme historiquement, celui des sommes de Gauss/Jacobi, se décline en deux versions, déterministe ou randomisé, avec un temps de calcul quasi-polynomial, mais ne fournit pas de certificat. La recherche d'algorithmes polynomiaux a été ainsi une motivation pour chercher de nouvelles approches.

Le second algorithme utilise les courbes elliptiques et hyperelliptiques. Il est de type Monte Carlo polynomial. Il en existe également une version pratique utilisant des courbes elliptiques à multiplication complexe, fournissant un certificat, mais dont la complexité polynomiale est heuristique.

4.1. Sommes de Gauss, Jacobi

En 1980, Adleman [1] ébauche un algorithme de primalité utilisant des lois de réciprocité d'ordre supérieur, dans une version randomisée. Rejoint par Pomerance et Rumely, ces travaux conduisent à l'article [3] dans lequel une version déterministe est également présentée. Un temps de calcul $O((\log N)^{c \log \log \log N})$ est prouvé pour les deux versions.

Décrivons brièvement les idées de base. Il n'est pas dans notre propos de donner tous les détails, contenus dans les articles déjà mentionnés ou bien dans [24]. Soient p, q des nombres premiers, $p^k \parallel q - 1$, $\text{pgcd}(pq, N) = 1$. Soit χ un caractère multiplicatif d'ordre p^k et de conducteur q :

$$\chi : \mathbf{F}_q^* \rightarrow \mathbf{C}$$

défini par sa valeur en g , un générateur de \mathbf{F}_q^* , par $\chi(g) = \zeta_{p^k}$, une racine primitive p^k -ième de l'unité fixée. La somme de Gauss attachée à χ est :

$$\tau(\chi) = \sum_{x=1}^{q-1} \chi(x) \zeta_q^x$$

à valeur dans $R = \mathbf{Z}[\zeta_{p^k}, \zeta_q]$. C'est un objet bien connu en théorie de la cyclotomie (voir par exemple [38]). On construit alors un test de pseudoprimauté en utilisant la proposition élémentaire suivante :

PROPOSITION 4.1. — *Si N est premier, $\text{pgcd}(N, pq) = 1$, alors*

$$(2) \quad \frac{\tau(\chi)^N}{\tau(\chi^N)} = \chi(N)^{-N},$$

la relation s'interprétant dans $\mathbf{Z}/N\mathbf{Z}[\zeta_{p^k}, \zeta_q]$.

L'idée de l'algorithme est alors de tester les identités (2) pour tous les couples (p, q) avec $p^k \parallel q - 1$ choisis de sorte que $s = \prod_{q \in \mathcal{Q}} > \sqrt{N}$. Si elles sont toutes vérifiées (ainsi

que des conditions techniques négligées ici), alors tout diviseur r de N appartient au groupe multiplicatif $\langle N \bmod s \rangle$, d'ordre $t = \text{ppcm}_{q \in \mathcal{Q}}(q - 1)$. Il ne reste plus qu'à vérifier qu'aucun élément de cet ensemble n'est un diviseur premier non trivial de N .

Le coût de l'algorithme est déterminé par cette dernière phase, avec un coût t , qui domine le reste de l'algorithme, constitué de (nombreuses) opérations sur des polynômes de relativement petit degré. Pour diminuer le coût, on commence par chercher t le plus petit possible, avec le plus de diviseurs possibles, et tel que

$$s(t) = \prod_{q-1|t} q$$

soit plus grand que \sqrt{N} . Un théorème d'Odlyzko et Pomerance prouve le temps de calcul intrinsèque de l'algorithme.

THÉORÈME 4.2. — *Il existe $c_1, c_2 > 0$ tels qu'on puisse trouver t convenable avec*

$$(\log N)^{c_1 \log \log \log N} \leq t \leq (\log N)^{c_2 \log \log \log N}.$$

H.W. Lenstra, Jr., a amélioré l'algorithme dans une série d'articles dont le premier est [47]. Avec H. Cohen [26], il donne une version beaucoup plus pratique de l'algorithme, remplaçant l'utilisation originale des sommes de Gauss par celle des sommes de Jacobi, qui se prêtent beaucoup mieux aux calculs (elles sont naturellement dans $\mathbf{Z}[\zeta_{p^k}]$). Cet algorithme, implanté par H. Cohen et A.K. Lenstra [25], est le premier à avoir prouvé efficacement la primalité de nombres quelconques de 100 à 200 chiffres décimaux, ce qui représentait un progrès remarquable. Par la suite, les pistes esquissées à la fin de [47] et développées dans [49, 50] ont été explorées avec succès par W. Bosma & M.-P. van der Hulst [18], ainsi que par P. Mihăilescu [54]. Elles concernent principalement l'utilisation pratique du concept d'extension cyclotomique d'un anneau.

4.2. Courbes elliptiques

4.2.1. *Préparation théorique.* Soit E une courbe elliptique définie sur $\mathbf{Z}/p\mathbf{Z}$, $p > 3$. L'ensemble des points de E est $E(\mathbf{Z}/p\mathbf{Z}) = \{(x : y : z) \in \mathbf{P}^2(\mathbf{Z}/p\mathbf{Z}), y^2z = x^3 + axz^2 + bz^3\}$, sur lequel on définit la loi de groupe habituelle, notée $+$. La multiplication par k sur E sera notée $[k]$.

On sait d'après Hasse que le cardinal de $E(\mathbf{Z}/p\mathbf{Z})$ vérifie :

$$|\#E(\mathbf{Z}/p\mathbf{Z}) - (p + 1)| < 2\sqrt{p}.$$

Inversement, d'après Deuring (cf. aussi [72]), pour tout entier t vérifiant $|t| < 2\sqrt{p}$, il existe une courbe E dont l'ensemble des points sur $\mathbf{Z}/p\mathbf{Z}$ a cardinal $p + 1 - t$. Malheureusement, ce théorème ne fournit pas d'algorithme efficace pour trouver E en fonction de t . Si c'était le cas, la primalité serait facile [60].

Pour pouvoir utiliser des courbes elliptiques en primalité, il est nécessaire de pouvoir calculer la cardinalité d'une courbe dans $\mathbf{Z}/p\mathbf{Z}$. Les méthodes élémentaires (formule dite

de Lang et Trotter, algorithme de Shanks) pour cela ne marchent pas en temps polynomial déterministe, et il a fallu attendre l'algorithme de Schoof [67] pour résoudre le problème. En quelques mots, cet algorithme consiste à calculer $t_\ell = t \bmod \ell$ pour de petits nombres premiers ℓ en nombre suffisant pour que $\prod \ell > 4\sqrt{p}$, ce qui permet de déterminer t par application du théorème Chinois.

Il faut également définir ce qu'est une courbe elliptique sur $\mathbf{Z}/N\mathbf{Z}$ quand N est composé, ce que Lenstra a fait dans [51]. On définit

$$E(\mathbf{Z}/N\mathbf{Z}) = \{(x : y : z) \in \mathbf{P}^2(\mathbf{Z}/N\mathbf{Z}), y^2z = x^3 + axz^2 + bz^3\}.$$

On notera $\Delta(E) = 4a^3 + 27b^2$ et on demande que Δ soit inversible modulo N . Notons que le point $O_E = (0 : 1 : 0)$ appartient à $E(\mathbf{Z}/N\mathbf{Z})$. On l'appelle point origine de la courbe. On définit alors une loi de groupe sur $E(\mathbf{Z}/N\mathbf{Z})$ en utilisant des formules données dans [42].

On remarque que si p premier divise N , alors on peut réduire E modulo p et envoyer $E(\mathbf{Z}/N\mathbf{Z})$ dans $E(\mathbf{Z}/p\mathbf{Z})$ par réduction des coordonnées modulo p ; le point origine O_E se réduit en le point à l'infini de la réduction de E modulo p .

Il nous reste maintenant à donner un théorème de primalité.

THÉORÈME 4.3. — *Soient m et s deux entiers tels que $s \mid m$, E une courbe elliptique sur $\mathbf{Z}/N\mathbf{Z}$ et P un point de $E(\mathbf{Z}/N\mathbf{Z})$. Si $[m]P = O_E$ et si pour tout diviseur premier q de s , on a $[m/q]P = (X : Y : Z)$ et $\text{pgcd}(Z, N) = 1$, alors pour tout diviseur premier p de N , on a $\#E(\mathbf{Z}/p\mathbf{Z}) \equiv 0 \pmod{s}$.*

COROLLAIRE 4.4. — *Si $s > (\sqrt[4]{N} + 1)^2$, alors N est premier.*

Ces résultats généralisent les énoncés de Pocklington. On peut désormais utiliser des facteurs premiers de cardinaux de courbes elliptiques pour fabriquer un DOWNRUN. On en déduit également un certificat de primalité généralisé.

4.2.2. L'algorithme de Goldwasser et Kilian. Cet algorithme a été introduit dans [35] (voir la version finale en [36]). Il est décrit à la figure 1. Cet algorithme est paramétré par les deux nombres B_1 et B_2 . Le nombre B_1 gouverne le nombre d'essais que l'on s'autorise dans l'algorithme. Pour avoir un algorithme dans \mathbf{RP} , ce nombre doit être polynomial en $\log N$. Nous y revenons plus loin. Le nombre B_2 sert à contrôler le nombre d'essais nécessaires à trouver un bon point sur E . Si N est premier, alors la probabilité de trouver un P convenable à cette étape est plus grande que $1/2$.

Que se passe-t-il quand N est composé ? L'algorithme de Schoof teste des identités entre polynômes, à la recherche de t_ℓ . Lors des calculs, il peut rarement apparaître un facteur premier de N , quand une inversion modulo N échoue. Plus sûrement, aucune valeur de t_ℓ ne sera trouvée et cela fournira une preuve de composition pour N . On pourrait imaginer de définir un nombre pseudopremier de Schoof pour (E, ℓ) comme étant un nombre pour

fonction ESTPREMIERAVECGK(N)

1. **répéter** B_1 fois

- (a) choisir a, b au hasard modulo N et calculer $g = \text{pgcd}(4a^3 + 27b^2, N)$;
- (b) **si** $g = N$ **alors** aller à (a) ;
- (c) **si** $g \neq 1$ **alors** retourner (non, g) ;
- (d) soit E la courbe elliptique d'équation $y^2 = x^3 + ax + b$ définie sur $\mathbf{Z}/N\mathbf{Z}$;
- (e) calculer $m = \#E(\mathbf{Z}/N\mathbf{Z})$ avec l'algorithme de Schoof ;
- (f) **si** $m = 2q$, avec q probablement premier **alors**

(α) **répéter** B_2 fois

- choisir au hasard $P \neq O_E$ sur E ;
- **si** $[m]P \neq O_E$ **alors** retourner non ;
- calculer $[q]P = (X_q : Y_q : Z_q)$ et $g = \text{pgcd}(Z_q, N)$;
- **si** $g = 1$ **alors** aller à (β) ;
- **si** $1 < g < N$ **alors** retourner (non, g) ;

(β) **si** GK(q)==oui **alors** retourner oui ;

2. retourner je ne sais pas.

FIG. 1. L'algorithme GK.

lequel on trouve une valeur de t_ℓ . À titre d'exemple, le plus petit N pseudopremier pour $E : y^2 = x^3 + x + 1$ et $\ell = 3$ est $N = 3481 = 59^2$.

L'analyse de l'algorithme GK fait appel au résultat suivant, prouvé par Lenstra [52].

THÉORÈME 4.5. — Pour $S \subset \mathbf{N}$, soit $S'_p = S \cap [p - \sqrt{p}, p + \sqrt{p}]$. Il existe $c > 0$ tel que

$$\text{Prob}(\#E(\mathbf{Z}/p\mathbf{Z}) \in S) \geq \frac{c(\#S'_p - 2)}{\sqrt{p} \log p},$$

la probabilité se calculant en choisissant (a, b) uniformément dans $\mathbf{Z}/p\mathbf{Z}$ de sorte que $4a^3 + 27b^2 \neq 0$.

Pour analyser GK, on s'intéresse alors à $S = \{m = 2q, q \text{ premier}\}$, ce qui nous ramène à étudier S'_p . Pour assurer la terminaison de GK dans tous les cas, il faut conjecturer le résultat suivant :

CONJECTURE 4.6. — Il existe $c_1, c_2 > 0$ tels que pour $x > x_0 : \pi(x + \sqrt{x}) - \pi(x) \geq \frac{c_2 \sqrt{x}}{\log^{c_1} x}$.

Avec cette conjecture, il nous suffirait de prendre pour B_1 une quantité polynomiale en $\log N$. On obtiendrait alors :

THÉORÈME 4.7. — Si la conjecture 4.6 est vraie, alors GK termine en temps polynomial pour tous les nombres.

La conjecture est beaucoup plus forte que ce que peut apporter l'hypothèse de Riemann classique, mais suivrait de la conjecture de Cramér.

On ne sait d'ailleurs pas prouver que S'_p est non vide. Rappelons que le plus petit exposant δ pour lequel on puisse prouver que $\pi(x + x^\delta) - \pi(x) > 0$ est $\delta = 0.525$ (voir [12]). Il est clair cependant que certains petits intervalles doivent contenir au moins un nombre premier. En utilisant les résultats de Heath-Brown [37] sur la différence entre nombres premiers consécutifs, il est possible de montrer :

THÉORÈME 4.8. — GK termine en temps $O((\log N)^9)$ en moyenne pour les nombres premiers $\leq x$, sauf pour ceux de l'ensemble $\mathcal{E}(x)$ de cardinal

$$\#\mathcal{E}(x) \ll \frac{x/\log x}{2^{\frac{\log \log x}{\log \log \log x}}}.$$

4.2.3. *L'amélioration d'Adleman et Huang.* Superficiellement, cette amélioration (décrite en [2, p. 136]), consiste à remplacer la condition $m = 2q$ par $m = \ell q$ avec ℓ premier petit. En utilisant les mêmes travaux de Heath-Brown, et d'autres de Pomerance [59], ils sont à même de montrer que si $\mathcal{E}'(x)$ est l'ensemble des entiers plus petits que x non prouvables par leur algorithme GK-AH, alors :

THÉORÈME 4.9. — Pour x suffisamment grand, $\#\mathcal{E}'(x) < x^{15/16}$.

4.3. Courbes de genre 2

Comme l'algorithme GK-AH pourrait ne pas terminer, Adleman et Huang ont eu l'idée d'utiliser les courbes de genre 2, pour lesquelles les cardinalités des groupes associés varient cette fois dans un intervalle beaucoup plus grand, $[x^2 - x^{1.5}, x^2]$ au lieu de $[x, x + x^{0.5}]$ comme en genre 1.

Soit p un nombre premier > 3 . Nous nous intéresserons aux courbes $C(f)$ définies sur $\mathbf{Z}/p\mathbf{Z}$ par une équation du type $y^2 = f(x)$ où f est un polynôme de degré 6 à coefficients dans $\mathbf{Z}/p\mathbf{Z}$ et sans racines multiples. Nous noterons $\mathcal{D}(f)$ le cardinal de la jacobienne de $C(f)$. Par Hasse-Weil, on sait que $(\sqrt{p} - 1)^4 < \mathcal{D}(f) < (\sqrt{p} + 1)^4$.

On peut maintenant imaginer généraliser l'algorithme de Schoof au calcul de $\mathcal{D}(f)$, ce qui fournit là encore un algorithme polynomial déterministe. Comme dans le cas du genre 1, l'algorithme ne fournit pas de résultat exploitable quand N est composé, ce qui permet de le repérer.

Nous avons maintenant tous les outils pour pallier le problème de terminaison (hypothétique, rappelons-le) de GK. L'idée de base consiste à construire une suite de nombres premiers $(q_i)_{1 \leq i \leq r}$ tels que la primalité de q_r entraîne celle de q_{r-1}, \dots , celle de q_1 celle de N . Contrairement à ce qu'il se passe classiquement, cette fois, la suite q_i sera *croissante* ! En effet, on commence par construire q_1 comme le cardinal $\mathcal{D}(f_1)$ d'une jacobienne définie sur $\mathbf{Z}/N\mathbf{Z}$. De même q_2 est construit comme un cardinal premier $\mathcal{D}(f_2)$ d'une jacobienne définie sur $\mathbf{Z}/q_1\mathbf{Z}$, etc. L'algorithme s'arrête quand un des q_i est prouvable par GK-AH.

Nous utilisons ainsi dans notre algorithme un sous-algorithme qui est de type Monte Carlo polynomial.

Le calcul de la probabilité de succès de cet algorithme n'est pas une mince affaire, et nécessite de commencer par généraliser le théorème 4.5. Une fois cela fait, il ne reste plus qu'à estimer le nombre de nombres premiers dans un intervalle très grand, cette fois $I(x) = [x^2 - x^{1.5}, x^2]$. Il suffit d'utiliser un résultat d'Iwaniec et Jutila [39], pour minorer le nombre de nombres premiers dans $I(x)$ par $cx^{1.5}/\log x$ pour x assez grand. C'est là l'une des clefs de l'algorithme AH.

On calcule alors le nombre de r -uplets (q_1, \dots, q_r) construits par l'algorithme à partir d'un nombre initial $q_0 = p$. Le calcul montre que, pour $r = 3$, la proportion des bons r -uplets devient significative, et l'algorithme a une probabilité de réussite plus grande que $1/(\log p)^c$ pour un certain entier c , ce qui achève de démontrer :

THÉORÈME 4.10. — *Il existe un algorithme de type Monte Carlo polynomial qui prouve la primalité de tous les entiers.*

Il existe donc finalement un algorithme de Las Vegas pour la primalité.

4.4. La primalité en pratique

4.4.1. *Les sommes de Jacobi.* L'algorithme des sommes de Jacobi a été relativement peu implanté. Une des raisons est qu'il s'avère difficile de convaincre de la véracité des calculs effectués, puisqu'aucun certificat n'est délivré. Il faut programmer soi-même l'algorithme, ce qui est chose délicate et qui ne convainc que son auteur. C'est un cas où on rêverait que la technologie de la preuve de correction des programmes soit opérationnelle à ce niveau (voir [21] pour une piste).

L'entier t utilisé dans la description de l'algorithme est le paramètre critique. On doit calculer des sommes de Jacobi associées à des nombres premiers q tels que $q - 1 \mid t$, donc potentiellement de la taille de t . Pour fixer les idées, un nombre de 100 chiffres décimaux nécessite $t = 8400$; si $N \approx 10^{1000}$, alors $t \approx 10^8$ et $N \approx 10^{6000}$ nécessite plutôt $t \approx 6 \cdot 10^9$. L'une des caractéristiques de l'algorithme est la possibilité de précalculer ces sommes une fois pour toutes. Notons que la taille mémoire du résultat n'est pas un problème, même si le temps de calcul et la mémoire nécessaire pour ces précalculs peuvent être très grands. Pour ces grandes tailles, l'utilisation de la factorisation dans les corps cyclotomiques devient rentable [71].

Au-delà de ces premières remarques, il faut utiliser les améliorations présentées dans [18, 54] pour aller plus vite. Parmi celles-ci, notons l'utilisation de facteurs de $N^w - 1$ pour w petit. P. Mihăilescu détient le record actuel avec le nombre $N = 2^{10000} + 177$ (3011 chiffres décimaux) depuis novembre 1997.

4.4.2. *Les courbes.* L'algorithme de Schoof n'est pas implantable tel quel. Il a fallu attendre les améliorations d'Elkies et Atkin pour disposer d'un algorithme de calcul de la

cardinalité de courbes elliptiques fonctionnant en temps polynomial, pratique et efficace, mais cette fois randomisé (voir [68, 30, 56, 17] par exemple). Malgré ces améliorations, l'algorithme SEA ne permet pas d'envisager de traiter des nombres de plus de quelques centaines de chiffres décimaux, et l'algorithme GK, même dopé par ces changements, n'est pas efficace.

Dès 1986, Atkin a eu l'idée d'utiliser la réduction de courbes elliptiques à multiplication complexe, dont on sait calculer la cardinalité facilement. Ces courbes remplacent les courbes aléatoires utilisées dans GK-AH. L'algorithme, appelé ECPP, est décrit dans [9, 24, 29] (voir également [28]), des améliorations ayant été données au cours du temps par l'auteur (consulter sa page web par exemple). ECPP a été analysé de façon heuristique dans [44] : la complexité en temps devrait être de l'ordre de $O((\log N)^{6+\varepsilon})$ ou $\tilde{O}((\log N)^5)$ avec de l'arithmétique rapide ($\tilde{O}((\log N)^4)$ avec des précalculs, voir également [57]).

Deux implantations sont connues : celle de l'auteur, maintenue depuis 12 ans, et dont une version se trouve dans le logiciel MAGMA, la seconde réalisée par Marcel Martin (PRIMO⁴). Le record actuel, dû à l'auteur, est la primalité du nombre $2177^{580} + 580^{2177}$, qui a 6016 chiffres décimaux (cf. [57])⁵.

L'algorithme AH est-il pratique? Rappelons qu'il sert essentiellement de bouée de sauvetage théorique pour l'algorithme GK-AH. Bien sûr, on pourrait imaginer y avoir recours dans le cas où on ne trouverait pas de bonnes courbes dans ECPP. Pour cela, il faudrait songer à remplacer la recherche des petits facteurs de nombres de la taille de N , par des nombres de taille N^2 , ou même N^8 , tout cela avec de multiples allers et retours entre les différents algorithmes...

Les problèmes d'implantation sont pour le moment formidables. Les meilleures implantations d'algorithmes à la Schoof pour des courbes de genre 2 permettent au plus de traiter le cas de $p = 5 \cdot 10^{24} + 41$ (voir le récent record de Gaudry et Schost annoncé dans la liste NMBRTHRY@LISTSERV.NODAK.EDU allant au-delà de [33]). Quand bien même on mettrait au point une variante d'ECPP en genre 2, appelons-là HECPP (et qui serait elle aussi heuristique), les calculs liés à la multiplication complexe dans ce cadre sont encore loin d'être aussi faciles et agréables que dans le cas du genre 1 (cf. [73]).

4.5. Conclusion sur la pratique

Pour des nombres relativement petits, disons plus petits que 10^{1000} , les deux algorithmes (sommées de Jacobi, ECPP) ont des temps de calcul raisonnables, n'excédant pas une dizaine d'heures sur une machine normale. Le problème pratique de prouver la primalité est donc résolu. Pour des nombres plus grands, les résultats dépendent de beaucoup trop de paramètres pour que l'auteur se risque à des réponses définitives.

⁴<http://www.ellipsa.net/>

⁵Le précédent record était détenu par Jose Luis Gomez Pardo, avec PRIMO, pour un nombre de 5878 chiffres décimaux, en février 2003; cf. <http://www.ellipsa.net/pages/primorecord.html>.

5. LES TRAVAUX D'AGRAWAL, KAYAL, SAXENA

5.1. Première idée

Dans [5], les auteurs présentent un test de composition très simple utilisant la proposition suivante :

PROPOSITION 5.1. — *N est premier si et seulement si le polynôme $P(X) = (X + 1)^N - X^N - 1$ est identiquement nul modulo N .*

Démonstration. — Il suffit pour cela de se rappeler que si N est premier, alors $N \mid \binom{N}{k}$ pour tout entier k , $0 < k < N$. Si N est composé, soit p un de ses facteurs premiers, et a la valuation p -adique de N . Alors $\binom{N}{p}$ est divisible exactement par p^{a-1} . \square

Cette proposition ne conduit pas à un algorithme efficace, puisque le polynôme $P(X)$ a un degré trop élevé. Par contre, on peut élaborer un test de composition de la façon suivante :

fonction ESTPREMIERAB(N)

1. Si N est une puissance de nombre entier, retourner **non**.
2. Choisir un polynôme aléatoire $Q(X)$ de degré $O(\log N)$ à coefficients dans $\mathbf{Z}/N\mathbf{Z}$. Si $(X + 1)^N \equiv X^N + 1 \pmod{(Q(X), N)}$ alors retourner **oui**, sinon retourner **non**.

Les auteurs montrent alors que la probabilité d'échec de l'algorithme est bornée par $1 - 1/(4 \log N)$, là où Solovay et Strassen donnent $1/2$. Ils conjecturent également :

CONJECTURE 5.2. — *Si N est composé, alors il existe $1 \leq r \leq \log N$ tel que $P(X)$ n'est pas divisible par $X^r - 1$ modulo N .*

Notons que le test généralise le test de Fermat pour $r = 1$, dans ce cas il est équivalent à $2^N \equiv 2 \pmod N$. Si N passe le test pour $r = 1$, il le passe aussi pour $r = 2$ (car $(X + 1)^N = X^N + 1 \pmod{(X + 1, N)}$ quand N est impair).

L'auteur a testé cette conjecture sur les nombres pseudopremiers⁶ en base 2 plus petits que 10^{12} à l'aide de MAPLE. Il y a 24 nombres qui passent les tests pour $1 \leq r \leq 8$, 8 étant la valeur maximale atteinte. Le plus petit est 597717121, le plus grand 880731910801 (notons que $\log 597717121 \approx 20.20$).

5.2. Le théorème original

C'est celui que l'on peut extraire de [6], suivant en cela [13] (remplacé depuis par [14]) et inspiré en partie par [64]. L'idée est de réussir à combiner des tests de pseudoprimauté pour obtenir une preuve de primalité pour N .

⁶Grâce au fichier envoyé il y a fort longtemps par S. S. Wagstaff, Jr.

THÉORÈME 5.3. — Soient N un entier qui ne s'écrit pas M^k pour M et k entiers, $k > 1$. Soient s un entier positif, r un nombre premier et q le plus grand facteur premier de $r - 1$.

On suppose que

(i) [condition arithmétique] $N^{(r-1)/q} \bmod r \notin \{0, 1\}$;

(ii) [condition combinatoire]

$$(3) \quad \binom{q-1+s}{s} > N^{2\lfloor\sqrt{r}\rfloor}.$$

(iii) [divisibilité élémentaire] N n'a pas de facteur premier $p \leq s$;

(iv) [tests de pseudoprimauté] $(X-a)^N \equiv X^N - a \bmod (X^r - 1, N)$ pour tout $1 \leq a \leq s$.

Alors N est premier.

Démonstration. — Supposons que N soit composé. Il existe un facteur premier p de N tel que $p^{(r-1)/q} \bmod r \notin \{0, 1\}$, car sinon cela contredirait (i). D'après (iii), on a également $p > s$.

Par réduction modulo p , la propriété (iv) implique que pour tout a dans $\mathcal{A} = \{1, \dots, s\}$:

$$(X-a)^N \equiv X^N - a \bmod (X^r - 1, p).$$

Comme p est premier, on a aussi

$$(X-a)^p \equiv X^p - a \bmod (X^r - 1, p).$$

L'idée est de combiner ces deux relations.

LEMME 5.4. — Soient m_1 et m_2 deux entiers. Si

$$(X-a)^{m_1} \equiv X^{m_1} - a \bmod (X^r - 1, p),$$

$$(X-a)^{m_2} \equiv X^{m_2} - a \bmod (X^r - 1, p),$$

alors $(X-a)^{m_1 m_2} \equiv X^{m_1 m_2} - a \bmod (X^r - 1, p)$.

Démonstration. — Il existe $g(X) \in \mathbf{F}_p[X]$ tel que

$$(X-a)^{m_2} - (X^{m_2} - a) = (X^r - 1)g(X).$$

D'où :

$$(X^{m_1} - a)^{m_2} - (X^{m_1 m_2} - a) = (X^{m_1 r} - 1)g(X^{m_1}).$$

Comme $X^r - 1 \mid X^{m_1 r} - 1$, on en déduit

$$(X-a)^{m_1 m_2} \equiv (X^{m_1} - a)^{m_2} \equiv X^{m_1 m_2} - a \bmod (X^r - 1, p).$$

□

On déduit de cela qu'en fait, pour tous les entiers positifs i, j , et tout $a \in \mathcal{A}$:

$$(X - a)^{p^{iN^j}} \equiv X^{p^{iN^j}} - a \pmod{(X^r - 1, p)}.$$

On utilise alors l'argument combinatoire suivant. Posons $L = \{p^{iN^j}, 0 \leq i, j \leq \lfloor \sqrt{r} \rfloor\}$. Comme N n'est pas une puissance d'un nombre entier, il ne s'écrit *a fortiori* pas comme p^k . Par suite, tous les éléments de L sont distincts et $\#L = (\lfloor \sqrt{r} \rfloor + 1)^2 > r$. Il existe donc deux éléments qui sont congrus modulo r , ce qu'on écrit :

$$m_1 = p^{i_1 N^{j_1}}, m_2 = p^{i_2 N^{j_2}} = m_1 + kr, (i_1, j_1) \neq (i_2, j_2)$$

avec par exemple $m_1 \leq m_2$. On obtient alors pour tout $a \in \mathcal{A}$:

$$(X - a)^{m_2} \equiv X^{m_1+kr} - a \equiv X^{m_1} - a \equiv (X - a)^{m_1} \pmod{(X^r - 1, p)}.$$

Il ne nous reste plus qu'à démontrer que $m_1 = m_2$, ce qui impliquera $N = p^t$, contredisant l'hypothèse faite.

À ce stade, on choisit un facteur irréductible du r -ième polynôme cyclotomique modulo p , soit $h(X)$. On sait que $h(X)$ est de degré d , l'ordre de p modulo r (voir par exemple [38]). Par le choix de p , $d \geq q$. On considère $F = \mathbf{F}_p[X]/(h(X))$ le corps fini à p^d éléments, et $\theta = X \pmod{(h(X), p)}$. Par hypothèse :

$$\forall a \in \mathcal{A}, (X - a)^{m_1} \equiv (X - a)^{m_2} \pmod{(h(X), p)}.$$

On introduit alors le monoïde S de F^* engendré par les $\theta - a$ pour $a \in \mathcal{A}$.

LEMME 5.5. — Soit T le sous-ensemble de S formé des produits $\prod_{a=1}^s (\theta - a)^{\alpha_a}$ avec

$$(4) \quad \sum_{a=1}^s \alpha_a \leq q - 1, \alpha_a \geq 0.$$

Alors

- (i) les éléments de T sont tous distincts;
- (ii) le cardinal de T est $\binom{q-1+s}{s}$.

Démonstration. — (i) Tous les $X - a$ pour $a \in \mathcal{A}$ sont irréductibles et distincts dans $\mathbf{F}_p[X]$, puisque $p > s$. Tous les polynômes $\prod_{a=1}^s (X - a)^{\alpha_a}$ avec les α_a satisfaisant (4) sont de degré strictement plus petit que d , et deux à deux distincts dans $\mathbf{F}_p[X]$. Deux d'entre eux ne peuvent être égaux dans F que s'ils diffèrent d'un multiple de $h(X)$, ce qui est impossible car leur degré est $< q \leq d$.

(ii) C'est un résultat classique de combinatoire. On met les solutions de (4) en bijection avec les parties à s éléments de l'intervalle $[1, q - 1 + s]$ de la façon suivante. Si (α_a) est une solution, on pose $\beta_1 = \alpha_1 + 1, \beta_2 = \alpha_1 + \alpha_2 + 2, \dots, \beta_s = \alpha_1 + \alpha_2 + \dots + \alpha_s + s$. La suite β_a vérifie $1 \leq \beta_1 < \beta_2 < \dots < \beta_s \leq q - 1 + s$. Le nombre de suites (β_a) possibles est bien la quantité annoncée, appelée traditionnellement nombre de combinaisons avec répétition [27]. \square

Retour au théorème.

Les éléments de S et *a fortiori* de T sont tous racines du polynôme $Y^{m_1} - Y^{m_2} = Y^{m_1}(Y^{m_2-m_1} - 1) = Y^{m_1}P(Y)$. Comme $m_2 - m_1 \leq N^{2\lfloor\sqrt{r}\rfloor} < \binom{q-1+s}{s} \leq \#T$, le polynôme $P(Y)$ a plus de racines que son degré, donc il est identiquement nul et $m_1 = m_2$. \square

Il est facile de déduire un algorithme du théorème 5.3, si on se donne r et s , ce que nous ferons dans la section qui suit.

PROPOSITION 5.6. — *Le temps de calcul de AKS est $O(sr^2(\log N)^3)$ si on utilise des algorithmes classiques de multiplication et $\tilde{O}(sr(\log N)^2)$ si on utilise de la multiplication rapide, aussi bien pour les polynômes que pour les entiers.*

Démonstration. — Le temps de calcul est largement dominé par le temps passé à vérifier la condition (iv). Pour chaque valeur de a , le calcul de $(X-a)^N \bmod (X^r-1, N)$ demande $O(\log N)$ multiplications $A(X)B(X) \bmod (X^r-1)$ où $A(X)$ et $B(X)$ ont degré $O(r)$. La réduction ne coûte rien car X^r-1 est creux. Notant $\mathcal{P}(N, r)$ le temps nécessaire à multiplier deux polynômes de degré r à coefficients dans $\mathbf{Z}/N\mathbf{Z}$, le temps de calcul est donc $O(s(\log N)\mathcal{P}(N, r))$. En utilisant les algorithmes naïfs de multiplication, on obtient un temps de calcul $O(sr^2(\log N)^3)$. Avec des FFT partout, cela devient $\tilde{O}(sr(\log N)^2)$. \square

5.3. Le choix de r et s

Pour achever de démontrer que la primalité est décidable en temps polynomial déterministe, il nous faut montrer que l'on peut choisir r et s comme puissances de $\log N$.

On commence par minorer brutalement $\binom{q-1+s}{s}$ par $(q/s)^s$, puis on impose $q \geq 2s$. Si on choisit s tel que $2^s \geq N^{2\lfloor\sqrt{r}\rfloor}$, alors la condition (ii) est certainement vérifiée. Par suite, il suffit d'assurer l'existence d'un nombre premier r tel que $r-1$ ait un facteur premier q plus grand que $4\sqrt{r} \log N / \log 2$.

Notant $P(n)$ le plus grand facteur premier de n , il faut donc être capable d'estimer le nombre de nombres premiers r ayant une grande valeur de $P(r-1)$. Intéressons-nous à la quantité

$$\mathcal{P}_\delta(x) = \#\{p \text{ premier} \leq x, P(p-1) > x^\delta\}$$

pour $\delta > 0$. Nous nous intéressons ici au cas où il y a suffisamment de nombres premiers dans l'ensemble, c'est-à-dire au cas où $\mathcal{P}_\delta(x) \geq c_\delta \pi(x)$ pour x assez grand. Suite aux travaux de Fouvry [31] ainsi que de Baker et Harman [11], on sait que la plus grande valeur de δ pour laquelle $\mathcal{P}_\delta(x) \geq c_\delta \pi(x)$ est $\delta = 0.676$. Pour notre propos, toute valeur de $\delta \geq 2/3$ ou même $> 1/2$ (en utilisant [34] comme suggéré par Pomerance) suffirait.

Nous aurons également besoin de majorations effectives de $\pi(x)$. Celles de [66] donnent : $x/\log x < \pi(x) \leq \gamma x/\log x$ pour $x \geq 114$.

Nous sommes maintenant prêts à montrer :

THÉORÈME 5.7. — Soient $\delta \in]1/2, 0.676]$ et $\alpha = 2/(2\delta - 1)$. Il existe deux constantes c_1 et $c_2 > c_1$ telles qu'il existe un nombre premier r dans l'intervalle $I_\alpha = [c_1(\log N)^\alpha, c_2(\log N)^\alpha]$ tel que $r - 1$ ait un facteur premier $q \geq 4\sqrt{r} \log N / \log 2$ et pour lequel l'ordre de N modulo r soit divisible par q .

Démonstration. — On commence par remarquer que si $r \in I_\alpha$, on a $r^\delta \geq 4\sqrt{r} \log N / \log 2$, les exposants des $\log N$ étant égaux, à condition de prendre $c_1^\delta \geq 4c_2^{1/2} / \log 2$ et aussi $c_2 > (4/\log 2)^{2/(2\delta-1)} = (4/\log 2)^\alpha$ (puisque c_2 doit être plus grand que c_1).

Appelons *convenables* les nombres premiers cherchés et comptons leur nombre $\mathcal{R}(N)$. Nous noterons pour alléger $L = \log N$. Ce nombre est :

$$\begin{aligned} \mathcal{R}(N) &= \mathcal{P}_\delta(c_2 L^\alpha) - \mathcal{P}_\delta(c_1 L^\alpha) \\ &\geq \mathcal{P}_\delta(c_2 L^\alpha) - \pi(c_1 L^\alpha) \\ &\geq c_\delta \frac{c_2 L^\alpha}{\log c_2 + \alpha \log L} - \pi(c_1 L^\alpha) \\ &\geq \frac{L^\alpha}{(\log c_2 + \alpha \log L)(\log c_1 + \alpha \log L)} (c_\delta c_2 \log c_1 - \gamma c_1 \log c_2 + (c_\delta c_2 - \gamma c_1) \alpha \log L). \end{aligned}$$

Il suffit de trouver c_1 tel que $c_1 < c_\delta c_2 / \gamma$. N'oublions pas que $c_1 \geq (4c_2^{1/2})^{1/\delta}$. Ces deux relations sont compatibles à partir du moment où $c_2^{1-1/(2\delta)} \geq 4^{1/\delta} \gamma / c_\delta$, ce qui est faisable quand $1 - 1/(2\delta) > 0$. Par suite

$$\mathcal{R}(N) \geq c_3 \frac{(\log N)^\alpha}{\log \log N}$$

pour une certaine constante $c_3 > 0$ et N assez grand.

On cherche maintenant un r de l'intervalle tel que $r - 1$ ait un facteur premier q vérifiant $N^{(r-1)/q} \not\equiv 1 \pmod{r}$. Pour cela, on considère le produit :

$$\Pi = (N - 1)(N^2 - 1) \cdots (N^v - 1).$$

Il suffit de trouver r convenable et $q \mid r - 1$ tels que $(r - 1)/q \leq v$ et r ne divise pas Π . Or le nombre de diviseurs premiers de Π est au plus $\log N^{v(v+1)/2} / \log 2 \leq v^2 \log N$. Si $v^2 \log N < \mathcal{R}(N)$, alors on est sûr de l'existence d'un r convenable ne divisant pas Π . Il suffit de prendre $v = c_4(\log N)^{(\alpha-2)/2}$ pour cela.

La condition $(r - 1)/q \leq v$ sera satisfaite si

$$c_2^{1-\delta} (\log N)^{(1-\delta)\alpha} \leq c_4 (\log N)^{(\alpha-2)/2}$$

c'est-à-dire $c_4 \geq c_2^{1-\delta}$. □

On a donc montré, en prenant $r = O((\log N)^\alpha)$, $s = O((\log N)^{\alpha/2+1})$, que :

THÉORÈME 5.8. — Pour tout $\delta \in]1/2, 0.676]$, il existe un algorithme de primalité déterministe dont le temps de calcul est $O((\log N)^{(8\delta+1)/(2\delta-1)})$ si on utilise de l'arithmétique naïve, et $\tilde{O}((\log N)^{6\delta/(2\delta-1)})$ si on utilise les FFT.

COROLLAIRE 5.9. — $\text{estPremier?} \in P$.

Si l'on prend $\delta = 2/3$, on obtient les exposants 19 et 12. Remarquons que si on pouvait faire tendre δ vers 1, on obtiendrait au mieux $\tilde{O}((\log N)^6)$. Notons qu'une façon d'obtenir $\delta = 1$ serait de prouver l'existence d'un nombre suffisant de nombres premiers de Sophie Germain, i.e. r tel que $(r - 1)/2$ soit également premier.

5.4. L'après AKS

De multiples améliorations de toutes natures ont fleuri depuis la parution de l'article, concernant des généralisations et améliorations de l'algorithme de base. D. Bernstein maintient une page résumant les différents travaux [14]. La situation est loin d'être stabilisée à la date où je tape ces lignes.

Les résultats basés sur les théorèmes de Fouvry ou Baker/Harman ne sont pas effectifs, ce qui fait qu'on ne sait pas à partir de quelle taille de N les estimations s'appliquent. De plus, les contraintes imposées au nombre premier r auxiliaire sont très fortes.

H.W. Lenstra, Jr., (communication personnelle, cf. également [14]) a amélioré l'algorithme au point de prouver un temps de calcul effectif en $\tilde{O}((\log N)^{12})$ ou bien $\tilde{O}((\log N)^8)$ en utilisant de nouveau le théorème de Fouvry. Des résultats similaires ont été trouvés par S. David (communication personnelle). Rapidement, l'amélioration essentielle tient à une relaxation des contraintes sur r , qui peut être maintenant un entier quelconque. Une condition combinatoire ressemblant à (3) est utilisée en ses lieu et place.

D'autres améliorations ont été apportées par d'autres chercheurs, comme la lecture de [14] en fait foi.

5.5. Vers un algorithme pratique

L'algorithme AKS n'est pas un algorithme efficace dans sa version de base, essentiellement parce que les degrés des polynômes en jeu sont bien trop grands. En effet, au minimum, r doit être de taille $O((\log N)^2)$ pour satisfaire les conditions du théorème. On constate de plus que la complexité prouvée est pire que celle des sommes de Jacobi pour les nombres de taille raisonnable, ou bien celle d'ECPP, même heuristique. Les récentes avancées semblent augurer de l'existence d'une version randomisée d'AKS qui sera peut-être plus rapide dans la pratique.

Signalons pour commencer que l'article [6] contient une conjecture intrigante, tant elle paraît simple.

CONJECTURE 5.10. — *Soit r un entier tel que $r - N^2 - 1$. Si*

$$(5) \quad (X + 1)^N \equiv X^N + 1 \pmod{(N, X^r + 1)}$$

alors N est premier.

Si cela était vrai, alors on disposerait d'un algorithme déterministe en $\tilde{O}((\log N)^3)$. Cette conjecture a été testée numériquement pour des nombres plus petits que 10^{10} (cf. [40]). Nous extrayons de ce travail le résultat suivant. Supposons que $N = 10k + 3$. Alors la relation (5) est équivalente aux deux équations :

$$5^{(N-1)/2} \equiv -1 \pmod{N}, \left(\frac{\theta+1}{\theta-1}\right)^{(N+1)/2} \equiv -1 \pmod{N}$$

où $\theta^2 - 5 = 0$. Ces deux relations demandent donc à N d'être pseudopremier d'Euler en base 5 et pseudopremier de Lucas pour θ . Cela n'est pas sans rappeler les combinaisons de tests proposés dans [61] (voir aussi [29, Ex. 3.41, p. 156]) et pour lesquels aucun contre-exemple n'est connu à ce jour.

P. Berrizbeitia [16], suivi de Q. Cheng [23], a proposé une amélioration dans le cas où $N - 1$ admet un facteur premier r pour lequel $r^\alpha \parallel N - 1$, $r \geq \log^2 N$. On provoque une explosion combinatoire à l'aide des racines r -ièmes de l'unité dans $\mathbf{Z}/N\mathbf{Z}$, ce qui donne une complexité en $\tilde{O}((\log N)^4)$ pour l'algorithme qui est randomisé.

Suivant en cela un schéma classique, passer de $N - 1$, puis à $N + 1$ (comme Berrizbeitia), il ne reste plus qu'à généraliser au cas où il existe deux entiers e et d tels que $e \mid N^d - 1$, ce qu'a fait Bernstein [15]. La complexité tombe encore à $\tilde{O}((\log N)^4)$ pour une version randomisée, cette fois pour tous les nombres. Dans la suite logique, P. Mihăilescu (communications personnelles) a annoncé très récemment un algorithme combinant AKS et cyclotomie.

6. CONCLUSIONS

Nous avons tenté de faire un tour d'horizon du problème de la primalité, en mêlant considérations théoriques (complexité, mathématiques), et algorithmiques, voire pratiques. Maintenant que `estPremier?` est montré être dans \mathbf{P} , l'histoire s'arrête-t-elle là ? Sans doute pas. Tout d'abord, la chasse au meilleur exposant ne fait que commencer. En outre, une version pratique de AKS n'est encore que frémissante. L'avenir dira quand elle détrônera les méthodes ancestrales.

Remerciements. Je remercie G. Hanrot pour sa relecture précise et ses nombreuses suggestions, P. Gaudry pour sa relecture et de nombreuses discussions sur le sujet. Merci à tous ceux avec qui j'ai eu des discussions sur certains points des différents algorithmes : H. Cohen, S. David, H. W. Lenstra, Jr., P. Mihăilescu, J.-L. Nicolas, C. Pomerance, O. Ramaré, J. Rivat.

RÉFÉRENCES

- [1] L. M. Adleman. On distinguishing prime numbers from composite numbers. In *Foundations of computer science*, pages 387–406. IEEE, 1980. 21st FOCS, Syracuse, USA, Proceedings.
- [2] L. M. Adleman and M.-D. A. Huang. *Primality testing and Abelian varieties over finite fields*, volume 1512 of *Lecture Notes in Math*. Springer-Verlag, 1992.
- [3] L. M. Adleman, C. Pomerance, and R. S. Rumely. On distinguishing prime numbers from composite numbers. *Ann. of Math. (2)*, 117:173–206, 1983.
- [4] L. M. Adleman, R. L. Rivest, and A. Shamir. A method for obtaining digital signatures and public-key cryptosystems. *Comm. ACM*, 21(2):120–126, 1978.
- [5] M. Agrawal and S. Biswas. Primality and identity testing via Chinese remaindering. In *Proc. Ann. IEEE Symp. Found. Comp. Sci.*, pages 202–209, 1999.
- [6] M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in P. Preprint; available at <http://www.cse.iitk.ac.in/primality.pdf>, August 2002.
- [7] W. R. Alford, A. Granville, and C. Pomerance. There are infinitely many Carmichael numbers. *Ann. of Math. (2)*, 139(3):703–722, May 1994.
- [8] M. Artjuhov. Certain criteria for the primality of numbers connected with the little Fermat theorem (Russian). *Acta Arith.*, 12:355–364, 1966/67.
- [9] A. O. L. Atkin and F. Morain. Elliptic curves and primality proving. *Math. Comp.*, 61(203):29–68, July 1993.
- [10] E. Bach. Explicit bounds for primality testing and related problems. *Math. Comp.*, 55(191):355–380, July 1990.
- [11] R. C. Baker and G. Harman. The Brun-Titchmarsh theorem on average. In *Proceedings of a conference in Honor of Heini Halberstam*, volume 1, pages 39–103, 1996.
- [12] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc. (3)*, 83(3):532–562, 2001.
- [13] D. Bernstein. An exposition of the Agrawal-Kayal-Saxena primality-proving theorem. Preprint, August 2002.
- [14] D. Bernstein. Proving primality after Agrawal-Kayal-Saxena.
<http://cr.yp.to/papers/aks.ps>, January 2003.
- [15] D. Bernstein. Proving primality in essentially quartic expected time.
<http://cr.yp.to/papers/quartic.ps>, January 2003.
- [16] P. Berrizbeitia. Sharpening “Primes is in P” for a large family of numbers.
<http://arxiv.org/abs/math.NT/0211334>, November 2002.

- [17] I. Blake, G. Seroussi and N. Smart. *Elliptic curves in cryptography*, volume 265 of *London Math. Soc. Lecture Note Ser.* Cambridge University Press, 1999.
- [18] W. Bosma and M.-P. van der Hulst. *Primality proving with cyclotomy*. PhD thesis, Universiteit van Amsterdam, December 1990.
- [19] J. Brillhart, D. H. Lehmer, and J. L. Selfridge. New primality criteria and factorizations of $2^m \pm 1$. *Math. Comp.*, 29(130):620–647, 1975.
- [20] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman and S. S. Wagstaff, Jr. *Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ up to high powers*. Number 22 in *Contemporary Mathematics*. AMS, 2nd edition, 1988.
- [21] O. Caprotti and M. Oostdijk. Formal and efficient primality proofs by use of computer algebra oracles. *J. Symbolic Comput.*, 32:55–70, 2001.
- [22] H. F. Chau and H.-K. Lo. Primality test via quantum factorization. *Internat. J. Modern Phys. C*, 8(2):131–138, 1997.
- [23] Q. Cheng. Primality proving via one round in ECPP and one iteration in AKS. www.cs.ou.edu/~qcheng/paper/aksimp.pdf, January 2003.
- [24] H. Cohen. *A course in algorithmic algebraic number theory*, volume 138 of *Graduate Texts in Mathematics*. Springer–Verlag, 1996. Third printing.
- [25] H. Cohen and A. K. Lenstra. Implementation of a new primality test. *Math. Comp.*, 48(177):103–121, 1987.
- [26] H. Cohen and H. W. Lenstra, Jr. Primality testing and Jacobi sums. *Math. Comp.*, 42(165):297–330, 1984.
- [27] L. Comtet. *Analyse combinatoire*. Presses Universitaires de France, 1970.
- [28] D. A. Cox. *Primes of the form $x^2 + ny^2$* . John Wiley & Sons, 1989.
- [29] R. Crandall and C. Pomerance. *Prime numbers – A Computational Perspective*. Springer Verlag, 2000.
- [30] N. D. Elkies. Elliptic and modular curves over finite fields and related computational issues. In D. A. Buell and J. T. Teitelbaum, editors, *Computational Perspectives on Number Theory: Proceedings of a Conference in Honor of A.O.L. Atkin*, volume 7 of *AMS/IP Studies in Advanced Mathematics*, pages 21–76. American Mathematical Society, International Press, 1998.
- [31] E. Fouvry. Théorème de Brun-Titschmarsh; application au théorème de Fermat. *Invent. Math.*, 79:383–407, 1985.
- [32] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 1999.
- [33] P. Gaudry and R. Harley. Counting points on hyperelliptic curves over finite fields. In W. Bosma, editor, *Algorithmic Number Theory*, volume 1838 of *Lecture Notes in*

- Comput. Sci.*, pages 313–332. Springer Verlag, 2000. 4th International Symposium, ANTS-IV, Leiden, The Netherlands, July 2000, Proceedings.
- [34] M. Goldfeld. On the number of primes p for which $p + a$ has a large prime factor. *Mathematika*, 16:23–27, 1969.
- [35] S. Goldwasser and J. Kilian. Almost all primes can be quickly certified. In *Proc. 18th STOC*, pages 316–329. ACM, 1986. May 28–30, Berkeley.
- [36] S. Goldwasser and J. Kilian. Primality testing using elliptic curves. *Journal of the ACM*, 46(4):450–472, July 1999.
- [37] D. R. Heath-Brown. The differences between consecutive primes. *J. London Math. Soc. (2)*, 18(1):7–13, 1978.
- [38] K. Ireland and M. Rosen. *A classical introduction to modern number theory*, volume 84 of *Graduate Texts in Mathematics*. Springer, 1982.
- [39] H. Iwaniec and M. Jutila. Primes in short intervals. *Ark. Mat.*, 17(1):167–176, 1979.
- [40] N. Kayal and N. Saxena. Towards a deterministic polynomial-time primality test. Technical report, IIT Kanpur, 2002.
<http://www.cse.iitk.ac.in/research/btp2002/primality.html>.
- [41] D. E. Knuth. *The Art of Computer Programming: Seminumerical Algorithms*. Addison-Wesley, 2nd edition, 1981.
- [42] H. Lange and W. Ruppert. Complete systems of addition laws on abelian variety. *Invent. Math.*, 79:603–610, 1985.
- [43] D. H. Lehmer. Strong Carmichael numbers. *J. Austral. Math. Soc. Ser. A*, 21:508–510, 1976.
- [44] A. K. Lenstra and H. W. Lenstra, Jr. Algorithms in number theory. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A: Algorithms and Complexity, chapter 12, pages 674–715. North Holland, 1990.
- [45] A. K. Lenstra and H. W. Lenstra, Jr., editors. *The development of the number field sieve*, volume 1554 of *Lecture Notes in Math.* Springer, 1993.
- [46] H. W. Lenstra, Jr. Miller’s primality test. *Inform. Process. Lett.*, 8(2):86–88, 1979.
- [47] H. W. Lenstra, Jr. Primality testing algorithms (after Adleman, Rumely, Williams). *Sém. Bourbaki*, 1980-1981, exp. n° 576, *Lecture Notes in Math.*, Springer, 901:243-257, 1981.
- [48] H. W. Lenstra, Jr. Primality testing. In *Computational methods in number theory, Part I*, pages 55–77. Math. Centrum, Amsterdam, 1982.
- [49] H. W. Lenstra, Jr. Primality testing with Artin symbols. In *Number theory related to Fermat’s last theorem (Cambridge, Mass., 1981)*, volume 26 of *Progr. Math.*, pages 341–347. Birkhäuser Boston, Mass., 1982.

- [50] H. W. Lenstra, Jr. Galois theory and primality testing. In I. Reiner and K. W. Roggenkamp, editors, *Orders and their applications*, volume 1142 of *Lecture Notes in Math.*, pages 169–189. Springer, 1984. Proc. of a conference, Oberwolfach, June 3–9, 1984.
- [51] H. W. Lenstra, Jr. Elliptic curves and number-theoretic algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, pages 99–120, Providence, RI, 1987. Amer. Math. Soc.
- [52] H. W. Lenstra, Jr. Factoring integers with elliptic curves. *Ann. of Math. (2)*, 126:649–673, 1987.
- [53] A. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of applied cryptography*. CRC Press, 1997.
- [54] P. Mihăilescu. *Cyclotomy of rings and primality testing*. Diss. ETH No. 12278, Swiss Federal Institute of Technology Zürich, 1997.
- [55] G. L. Miller. Riemann’s hypothesis and tests for primality. In *Proc. 7th STOC*, pages 234–239, 1975.
- [56] F. Morain. Calcul du nombre de points sur une courbe elliptique dans un corps fini : aspects algorithmiques. *J. Théor. Nombres Bordeaux*, 7:255–282, 1995.
- [57] F. Morain. Implementing the asymptotically fast version of the elliptic curve primality proving algorithm. En préparation, Juin 2003.
- [58] C. H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1995.
- [59] C. Pomerance. Analysis and comparison of some integer factoring algorithms. In H. W. Lenstra, Jr. and R. Tijdeman, editors, *Computational methods in number theory*, pages 89–140. Mathematisch Centrum, Amsterdam, 1982. Mathematical Center Tracts 154/155.
- [60] C. Pomerance. Very short primality proofs. *Math. Comp.*, 48(177):315–322, 1987.
- [61] C. Pomerance, J. L. Selfridge, and Samuel S. Wagstaff, Jr. The pseudoprimes to $25 \cdot 10^9$. *Math. Comp.*, 35(151):1003–1026, 1980.
- [62] V. R. Pratt. Every prime has a succinct certificate. *SIAM J. Comput.*, 4:214–220, 1975.
- [63] M. Rabin. Probabilistic algorithms for testing primality. *J. Number Theory*, 12:128–138, 1980.
- [64] J. Radhakrishnan. Primes in P. *Bull. of the EATCS*, 78:61–65, October 2002.
- [65] P. Ribenboim. *The new book of prime number records*. Springer-Verlag, 1996.
- [66] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.*, 6:64–94, 1962.

- [67] R. Schoof. Elliptic curves over finite fields and the computation of square roots mod p . *Math. Comp.*, 44:483–494, 1985.
- [68] R. Schoof. Counting points on elliptic curves over finite fields. *J. Théor. Nombres Bordeaux*, 7:219–254, 1995.
- [69] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 26th Annual ACM Symposium on Theory of Computing (STOC)*, pages 124–134, Montreal, Canada, 1994. ACM.
- [70] R. Solovay and V. Strassen. A fast Monte-Carlo test for primality. *SIAM J. Comput.*, 6(1):84–85, 1977. Erratum, *ibid.*, volume 7, 1, 1978.
- [71] P. van Wamelen. Jacobi sums over finite fields. *Acta Arith.*, 102.1:1–20, 2002.
- [72] E. Waterhouse. Abelian varieties over finite fields. *Ann. Sci. Éc. Norm. Sup.*, 2:521–560, 1969.
- [73] A. Weng. Constructing hyperelliptic curves of genus 2 suitable for cryptography. *Math. Comp.*, 72:435–458, 2003.
- [74] H. C. Williams and J. O. Shallit. Factoring integers before computers. In *Mathematics of Computation 1943–1993: a half-century of computational mathematics (Vancouver, BC, 1993)*, volume 48 of *Proc. Sympos. Appl. Math.*, pages 481–531. Amer. Math. Soc., Providence, RI, 1994.

François MORAIN

Laboratoire d'Informatique

École polytechnique (LIX)

CNRS/UMR 7650, INRIA-Futurs

F-91128 Palaiseau Cedex

E-mail : morain@lix.polytechnique.fr

SYSTÈMES HYPERBOLIQUES ET VISCOSITÉ ÉVANESCENTE

[d'après S. Bianchini et A. Bressan]

par Frédéric ROUSSET

INTRODUCTION

De nombreux phénomènes physiques provenant de domaines variés comme la mécanique des fluides, l'élasticité ou l'électromagnétisme peuvent être décrits par des systèmes de lois de conservation

$$(1) \quad u_t + f(u)_x = 0,$$

où $u(t, x)$ est un vecteur de \mathbb{R}^n , $t > 0$ est la variable de temps, $x \in \mathbb{R}$ est la variable d'espace et le flux $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est régulier. Dans tout cet exposé, on notera les dérivées partielles par des indices : $u_t = \partial u / \partial t$, $f(u)_x = \partial f(u) / \partial x$. On considère le problème de Cauchy, où le système (1) est complété par une donnée initiale

$$(2) \quad u(0, x) = u_0(x).$$

Il est bien connu (même sur l'exemple très simple de l'équation de Burgers $u_t + (u^2/2)_x = 0$, voir par exemple [11], [21], [32], [40]) que pour des données initiales u_0 très régulières les solutions de (1) peuvent développer des discontinuités en temps fini. Il faut donc chercher les solutions de (1) dans un espace de fonctions discontinues et comprendre (1) au sens faible

$$\iint (u\phi_t + f(u)\phi_x) dx dt = 0,$$

pour toute fonction ϕ régulière à support compact dans $]0, T[\times \mathbb{R}$. Cependant ces solutions faibles ne sont en général pas uniques (cela peut encore se vérifier sur l'équation de Burgers). Il faut donc se donner des critères permettant d'isoler les solutions «physiquement» admissibles. Il existe de nombreux critères (conditions d'entropie, de Lax, de Liu, du profil de viscosité), voir par exemple le livre de C. Dafermos [21] pour une comparaison entre ces différents critères. Le moyen le plus naturel pour trouver certains d'entre eux est de considérer que les lois de conservation (1) sont obtenues à partir de systèmes plus complexes

$$(3) \quad u_t^\varepsilon + f(u^\varepsilon)_x = \varepsilon u_{xx}^\varepsilon, \quad \varepsilon > 0$$

en négligeant les effets de la diffusion. Une conjecture naturelle est donc que les solutions admissibles de (1) doivent être obtenues comme limites de solutions de (3) quand ε tend vers zéro.

Sur le plan mathématique cette brève introduction soulève deux problèmes importants :

1. Montrer que le problème de Cauchy pour (1) est bien posé (c'est-à-dire qu'il existe une unique solution admissible dépendant continûment des données),
2. Montrer que ces solutions peuvent effectivement s'obtenir comme limite de solutions de (3) lorsque ε tend vers zéro.

Avant de poursuivre, signalons que, dans le cas scalaire ($n = 1$), ces deux questions ont été résolues il y a longtemps dans un cadre L^∞ grâce au principe du maximum [30]. La méthode de compacité par compensation basée sur des estimations L^∞ a aussi permis de montrer pour certains systèmes 2×2 la convergence faible de u^ε vers une solution de (1) [23]. Enfin, pour des systèmes généraux, des méthodes de type «perturbations singulières» [26], [44] avaient permis de montrer que certaines solutions de (1) (régulières par morceaux avec des chocs sans interactions) étaient limites de solutions de (3).

Le but de cet exposé est de présenter les résultats obtenus dans la résolution du problème 2 par S. Bianchini et A. Bressan [7]. Néanmoins, les deux problèmes n'étant pas indépendants, nous rappelons d'abord les résultats obtenus dans la résolution du problème 1 par Bressan et divers collaborateurs pour des fonctions à variations bornées (notées VB).

1. LE PROBLÈME DE CAUCHY POUR (1)

Commençons par fixer le cadre de la théorie ; on considère des systèmes strictement hyperboliques

(SH) $df(u)$ est diagonalisable à valeurs propres réelles simples.

Ces valeurs propres seront notées $\lambda_1(u) < \dots < \lambda_n(u)$, les vecteurs propres correspondants $r_i(u)$.

Pour des données initiales u_0 à variation totale suffisamment petite, l'existence globale de solutions entropiques avait été montrée par Glimm dans les années 60 [24]. Une autre méthode de construction dite de «front tracking» a par la suite été proposée [20, 23, 8, 38, 1]. Sans entrer dans les détails, ces deux méthodes de construction reposent sur la résolution du problème de Riemann au moyen de n ondes simples due à Lax [31], *i.e.* la résolution de (1), (2) lorsque u_0 est constante par morceaux, de la forme

$$(4) \quad u_0(x) = \begin{cases} u_-, & x < 0, \\ u_+, & x > 0. \end{cases}$$

Cela demande une hypothèse supplémentaire que nous ferons tout au long de ce paragraphe :

(\diamond) les champs caractéristiques sont vraiment non linéaires $d\lambda_i(u)r_i(u) \neq 0, \forall u$ ou linéairement dégénérés $d\lambda_i(u)r_i(u) = 0, \forall u$.

Dans les deux méthodes, l'existence de solutions est obtenue par un argument de compacité («l'injection» de $VB(\mathbb{R})$ dans $L^1_{loc}(\mathbb{R})$ est compacte), ce qui ne garantit pas du tout l'unicité. Cette question a été résolue dans les dix dernières années par Bressan et ses collaborateurs [9, 13, 14, 18, 16, 15, 17]. On peut résumer les résultats obtenus comme suit :

- Les solutions obtenues comme limite d'approximation par les schémas de Glimm ou de «front tracking» sont uniques et dépendent de manière lipschitzienne du temps et de la donnée initiale [35], [36] [18]. Plus précisément, la méthode de front-tracking permet de construire de manière unique un semi-groupe lipschitzien $(t, \bar{u}) \mapsto S_t \bar{u}$ défini pour $t \geq 0$ et $\bar{u} \in \mathcal{D} \subset L^1$, qui résout le problème de Cauchy pour (1) et dont on peut en fait caractériser les trajectoires. Soit $u \in \text{Lip}([0, T], L^1)$ telle que $u(t, \cdot) \in \mathcal{D}$; alors u est une trajectoire du semi-groupe si et seulement si, au voisinage de chaque point (τ, ξ) , pour des temps positifs, u peut être bien approchée par une solution du problème de Riemann et par une solution d'un problème linéaire à coefficients constants là où la variation totale est petite. Pour énoncer un théorème précis, notons $u^- = \lim_{x \rightarrow \xi^-} u(\tau, x)$, $u^+ = \lim_{x \rightarrow \xi^+} u(\tau, x)$ et $U^1_{(u, \tau, \xi)}$ la solution du problème de Riemann avec donnée (u^-, u^+) . On définit aussi $U^2_{(u, \tau, \xi)}$ la solution du problème de Cauchy linéaire à coefficients constants à donnée en $t = \tau$

$$w_t + df(u(\tau, \xi))w_x = 0, w(\tau, x) = u(\tau, x) ;$$

on a alors :

THÉORÈME 1.1 ([9], [11]). — Soit $u \in \text{Lip}([0, T], L^1)$ telle que $u(t, \cdot) \in \mathcal{D}$ pour tout t ; alors, il existe $\bar{u} \in \mathcal{D}$ tel que $u = S_t \bar{u}$ si et seulement si

i) $\forall (\tau, \xi), \beta > 0,$

$$\lim_{h \rightarrow 0^+} \frac{1}{h} \int_{\xi-h, \beta}^{\xi+h, \beta} |u(\tau+h, x) - U^1_{(u, \tau, \xi)}(h, x - \xi)| dx = 0,$$

ii) $\exists C, \beta > 0, \forall a < \xi < b, \forall \tau \geq 0,$

$$\limsup_{h \rightarrow 0^+} \frac{1}{h} \int_{a+\beta h}^{b-\beta h} |u(\tau+h, x) - U^2_{(u, \tau, \xi)}(h, x)| dx \leq C \text{VT}(u(\tau)_{/[a, b]}).$$

Dans tout cet exposé, on notera VT la variation totale d'une fonction.

- Le fait surprenant du théorème précédent est que toute fonction u vérifiant i) et ii) sera en fait une solution de (1). Un autre résultat d'unicité consiste en une caractérisation directe des solutions entropiques de (1) qui satisfont une condition supplémentaire [16], [17], [15]. Par exemple, on dit que u satisfait la condition d'oscillation douce («tame oscillation») s'il existe $C > 0, \delta > 0$ tels que pour tout

$t > 0$ et $a, b, a < b$ l'oscillation $\omega(u, \Delta) = \sup_{(s,y), (s',y') \in \Delta} \{|u(s, y) - u(s', y')|\}$ de u sur le triangle

$$\Delta = \{(s, y), s \geq t, a + (s - t)/\delta < y < b - (s - t)/\delta\}$$

vérifie :

$$\omega(u, \Delta) \leq C \text{ VT}(u(\tau)_{/[a,b]}).$$

THÉORÈME 1.2 ([15]). — *Toute solution faible de (1), (2) qui satisfait la condition de choc de Lax et la condition d'oscillation douce coïncide avec une trajectoire du semi-groupe construit par le schéma de front-tracking.*

Enfin, on peut montrer que le fait de travailler dans VB et les hypothèses « donnée initiale à variation totale petite » et « système strictement hyperbolique » (SH) sont cruciales. Il existe des contre-exemples (voir par exemple [10]) montrant :

- (1) si le système n'est pas strictement hyperbolique, les solutions peuvent ne pas dépendre continûment de la donnée initiale en norme L^1 ,
- (2) si la donnée initiale est seulement dans L^∞ , la solution peut devenir à valeurs mesurées en temps fini. L'unicité et la dépendance continue par rapport aux données sont alors perdues,
- (3) si la variation totale à $t = 0$ est grande, alors la norme L^∞ peut exploser en temps fini. En particulier, il n'existe pas de solution globale dans VB [29].

2. LES RÉSULTATS DE BIANCHINI ET BRESSAN

THÉORÈME 2.1 ([7]). — *Considérons le problème de Cauchy*

$$(5) \quad u_t^\varepsilon + A(u^\varepsilon)u_x^\varepsilon = \varepsilon u_{xx}^\varepsilon, \quad u^\varepsilon(0, x) = \bar{u}(x),$$

où A est régulière et strictement hyperbolique (i.e. vérifie (SH)) alors il existe des constantes C, L, L' et $\delta > 0$ telles que si $\text{VT}(\bar{u}) < \delta$ alors pour tout $\varepsilon > 0$, (5) a une unique solution u^ε définie pour tout $t \geq 0$. En notant $S_t^\varepsilon \bar{u}$ cette solution, on a :

- une **borne uniforme sur la variation totale**

$$(6) \quad \text{VT}(S_t^\varepsilon \bar{u}) \leq C \text{ VT}(\bar{u}),$$

- des **estimations de stabilité**

$$(7) \quad \|S_t^\varepsilon \bar{u} - S_t^\varepsilon \bar{v}\|_{L^1} \leq L \|\bar{u} - \bar{v}\|_{L^1},$$

$$(8) \quad \|S_t^\varepsilon \bar{u} - S_s^\varepsilon \bar{u}\|_{L^1} \leq L'(|t - s| + |\sqrt{\varepsilon t} - \sqrt{\varepsilon s}|),$$

- la **convergence** : quand ε tend vers zéro, u^ε converge vers une trajectoire d'un semi-groupe S tel que

$$(9) \quad \|S_t \bar{u} - S_s \bar{v}\|_{L^1} \leq L \|\bar{u} - \bar{v}\| + L'|t - s|.$$

Appelons les trajectoires de S « solutions de viscosité évanescence » pour le système hyperbolique $u_t + A(u)u_x = 0$.

Dans le cas conservatif $A = df$, toute solution de viscosité évanescence est une solution faible de (1) qui satisfait la condition d'admissibilité de Liu. En supposant de plus (\diamond) , les solutions de viscosité évanescence coïncident avec les limites uniques des schémas de Glimm et de front tracking.

Ce théorème provenant de [7] est l'aboutissement de toute une série de travaux [3, 6, 4, 5] ; il contient en fait plusieurs résultats, chacun étant intéressant par lui-même. L'idée directrice de ces travaux était l'obtention d'estimations VB permettant d'obtenir la convergence sans hypothèse très particulière sur la structure de la solution de (1) qu'il est nécessaire de faire pour utiliser les méthodes de perturbation singulière [26], [44].

Le théorème 2.1 montre, pour des données initiales quelconques à variation totale petite, l'existence d'une solution globale pour le problème de Cauchy (5) vérifiant les trois estimations (6), (7), (8). Cela généralise considérablement des travaux antérieurs où la stabilité était obtenue au voisinage de solutions particulières (chocs ou détentes) à variations totales petites [25, 26, 34, 41, 42, 44]. Dans certains cas particuliers, d'autres méthodes permettent de supprimer l'hypothèse de petitesse : pour des perturbations régulières d'un choc de forte amplitude, on a des résultats de stabilité et d'approximation par viscosité sous hypothèse spectrale [46, 39] même en plusieurs dimensions d'espace [45, 27]. Nous donnerons dans la suite des éléments de preuve de cette partie du théorème.

En ce qui concerne la partie convergence, on peut facilement montrer à partir du théorème de Helly (qui affirme que l'injection de $BV(\mathbb{R})$ dans $L^1_{loc}(\mathbb{R})$ est compacte) et des estimations (6) et (7), (8) qu'il existe une sous-suite ε_n telle que $S_t^{\varepsilon_n} \bar{u}$ converge dans $L^1_{loc}(\mathbb{R})$ pour tout $t \geq 0$ et tout \bar{u} tel que $VT(\bar{u}) \leq \delta$. Ce même argument de compacité était déjà utilisé dans les méthodes de Glimm et de front-tracking (voir par exemple le livre de Bressan [11]). On remarque que dans le cas conservatif $A = df$, cela donne un procédé de construction de solutions faibles pour (1), (2) qui n'utilise pas l'hypothèse (\diamond) . Cela généralise une série de résultats [33, 37, 28] où l'hypothèse était relaxée mais pas complètement supprimée. La convergence reposant sur un argument de compacité, le problème est ensuite de montrer l'unicité des solutions obtenues. Dans le cas conservatif $A = df$, les solutions de viscosité évanescence sont des solutions faibles de (1). Avec l'hypothèse (\diamond) , il suffit alors de montrer que les trajectoires des semi-groupes obtenus comme limites de S^{ε_n} possèdent la liste de propriétés permettant d'utiliser le théorème 1.2. La méthode consiste à montrer que les solutions de (5) vérifient « presque » ces propriétés lorsque $\varepsilon > 0$ est petit. Par exemple, on montre que si deux solutions $u^\varepsilon, v^\varepsilon$ de (5) ont des données initiales vérifiant $u_0(x) = v_0(x), \forall x \in [a, b]$, on a l'estimation

$$|u^\varepsilon(t, x) - v^\varepsilon(t, x)| \leq \alpha \|u_0 - v_0\|_{L^\infty} \left(e^{\frac{\beta t - (x-a)}{\varepsilon}} + e^{\frac{\beta t + (x-b)}{\varepsilon}} \right).$$

Ainsi, en passant à la limite quand ε_n tend vers zéro, on retrouve la propriété classique de propagation à vitesse finie pour les solutions de (1) :

$$u(t, x) = v(t, x), \text{ si } a + \beta t < x < b - \beta t.$$

On utilise la même idée pour montrer la condition d'oscillation douce.

Le cas général non conservatif est beaucoup plus délicat. Les auteurs construisent d'abord pour une donnée initiale de Riemann (*i.e.* de la forme (4)) une solution auto-similaire du problème de Cauchy non conservatif pour le système $u_t + A(u)u_x = 0$ qui est l'unique solution pouvant s'obtenir comme limite de solutions u^ε de (5). Cette construction utilise les courbes de la décomposition en ondes progressives que nous décrirons plus loin ; dans l'esprit, elle est donc proche du procédé proposé dans [22]. Cela permet de définir une notion de solutions de viscosité en termes d'intégrales locales (u est solution de viscosité si u vérifie i) et ii) du théorème 1.1) comme dans le cas conservatif. Il s'agit ensuite de montrer par une modification de la preuve du théorème 1.1 que ces solutions de viscosité sont uniques et coïncident avec les trajectoires de n'importe quel semi-groupe $S = \lim S^{\varepsilon_n}$ obtenu par viscosité évanescence. La limite étant indépendante de la suite ε_n , cela montre la convergence vers une unique limite de toute la famille d'approximations visqueuses S^ε .

3. LE PROBLÈME DE CAUCHY POUR (5)

Dans ce paragraphe on s'attache à décrire la construction d'une solution globale pour (5) vérifiant la borne (6). On remarque tout d'abord que si on pose $u^\varepsilon(t, x) = u(\frac{t}{\varepsilon}, \frac{x}{\varepsilon})$, u est une solution de

$$(10) \quad u_t + A(u)u_x = u_{xx}$$

avec donnée initiale $u(0, x) = \bar{u}(\varepsilon x)$. De plus la variation totale est invariante par le changement d'échelle ; ainsi on a toujours

$$VT(\bar{u}(\varepsilon \cdot)) = VT(\bar{u}) < \delta.$$

Ainsi, pour obtenir les estimations (6) et (7), (8), il suffit de considérer le système (10).

Pour présenter les nouvelles techniques de [7], on va donner des éléments de preuve de

THÉORÈME 3.1. — *Il existe des constantes $\delta_0 > 0$ et $\kappa > 0$ telles que si $VT(u_0) \leq \frac{\delta_0}{\kappa}$ alors il existe une unique solution de (10) vérifiant $u(0, x) = u_0(x)$ et $VT(u(t)) \leq \delta_0$ pour tout $t \geq 0$.*

La preuve de ce théorème comporte plusieurs étapes. Tout d'abord, l'équation (10) a une forme classique, c'est une équation parabolique semilinéaire. On dispose donc d'un théorème d'existence et d'unicité locale pour le problème de Cauchy (voir par exemple le livre de D. Serre [40]). Dans ce qui suit, on notera simplement $\|\cdot\|$ la norme $\|\cdot\|_{L^1(\mathbb{R})}$.

Pour montrer que les solutions sont en fait globales, on peut donc faire un raisonnement de type équations différentielles ordinaires. On définit

$$T^* = \sup\{T, \forall t \in [0, T[, \|u_x(t)\| \leq \delta_0\}$$

et on cherche à montrer que $T^* = +\infty$. Grâce au théorème d'existence locale, il suffit de montrer que l'on ne peut jamais avoir $\|u_x\| = \delta_0$.

Pour cela on utilise deux types d'estimations. Tout d'abord, on montre que pour $t \in [0, \hat{t}]$, où $\hat{t} \sim \frac{1}{\delta_0^2}$ on a l'estimation

$$\|u_x(t)\| \leq \frac{\delta_0}{2}.$$

En particulier cela implique que $T^* > \hat{t}$. Cette partie utilise des techniques «paraboliques», l'estimation pour le problème non linéaire (10) provient de la stabilité linéaire des constantes.

Il s'agit ensuite de faire une estimation pour $t \geq \hat{t}$ de $\|u_x\|$. Pour cela, des techniques «hyperboliques» sont utilisées : on trouve une décomposition adéquate de u_x sous la forme

$$(11) \quad u_x = \sum_{i=1}^n v_i \tilde{r}_i$$

de telle sorte que

$$(12) \quad \|u_x\| \sim \sum_{i=1}^n \|v_i\|.$$

Les composantes v_i sont alors solutions d'équations de transport diffusion-scalaire

$$(13) \quad v_{it} + (c_i v_i)_x = v_{ixx} + \varphi_i$$

où le terme source φ_i décrit des interactions entre les différentes composantes. Pour cette équation scalaire, on a la propriété classique de contraction dans L^1

$$(14) \quad \|v_i(t)\| \leq \|v_i(\hat{t})\| + \int_{\hat{t}}^t \|\varphi_i(s)\| ds, \forall t \in [\hat{t}, T].$$

Par conséquent, si on trouve une décomposition (11) telle que les termes sources vérifient

$$(15) \quad \forall t, \hat{t} \leq t \leq T, \forall i, \quad \int_{\hat{t}}^T \|\varphi_i\| \leq \frac{\delta_0}{2C},$$

on peut déduire de l'équivalence (12) que l'on garde

$$(16) \quad \|u_x(t)\| \leq \delta_0.$$

On peut donc modifier la définition de T^* et prendre

$$T^* = \sup\left\{T \geq \hat{t}, \forall i, \int_{\hat{t}}^T \|\varphi_i\| dt \leq \frac{\delta_0}{2C}\right\}.$$

Pour conclure, il suffit alors de montrer que si on a l'estimation (15), et donc aussi (16), les termes d'interactions entre solutions de (13) sont en fait meilleurs et vérifient une estimation du type

$$(17) \quad \int_{\hat{t}}^T \|\varphi_i(s)\| ds \leq C\delta_0^2, \quad \forall t \in [\hat{t}, T].$$

pour C indépendant de T .

4. ESTIMATIONS PARABOLIQUES

PROPOSITION 4.1. — *Il existe $\kappa > 0$, $\delta_0 > 0$ et \hat{t} telles que si $\|u_{0x}\| \leq \frac{\delta_0}{2\kappa}$, alors la solution de (10) telle que $u(0, x) = u_0(x)$ vérifie*

$$(18) \quad \|u_x(t)\| \leq \frac{\delta_0}{2}, \quad \forall t \in [0, \hat{t}].$$

De plus \hat{t} est de l'ordre de $\frac{1}{\delta_0^2}$.

Enfin, si on conserve la borne

$$(19) \quad \|u_x(t)\| \leq \delta_0, \quad \forall t \in [0, T]$$

pour $T > \hat{t}$, on a alors de meilleures estimations pour les dérivées d'ordre supérieur, par exemple

$$(20) \quad \|u_{xx}(t)\|, \|u_x(t)\|_\infty = \mathcal{O}(1)\delta_0^2.$$

La fin de la proposition montre que la partie difficile de la preuve du théorème 3.1 est vraiment l'obtention de la borne (19), puisque cela entraîne automatiquement un meilleur comportement des dérivées d'ordre supérieur.

La preuve se fait par une méthode de point fixe classique. On remarque d'abord que pour une solution $u(t, x)$ de (10), l'état

$$u^* = \lim_{x \rightarrow -\infty} u(t, x)$$

est indépendant de t , on récrit alors l'équation (10) comme

$$(21) \quad u_t + A(u^*)u_x - u_{xx} = \left(A(u^*) - A(u) \right) u_x$$

c'est-à-dire comme une perturbation du système parabolique linéaire

$$u_t + A(u^*)u_x - u_{xx}.$$

Appelons $G(t, x)$ la matrice de Green de ce système linéaire ; on peut alors récrire (21) comme une équation intégrale

$$(22) \quad u(t) - u^* = G(t) \star (u_0 - u^*) + \int_0^t G(t-s) \star \left(A(u^*) - A(u) \right) u_x(s) ds$$

$$(23) \quad u_x(t) = G(t) \star u_{0x} + \int_0^t G_x(t-s) \star (A(u^*) - A(u)) u_x(s) ds$$

où les convolutions sont des convolutions par rapport à la variable d'espace x . Pour ce système linéaire, la matrice de Green est explicite ; elle s'écrit

$$G(t, x) = \sum_{i=1}^n G_i(t, x) r_i(u^*) l_i(u^*),$$

où les $l_i(u^*)$ sont les vecteurs propres à gauche de $A(u^*)$ et

$$G_i(t, x) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{(x - \lambda_i(u^*)t)^2}{4t}\right).$$

En particulier on a les estimations

$$\|G(t)\| \leq \kappa, \|G_x(t)\| \leq \frac{\kappa}{\sqrt{t}}, \|G_{xx}(t)\| \leq \frac{\kappa}{t}.$$

On obtient facilement la preuve de la première partie de la proposition par un argument de point fixe et des estimations de convolutions. Pour la deuxième partie, on utilise la même méthode sur $[t - \hat{t}, t]$. On obtient par exemple

$$u_{xx}(t) = G_x(\hat{t}) \star u_x(\hat{t}) + \dots$$

et donc

$$\|u_{xx}(t)\| \leq \frac{C}{\sqrt{\hat{t}}} \delta_0 \leq C \delta_0^2$$

puisque $\hat{t} \sim \frac{1}{\delta_0^2}$.

5. DÉCOMPOSITION D'ONDES À L'AIDE DE VARIÉTÉS CENTRALES

On passe maintenant à la deuxième partie du programme qui consiste à trouver une décomposition (11) telle que les termes sources φ_i vérifient (17), c'est-à-dire

$$(24) \quad \int_0^{+\infty} \|\varphi_i(t)\| dt < +\infty.$$

En s'inspirant des techniques utilisées pour les systèmes hyperboliques, une première approche naturelle est d'essayer la décomposition

$$(25) \quad u_x = \sum_{i=1}^n v_i r_i(u)$$

où les $r_i(u)$ sont les vecteurs propres de $A(u)$. Cette décomposition peut effectivement être utilisée (voir [3]) lorsque le système (10) possède des propriétés très particulières : les courbes intégrales des champs de vecteurs r_i sont des droites. Cependant une propriété fondamentale des équations paraboliques (10) est l'existence d'ondes progressives du type $u(t, x) = V(x - \sigma t)$ avec V régulière. En général, lorsqu'on essaie la décomposition (25)

pour une onde progressive, on trouve des termes sources φ_i qui ne sont pas nuls et qui dépendent en fait de la seule variable $\xi = x - \sigma t$. Il est donc impossible d'avoir (24). Ce phénomène peut se vérifier [5] sur un système triangulaire

$$\begin{aligned} u_t + f(u)_x &= u_{xx}, \\ v_t + g(u, v)_x &= v_{xx}. \end{aligned}$$

L'idée pour surmonter cette difficulté est précisément de chercher en tout point x une décomposition de u_x en somme de dérivées d'ondes progressives U_i passant par $u(x)$. En fait, à cause de la présence de viscosité il faut chercher une décomposition simultanée de u_x et u_{xx} :

$$(26) \quad u_x = \sum_{i=1}^n U_i'(x), \quad u_{xx} = \sum_{i=1}^n U_i''(x).$$

Une onde progressive est solution de l'équation différentielle du second ordre

$$(27) \quad U_i'' = (A(U_i) - \sigma_i)U_i',$$

σ_i étant la vitesse. Ainsi, pour tout i , on peut trouver une famille à $n + 1$ paramètres d'ondes progressives en prescrivant σ_i et U_i' . Cela donne $n(n+1)$ paramètres à déterminer, ce qui est beaucoup trop puisque (26) est un système de $2n$ équations. Pour régler ce problème, l'idée est de se restreindre à des solutions U_i bornées sur \mathbb{R} et donc de les chercher dans une variété centrale. Plus précisément, pour avoir le bon nombre de paramètres, on va chercher n familles d'ondes progressives dépendant de 2 paramètres lorsqu'on a fixé $U_i(x) = u(x)$.

On peut récrire le système (27) comme un système du premier ordre sur $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$:

$$(28) \quad \begin{cases} u' = v, \\ v' = (A(u) - \sigma)v, \\ \sigma' = 0, \end{cases}$$

puis linéariser ce système au point d'équilibre $P_i^* = (u^*, 0, \lambda_i(u^*))$, ce qui donne

$$(29) \quad \begin{cases} u' = v, \\ v' = (A(u^*) - \lambda_i(u^*))v, \\ \sigma' = 0. \end{cases}$$

Décomposons un vecteur v de \mathbb{R}^n sur la base (r_1^*, \dots, r_n^*) ,

$$v = \sum_{j=1}^n V_j r_j^*$$

où $r_j^* = r_j(u^*)$. Le noyau (généralisé) \mathcal{N}_i du système linéaire (29) est de dimension $n + 2$, il est constitué des vecteurs (u, v, σ) tels que

$$V_j = 0, \quad \forall j \neq i.$$

Le théorème de variété centrale [43] donne l'existence d'une variété \mathcal{M}_i tangente à \mathcal{N}_i au point P_i^* et contenant toutes les trajectoires de (28) qui restent dans un petit voisinage de P_i^* . Cette variété peut donc être décrite localement par les équations

$$V_j = \varphi_j^i(u, V_i, \sigma), \quad j \neq i.$$

De plus, \mathcal{M}_i contient les équilibres $(u, 0, \sigma)$ au voisinage de P_i^* , on a donc $\phi_j^i(u, 0, \sigma) = 0$. Cela permet de factoriser φ_j^i et donc d'écrire pour $v \in \mathcal{M}_i$,

$$(30) \quad v = V_i \tilde{r}_i(u, V_i, \sigma)$$

avec (il est plus pratique d'avoir des vecteurs de norme 1)

$$\tilde{r}_i = \frac{r_i(u^*) + \sum_{j \neq i} \frac{\varphi_j^i(u, V_i, \sigma)}{V_i}}{\left| r_i(u^*) + \sum_{j \neq i} \frac{\varphi_j^i(u, V_i, \sigma)}{V_i} \right|}.$$

Sur la variété \mathcal{M}_i , on a donc une décomposition proche de (11) mais \tilde{r}_i est différent de $r_i(u)$. En effet, l'identité $(A(u) - \lambda_i)r_i = 0$ est remplacée par

$$(A(u) - \tilde{\lambda}_i)\tilde{r}_i = v_i(\tilde{r}_{i,u}\tilde{r}_i + \tilde{r}_{i,v}(\tilde{\lambda}_i - \sigma_i))$$

où $\tilde{\lambda}_i = (A(u)\tilde{r}_i, \tilde{r}_i)$. En dépit de sa complexité apparente, cette identité s'avère cruciale pour parvenir à des termes sources φ_i intégrables. Finalement, on remarque que lorsque $u(t, x)$ est une onde progressive dans la variété centrale \mathcal{M}_i , i.e $u(t, x) = U_i(x - \sigma_i t)$, on a alors $u_x = v = v_i \tilde{r}_i$ et donc en dérivant, on obtient

$$v_x = v_{ix}\tilde{r}_i + v_i\tilde{r}_{ix} = (A(u) - \sigma_i)v_i\tilde{r}_i.$$

Ce qui donne en prenant le produit scalaire par \tilde{r}_i

$$v_{ix} = (\tilde{\lambda}_i - \sigma_i)v_i$$

et donc en dérivant de nouveau

$$v_{it} + (\tilde{\lambda}_i v_i)_x = v_{ixx}$$

puisque $v_{it} = -\sigma_i v_{ix}$. On trouve donc que v_i évolue selon l'équation de transport diffusion souhaitée sans terme source φ_i dans ce cas particulier.

L'étape suivante est de parvenir à la décomposition (26). En se basant sur ce qui précède on voudrait prendre $U_i' = v_i \tilde{r}_i(u, v_i, \sigma_i)$: il faut donc identifier les paramètres v_i et σ_i en fonction de u , u_x et u_{xx} . Considérons de nouveau le cas où u est une onde progressive dans la variété centrale \mathcal{M}_i . On a alors $u_x = v_i \tilde{r}_i$ et donc \tilde{r}_i étant unitaire, $v_i = \pm |u_x|$. En ce qui concerne la vitesse, on a $u_t = -\sigma_i u_x$, donc u_t est colinéaire à \tilde{r}_i et on peut écrire

$$u_t = u_{xx} - A(u)u_x = \omega_i \tilde{r}_i.$$

La vitesse est alors donnée par $\sigma_i = -\frac{\omega_i}{v_i}$.

En s'inspirant de ce qui précède, pour le cas général, on a envie de définir $u_t = u_{xx} - A(u)u_x$ et d'essayer de trouver v_i, ω_i tels que

$$u_x = \sum_i v_i \tilde{r}_i(u, v_i, \sigma_i), \quad u_t = \sum_i \omega_i \tilde{r}_i(u, v_i, \sigma_i)$$

où $\sigma_i = -\frac{\omega_i}{v_i}$. Le problème de cette décomposition est que les vecteurs \tilde{r}_i ne sont définis que pour des vitesses σ_i proches de λ_i^* alors que le rapport ω_i/v_i peut devenir très grand lorsque $|u_x|$ est très petit. Pour surmonter cela, on utilise une fonction de troncature $\theta \in \mathcal{C}_c^\infty(\mathbb{R})$ telle que $\theta(x) = x$ dans un voisinage de l'origine. On pose alors $\omega_i = w_i - \lambda_i^* v_i$ et on cherche la décomposition

$$(31) \quad u_x = \sum_i v_i \tilde{r}_i(u, v_i, \sigma_i), \quad u_t = \sum_i (w_i - \lambda_i^* v_i) \tilde{r}_i(u, v_i, \sigma_i)$$

avec

$$(32) \quad u_t = u_{xx} - A(u)u_x, \quad \sigma_i = \lambda_i^* - \theta\left(\frac{w_i}{v_i}\right).$$

En fait, il y a toujours un problème lorsque $v_i = w_i = 0$, mais dans ce cas ce n'est pas gênant, car on a toujours $\tilde{r}_i(u, 0, \sigma_i) = r_i(u)$ quelle que soit la valeur de σ_i .

Le théorème des fonctions implicites permet ensuite de montrer que l'on peut effectivement parvenir à la décomposition (31), (32).

PROPOSITION 5.1. — *Pour $|u - u^*|, |u_x|$ et $|u_{xx}|$ suffisamment petits, le système d'équations (31) a une unique solution $(v, w) = (v_1, \dots, v_n, w_1, \dots, w_n)$. De plus l'application $(u, u_x, u_{xx}) \mapsto (v, w)$ est \mathcal{C}^∞ en dehors des $\mathcal{N}_i = \{v_i = w_i = 0\}$ et elle est $\mathcal{C}^{1,1}$ dans un voisinage de $(u^*, 0, 0)$.*

On peut également montrer que tant qu'on a l'estimation (19), les composantes v_i et w_i satisfont des estimations de type (20).

6. ESTIMATIONS DES TERMES SOURCES

Les composantes (v_i, w_i) de la décomposition (31) évoluent alors selon les équations

$$(33) \quad v_{i,t} + (\tilde{\lambda}_i v_i)_x - v_{i,xx} = \varphi_i, \quad w_{i,t} + (\tilde{\lambda}_i w_i)_x - w_{i,xx} = \psi_i.$$

On avait déjà vu que dans le cas idéal où u est une onde progressive, on trouvait ces équations sans termes sources.

Les termes sources ont donc trois origines possibles :

1. Il y a des ondes de deux différentes familles $j \neq k$ présentes dans la décomposition (31) au point x . Cela donne des termes quadratiques dus à des interactions transverses. Dans ce cas-là, leur contribution à φ_i, ψ_i est donnée par

$$(34) \quad T_1 = \mathcal{O}(1) \sum_{j \neq k} \left(|v_j v_k| + |v_{j,x} v_k| + |v_j w_k| + |v_{j,x} w_k| + |v_j w_{k,x}| + |w_j w_k| \right).$$

2. La décomposition (31) étant définie ponctuellement, deux ondes progressives de la même famille, peuvent avoir des vitesses différentes en des points différents, l'interaction infinitésimale entre des ondes voisines de la même famille sera donc déterminée par le taux de variation de la vitesse $\sigma_{j,x}$. Dans ce cas, la contribution aux termes sources peut être linéaire ou quadratique :

$$\mathcal{O}(1)v_j^2\sigma_{j,x} + \mathcal{O}(1)v_j^2\sigma_{j,x}^2.$$

En se souvenant que l'on a $\sigma_j = \lambda_j^* - \theta\left(\frac{w_j}{v_j}\right)$, on peut récrire la contribution aux termes sources comme

$$(35) \quad T_2^l = \mathcal{O}(1)|w_{j,x}v_j - v_{j,x}w_j|, \quad T_2^q = \mathcal{O}(1)\left|v_j\left(\frac{w_j}{v_j}\right)_x\right|^2 \chi_{\{\frac{w}{v} \in \text{Supp}(\theta)\}}.$$

3. Le dernier type de terme source apparaît lorsque la fonction de troncature θ est active, dans ce cas-là, il y a une erreur parce que l'on a fait un mauvais choix de la vitesse σ_j . On a alors une borne

$$(36) \quad T_3 = \mathcal{O}(1)(|v_{j,x}| + |w_{j,x}|)|w_j - \theta_j v_j|, \quad \theta_j = \theta\left(\frac{w_j}{v_j}\right).$$

Nous n'insisterons pas sur la preuve des bornes (34), (35), (36). C'est la partie la plus technique de la preuve du théorème 3.1.

La dernière partie de la démonstration du théorème 3.1 consiste à montrer une estimation du type (17). Pour cela on utilise différentes fonctions de Lyapounov pour contrôler les trois types de termes d'interactions que l'on vient de décrire.

6.1. Interactions transverses

On commence par l'estimation d'un terme de type T_1 i.e. $\int_0^{+\infty} \int_{\mathbb{R}} |v_j v_k| dx dt$ avec $j \neq k$. Le problème modèle à étudier est un système

$$(37) \quad \begin{cases} v_t + (\lambda v)_x = v_{xx}, & v(0, x) = v_0(x), \\ w_t + (\lambda w)_x = w_{xx}, & w(0, x) = w_0(x), \end{cases}$$

avec

$$(38) \quad \inf_{t,x} \lambda(t, x) - \sup_{t,x} \mu(t, x) \geq c > 0.$$

On peut montrer l'estimation

$$(39) \quad \int_0^T \int_{\mathbb{R}} |vw| dx dt \leq \|v_0\| \|w_0\|$$

en utilisant un potentiel d'interaction

$$Q(v, w) = \int_{\mathbb{R}^2} K(x - y)|v(y)||w(x)| dx dy$$

où K est de la forme $K(s) = \frac{1}{c}$ si $s \geq 0$, $K(s) = \frac{1}{c}e^{\frac{cs}{2}}$ si $s < 0$. Un calcul montre que

$$\frac{d}{dt}Q(v, w) + \int_{\mathbb{R}} |v(t, x)| |w(t, x)| dx \leq 0$$

ce qui donne (39). Ce potentiel peut être vu comme une version parabolique du potentiel pour les futures interactions d'onde [24]. On remarque que le potentiel a un rôle important dans la zone $x > y$, c'est-à-dire celle où on s'attend à une interaction compte tenu de (38).

6.2. Interactions d'ondes de la même famille

On passe maintenant à l'estimation des deux sortes de terme de type T_2 . Le problème modèle est encore (37), mais cette fois-ci $\lambda = \mu$. On définit alors une courbe plane $\gamma = (\int_{-\infty}^x v, \int_{-\infty}^x w)$, qui évolue selon

$$(40) \quad \gamma_t + \lambda \gamma_x = \gamma_{xx}.$$

D'après cette équation, γ évolue dans la direction de la courbure, une fonction de Lyapounov est alors la longueur de la courbe

$$L(\gamma(t)) = \int_{\mathbb{R}} |\gamma_x| dx = \int_{\mathbb{R}} (v^2 + w^2)^{\frac{1}{2}} dx.$$

Une autre fonction de Lyapounov introduite dans [6] est la « fonctionnelle d'aire »

$$\mathcal{A}(\gamma(t)) = \frac{1}{2} \iint_{x < y} |\gamma_x(t, x) \wedge \gamma_x(t, y)| dx dy.$$

Pour comprendre la signification de $\mathcal{A}(\gamma)$, on remarque que dans le cas d'une courbe fermée, on a

$$\int \gamma(y) \wedge \gamma_x(y) dy = \frac{1}{2} \iint_{x < y} \gamma_x(t, x) \wedge \gamma_x(t, y) dx dy,$$

$\mathcal{A}(\gamma)$ représente donc la somme des aires des régions entourées par la courbe multipliées par le nombre de tours. On remarque aussi que pour une loi de conservation scalaire $u_t + f(u)_x = u_{xx}$, la courbe $\gamma = (u, u_x - f(u))$ évolue selon l'équation (40) avec $\lambda = f'$. De plus, si on définit la vitesse infinitésimale d'une onde par $s(x) = -u_t(t, x)/u_x(t, x)$, on obtient

$$\begin{aligned} \mathcal{A}(\gamma) &= \frac{1}{2} \iint_{x < y} |u_x(t, x)u_t(t, y) - u_t(t, x)u_x(t, y)| dx dy \\ &= \frac{1}{2} \iint_{x < y} |u_x(t, x)| |u_x(t, y)| |s(x) - s(y)| dx dy \\ &= \frac{1}{2} \iint_{x < y} \left(\text{onde en } x \times \text{onde en } y \times \text{différence des vitesses} \right) dx dy \end{aligned}$$

ce qui indique bien que $\mathcal{A}(\gamma)$ peut être vu comme un potentiel d'interaction entre ondes de la même famille.

Le taux de dissipation de ces deux fonctions de Lyapounov permet de contrôler les termes de type T_1^l, T_1^q .

En utilisant la base de Frénet (τ, n) , et en notant c , la courbure de $\gamma(t, \cdot)$, on peut récrire (40) comme

$$\gamma_t = \left(|\gamma_x|_x - \lambda |\gamma_x| \right) \tau + |\gamma_x|^2 c n.$$

Ainsi, on trouve

$$\frac{d}{dt}L(\gamma(t)) = \int_{\mathbb{R}} \gamma_{xt} \cdot \tau \, dx = - \int_{\mathbb{R}} \gamma_t \cdot n \, c |\gamma_x| \, dx = - \int_{\mathbb{R}} |\gamma_x|^3 c^2 \, dx = - \int_{\mathbb{R}} \frac{(wv_x - vw_x)^2}{(v^2 + w^2)^{\frac{3}{2}}} \, dx,$$

ce qui s'écrit finalement

$$\frac{d}{dt}L(\gamma(t)) + \int_{\mathbb{R}} |v| \left[\left(\frac{w}{v} \right)_x \right]^2 \frac{1}{\left(1 + \left(\frac{w}{v} \right)^2 \right)^{\frac{3}{2}}} \, dx \leq 0.$$

Cela permet de contrôler les termes quadratiques T_2^q .

De même, un calcul plus long (voir [6]) montre que

$$\frac{d}{dt}\mathcal{A}(\gamma(t)) + \int_{\mathbb{R}} |w_x v - v_x w| \, dx \leq 0$$

ce qui donne un contrôle des termes de type T_2^l .

6.3. Estimations d'énergie

Il reste à contrôler le dernier type de terme qui correspond à un mauvais choix de la vitesse. Pour expliquer les idées, revenons au cas idéal où la décomposition (31) ne comporterait qu'une seule onde progressive. Dans ce cas $u_x = v_i \tilde{r}_i$ et par définition, $u_{xx} - A(u)u_x = \omega_i \tilde{r}_i = (w_i - \lambda_i^*) \tilde{r}_i$. On trouve donc finalement

$$(41) \quad v_{ix} - \tilde{\lambda}_i v_i = w_i - \lambda_i^* v_i.$$

En remplaçant dans (36), on voit que le terme prépondérant à majorer est $\mathcal{O}(1)|v_{ix}|^2$ lorsque $|w_i/v_i| \geq A$, où A détermine la taille du support de θ , c'est-à-dire en utilisant de nouveau (41) lorsque

$$(42) \quad |v_{ix}/v_i| \geq A/2.$$

Le terme à estimer correspond au taux de dissipation de l'énergie usuelle

$$E(t) = \frac{1}{2} \int_{\mathbb{R}} |v_i|^2 \, dx.$$

Cependant, pour l'équation

$$v_{it} + (\tilde{\lambda}_i v_i)_x - v_{ixx} = 0.$$

l'énergie n'est en général pas décroissante, on a seulement

$$\frac{d}{dt}E(t) + \int_{\mathbb{R}} |v_{ix}|^2 \, dx \leq \int_{\mathbb{R}} |\lambda| |v_{ix}| |v_i| \, dx,$$

mais en choisissant bien la taille du support de θ , on conclut quand même en utilisant (42).

7. ESTIMATIONS DE STABILITÉ

On a seulement insisté sur l'obtention d'une borne VT uniforme pour (10). L'estimation de stabilité (7) s'obtient en utilisant le même type de technique et un argument homotopique. Soient u, v deux solutions de (10) avec données initiales \bar{u} et \bar{v} . On considère un chemin régulier $\bar{u}^\theta = \theta\bar{u} + (1 - \theta)\bar{v}$ et u^θ la solution de (10) avec donnée initiale \bar{u}^θ . Puisqu'on a

$$\|u(t) - v(t)\| \leq \int_0^1 \left\| \frac{du^\theta}{d\theta} \right\| d\theta,$$

il suffit de montrer une estimation du type

$$(43) \quad \left\| \frac{du^\theta}{d\theta} \right\| \leq L \left\| \frac{du^\theta(\theta=0)}{d\theta} \right\| = L\|\bar{u} - \bar{v}\|$$

pour avoir (7).

Posons $z^\theta = \frac{du^\theta}{d\theta}$, z^θ est solution de l'équation linéaire

$$(44) \quad z_t^\theta + \left(dA(u^\theta) \cdot z^\theta \right) u_x^\theta + A(u^\theta) z_x^\theta = z_{xx}^\theta$$

avec donnée initiale $\bar{u}(x) - \bar{v}(x)$. Pour montrer que z^θ vérifie (43), on utilise une décomposition de z^θ et $z_x^\theta - A(u^\theta)z^\theta$ analogue à (31).

8. CONCLUSION

Les nouvelles méthodes issues de [7] devraient avoir très rapidement de nouvelles applications : S. Bianchini est parvenu [2] à montrer la convergence d'approximations semi-discrètes en espace de (1) :

$$\frac{\partial u}{\partial t}(x, t) + \frac{1}{\Delta x} \frac{\partial}{\partial x} \left(f(u(x, t)) - f(u(x - \Delta x, t)) \right) = 0.$$

En ce qui concerne la convergence des schémas numériques complètement discrets, il a été montré que l'on ne pouvait pas espérer d'estimations VB analogues à (6) [12]. Enfin, A. Bressan et T. Yang [19] ont très récemment estimé le taux de convergence de la solution u^ε de (5) vers la solution u de (1) sous les hypothèses (SH) et (\diamond). Ils obtiennent une estimation qui paraît presque optimale $\|u^\varepsilon(t) - u(t)\|_{L^1} = \mathcal{O}(\varepsilon^{\frac{1}{2}} |\ln \varepsilon|)$.

BIBLIOGRAPHIE

- [1] P. Baiti et H. K. Jenssen. On the front-tracking algorithm. *J. Math. Anal. Appl.*, 217(2):395–404, 1998.
- [2] S. Bianchini. BV solutions of the semidiscrete upwind scheme. *Preprint*, 2002.
- [3] S. Bianchini et A. Bressan. BV solutions for a class of viscous hyperbolic systems. *Indiana Univ. Math. J.*, 49(4):1673–1713, 2000.

- [4] S. Bianchini et A. Bressan. A case study in vanishing viscosity. *Discrete Contin. Dynam. Systems*, 7(3):449–476, 2001.
- [5] S. Bianchini et A. Bressan. A center manifold technique for tracing viscous waves. *Commun. Pure Appl. Anal.*, 1(2):161–190, 2002.
- [6] S. Bianchini et A. Bressan. On a Lyapunov functional relating shortening curves and viscous conservation laws. *Nonlinear Anal.*, 51(4, Ser. A: Theory Methods):649–662, 2002.
- [7] S. Bianchini et A. Bressan. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Preprint*, 2002.
- [8] A. Bressan. Global solutions of systems of conservation laws by wave-front tracking. *J. Math. Anal. Appl.*, 170(2):414–432, 1992.
- [9] A. Bressan. The unique limit of the Glimm scheme. *Arch. Rational Mech. Anal.*, 130(3):205–230, 1995.
- [10] A. Bressan. Hyperbolic systems of conservation laws. *Rev. Mat. Complut.*, 12(1):135–200, 1999.
- [11] A. Bressan. *Hyperbolic systems of conservation laws*, volume 20 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2000. The one-dimensional Cauchy problem.
- [12] A. Bressan, P. Baiti et H. K. Jenssen. An instability of the Godunov scheme. *Preprint*.
- [13] A. Bressan et R. M. Colombo. The semigroup generated by 2×2 conservation laws. *Arch. Rational Mech. Anal.*, 133(1):1–75, 1995.
- [14] A. Bressan, G. Crasta et B. Piccoli. Well-posedness of the Cauchy problem for $n \times n$ systems of conservation laws. *Mem. Amer. Math. Soc.*, 146(694):viii+134, 2000.
- [15] A. Bressan et P. Goatin. Oleinik type estimates and uniqueness for $n \times n$ conservation laws. *J. Differential Equations*, 156(1):26–49, 1999.
- [16] A. Bressan et P. Le Floch. Uniqueness of weak solutions to systems of conservation laws. *Arch. Rational Mech. Anal.*, 140(4):301–317, 1997.
- [17] A. Bressan et M. Lewicka. A uniqueness condition for hyperbolic systems of conservation laws. *Discrete Contin. Dynam. Systems*, 6(3):673–682, 2000.
- [18] A. Bressan, T.-P. Liu et T. Yang. L^1 stability estimates for $n \times n$ conservation laws. *Arch. Ration. Mech. Anal.*, 149(1):1–22, 1999.
- [19] A. Bressan et T. Yang. On the rate of convergence of the vanishing viscosity approximation. *Preprint*, 2003
- [20] C. M. Dafermos. Polygonal approximations of solutions of the initial value problem for a scalar conservation law. *J. Math. Anal. Appl.*, 38:33–41, 1972.
- [21] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2000.

- [22] G. Dal Maso, P.G. Le Floch et F. Murat. Definition and weak stability of nonconservative products. *J. Math. Pure Appl.*, 74:483–548, 1995.
- [23] R. J. DiPerna. Convergence of approximate solutions to conservation laws. *Arch. Rational Mech. Anal.*, 82(1):27–70, 1983.
- [24] J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Comm. Pure Appl. Math.*, 18:697–715, 1965.
- [25] J. Goodman. Nonlinear asymptotic stability of viscous shock profiles for conservation laws. *Arch. Rational Mech. Anal.*, 95(4):325–344, 1986.
- [26] J. Goodman et Z.-P. Xin. Viscous limits for piecewise smooth solutions to systems of conservation laws. *Arch. Rational Mech. Anal.*, 121(3):235–265, 1992.
- [27] O. Guès, G. Métivier, M. Williams et K. Zumbrun. Multidimensional viscous shocks I, II. *Preprint*, 2002.
- [28] T. Iguchi et P.G. Le Floch. Existence theory for hyperbolic systems of conservation laws with general flux-functions, *Preprint*, 2002.
- [29] H.K. Jenssen. Blowup for systems of conservation laws. *SIAM J. Math. Anal.*, 31(4):894–908 (electronic), 2000.
- [30] S. Khruzhkov. First order quasilinear equations with several space variables. *Math. USSR Sbornik*, 10:217–243, 1970.
- [31] P.D. Lax. Hyperbolic systems of conservation laws II. *Comm. Pure Appl. Math.*, 10:537–566, 1957.
- [32] P.G. Le Floch. *Hyperbolic systems of conservation laws*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2002. The theory of classical and nonclassical shock waves.
- [33] T.-P. Liu. Admissible solutions of hyperbolic conservation laws. *Mem. Amer. Math. Soc.*, 30(240):iv+78, 1981.
- [34] T.-P. Liu. Nonlinear stability of shock waves for viscous conservation laws. *Mem. Amer. Math. Soc.*, 56(328):v+108, 1985.
- [35] T.-P. Liu et T. Yang. A new entropy functional for a scalar conservation law. *Comm. Pure Appl. Math.*, 52(11):1427–1442, 1999.
- [36] T.-P. Liu et T. Yang. Well-posedness theory for hyperbolic conservation laws. *Comm. Pure Appl. Math.*, 52(12):1553–1586, 1999.
- [37] T.-P. Liu et T. Yang. Weak solutions of general systems of hyperbolic conservation laws. *Comm. Math. Phys.*, 230(2):289–327, 2002.
- [38] N.H. Risebro. A front-tracking alternative to the random choice method. *Proc. Amer. Math. Soc.*, 117:1125–1139, 1993.
- [39] F. Rousset. Viscous limits for strong shocks of systems of conservation laws. *SIAM J. Math. Anal.*, à paraître.
- [40] D. Serre. *Systems of conservation laws. 1, 2*. Cambridge University Press, Cambridge, 2000. Translated from the 1996 French original by I. N. Sneddon.

- [41] A. Szepessy et Z.-P. Xin. Nonlinear stability of viscous shock waves. *Arch. Rational Mech. Anal.*, 122(1):53–103, 1993.
- [42] A. Szepessy et K. Zumbrun. Stability of rarefaction waves in viscous media. *Arch. Rational Mech. Anal.*, 133(3):249–298, 1996.
- [43] A. Vanderbauwhede. Centre manifolds, normal forms and elementary bifurcations. In *Dynamics reported, Vol. 2*, volume 2 of *Dynam. Report. Ser. Dynam. Systems Appl.*, pages 89–169. Wiley, Chichester, 1989.
- [44] S.-H. Yu. Zero-dissipation limit of solutions with shocks for systems of hyperbolic conservation laws. *Arch. Ration. Mech. Anal.*, 146(4):275–370, 1999.
- [45] K. Zumbrun. Multidimensional stability of planar viscous shock waves. In *Advances in the theory of shock waves*, volume 47 of *Progr. Nonlinear Differential Equations Appl.*, pages 307–516. Birkhäuser Boston, Boston, MA, 2001.
- [46] K. Zumbrun et P. Howard. Pointwise semigroup methods and stability of viscous shock waves. *Indiana Univ. Math. J.*, 47(3):741–871, 1998.

Frédéric ROUSSET

CNRS, Université de Nice

Laboratoire J.A. Dieudonné

UMR 6621 du CNRS

28 avenue Valrose

B.P. 2135

F-06103 Nice Cedex 2

E-mail : frousset@math.unice.fr

LA CONJECTURE DE BIRCH ET SWINNERTON-DYER p -ADIQUE

par **Pierre COLMEZ**

Notations

Dans tout l'article, $\overline{\mathbf{Q}}$ désigne la clôture algébrique de \mathbf{Q} dans \mathbf{C} et un plongement de $\overline{\mathbf{Q}}$ dans $\overline{\mathbf{Q}}_p$ est fixé; en particulier, $\mathcal{G}_{\mathbf{Q}_p} = \text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ est un sous-groupe bien déterminé de $\mathcal{G}_{\mathbf{Q}} = \text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. On note χ_{cycl} le caractère cyclotomique. Si $M \geq 1$ est un entier, on note $\zeta_M \in \overline{\mathbf{Q}}$ la racine de l'unité $\exp(\frac{2i\pi}{M})$.

0. INTRODUCTION

Si \mathbb{M} est un motif défini sur un corps de nombres, on sait lui associer (au moins conjecturalement) une fonction analytique complexe $L(\mathbb{M}, s)$ définie par un produit eulérien convergeant dans un demi-plan. La quantité d'information arithmétique contenue dans les valeurs aux entiers de ces fonctions L est absolument fascinante, et on dispose d'un faisceau de conjectures la décrivant (conjectures de Deligne [64], de Beilinson [7, 178, 160], et de Bloch-Kato [27, 76, 77]). Ces conjectures sont de lointaines descendantes de la conjecture de Birch et Swinnerton-Dyer [25, 26, 182], elle-même inspirée par la formule analytique du nombre de classes d'idéaux qui reste le seul cas général que l'on sache traiter. En p -adique, on a plus de mal à définir les fonctions L mais, une fois définies, celles-ci livrent un peu plus facilement l'information qu'elles contiennent comme nous le verrons dans ce texte pour les fonctions L p -adiques de courbes elliptiques définies sur \mathbf{Q} ou, plus généralement, de formes modulaires.

0.1. La formule analytique du nombre de classes

Soit K un corps de nombres d'anneau des entiers \mathcal{O}_K . La fonction zêta de Dedekind ζ_K de K est définie, pour $\text{Re}(s) > 1$, par le produit eulérien $\zeta_K(s) = \prod_{\mathfrak{p}} \frac{1}{1-N_{\mathfrak{p}}^{-s}}$, le produit portant sur les idéaux maximaux de \mathcal{O}_K , et possède un prolongement à \mathbf{C} tout entier, holomorphe en dehors d'un pôle simple en $s = 1$. On dispose des résultats suivants :

THÉORÈME 0.1. — (i) *Le groupe \mathcal{O}_K^* des unités de \mathcal{O}_K est de type fini sur \mathbf{Z} . Son rang $r(K)$ est égal à $r_1 + r_2 - 1$, où r_1 est le nombre de places réelles de K et r_2 le nombre de ses places complexes.*

(ii) *Le régulateur¹ $R_\infty(K)$ est non nul.*

THÉORÈME 0.2. — *Le groupe $\text{Pic}(\mathcal{O}_K)$ des classes d'idéaux de \mathcal{O}_K est un groupe fini.*

THÉORÈME 0.3. — *La fonction $\zeta_K(s)$ a, en $s = 0$, un zéro d'ordre $r(K)$ et on a*

$$\lim_{s \rightarrow 0} s^{-r(K)} \zeta_K(s) = -|\text{Pic}(\mathcal{O}_K)| \cdot R_\infty(K).$$

0.2. La conjecture de Birch et Swinnerton-Dyer

Soit E une courbe elliptique définie² sur \mathbf{Q} de conducteur N_E . Si p est un nombre premier, on définit l'entier a_p par :

- si $p^2 \mid N_E$ (i.e. si E a réduction additive en p), alors $a_p = 0$;
- si $p \mid N_E$ et $p^2 \nmid N_E$ (i.e. si E a réduction multiplicative en p), alors $a_p = 1$ (resp. $a_p = -1$) si E a réduction multiplicative déployée (resp. non déployée) ;
- si $p \nmid N_E$ (i.e. si E a bonne réduction en p), alors $a_p = p + 1 - |E(\mathbf{F}_p)|$.

Ceci nous permet de définir la fonction L complexe $L(E, s)$ attachée à E par le produit eulérien (convergeant pour $\text{Re}(s) > \frac{3}{2}$ car $|a_p| \leq 2\sqrt{p}$ d'après le théorème de Hasse) :

$$L(E, s) = \prod_{p \nmid N_E} \frac{1}{1 - a_p p^{-s} + p^{1-2s}} \prod_{p \mid N_E} \frac{1}{1 - a_p p^{-s}} = \sum_{n=1}^{+\infty} a_n n^{-s}.$$

La fonction $L(E, s)$ est holomorphe sur le demi-plan $\text{Re}(s) > \frac{3}{2}$ et possède un prolongement analytique à tout le plan complexe (voir cor. 0.18).

Les rôles joués par ζ_K , \mathcal{O}_K^* et $\text{Pic}(\mathcal{O}_K)$ dans la formule analytique du nombre de classes sont ici joués respectivement par $L(E, s)$, le groupe $E(\mathbf{Q})$ des points de E rationnels sur \mathbf{Q} et le groupe de Tate-Shafarevitch³ $\sqcup(E)$ de E , et les théorèmes 0.1, 0.2 et 0.3 deviennent respectivement :

¹Ce régulateur est défini de la manière suivante : on part d'une famille $\varepsilon_1, \dots, \varepsilon_{r(K)}$ d'éléments de \mathcal{O}_K^* engendrant un sous-groupe d'indice fini U de \mathcal{O}_K^* et de plongements $\sigma_1, \dots, \sigma_{r(K)}$ de K dans \mathbf{C} induisant des places différentes et on pose $e_i = 1$ si σ_i induit une place réelle et $e_i = 2$ si σ_i induit une place complexe ; alors $R_\infty(K) = \frac{1}{[\mathcal{O}_K^* : U]} |\det(e_i \log |\sigma_i(\varepsilon_j)|)_{1 \leq i, j \leq r(K)}|$ ne dépend d'aucun des choix que l'on a faits.

²On se restreint aux courbes définies sur \mathbf{Q} car ce sont les seules pour lesquelles on peut démontrer quoi que ce soit grâce à leur modularité (cf. § 0.4).

³Ce groupe est un groupe de torsion qui représente en quelque sorte l'obstruction à la détermination du groupe $E(\mathbf{Q})$: si m est un entier $m \geq 1$, le sous-groupe de m -torsion $\sqcup_m(E)$ de $\sqcup(E)$ vit dans une suite exacte $0 \rightarrow E(\mathbf{Q})/mE(\mathbf{Q}) \rightarrow S_m(E) \rightarrow \sqcup_m(E) \rightarrow 0$, où le m -groupe de Selmer $S_m(E)$ est un groupe fini « effectivement calculable ».

THÉORÈME 0.4. — (i) Le groupe $E(\mathbf{Q})$ est un groupe abélien de type fini.

(ii) Le régulateur⁴ $R_\infty(E)$ est non nul.

CONJECTURE 0.5. — Le groupe $\sqcup(E)$ est un groupe fini. Plus généralement, si K est une extension finie de \mathbf{Q} , alors $\sqcup(E/K)$ est un groupe fini.

On note $r(E)$ le rang de $E(\mathbf{Q})$ et $r_\infty(E)$ l'ordre du zéro de $L(E, s)$ en $s = 1$.

CONJECTURE 0.6 (Birch et Swinnerton-Dyer). — On a $r_\infty(E) = r(E)$ et⁵

$$\lim_{s \rightarrow 1} (s-1)^{-r(E)} L(E, s) = \Omega_E^+ \cdot |\sqcup(E)| \cdot R_\infty(E) \cdot \prod_v m_v.$$

Les résultats concernant les conjectures 0.5 et 0.6 sont très partiels ; ce sont les suivants :

- $(\Omega_E^+)^{-1} L(E, 1)$ est un nombre rationnel (c'est une conséquence du théorème de Manin-Drinfeld) ;
- si $L(E, 1) = 0$, alors $\lim_{s \rightarrow 1} (s-1)^{-1} L(E, s)$ est un multiple rationnel de $\Omega_E^+ \cdot R_\infty(E)$ (cela suit du théorème de Gross-Zagier [85, 40]) ;
- si $r_\infty(E) \leq 1$, alors $\sqcup(E)$ est fini et $r(E) = r_\infty(E)$ (théorème de Kolyvagin ; la démonstration utilise le théorème de Gross-Zagier et la technique des dérivées de Kolyvagin introduite à cette occasion [106, 137]).

REMARQUE 0.7. — (i) En ce qui concerne le dernier point, on n'a pas de résultat dans l'autre sens : on ne sait pas démontrer que $r(E) = 0$ implique $r_\infty(E) = 0$ (de manière

⁴Ce régulateur est défini à partir de l'accouplement hauteur de Néron-Tate $\langle \cdot, \cdot \rangle_\infty$ sur l'espace vectoriel $\mathbf{R} \otimes_{\mathbf{Z}} E(\mathbf{Q})$. Si $y^2 = 4x^3 + ax + b$ est une équation de Weierstrass de E , la fonction $P = (x(P), y(P)) \mapsto h(P) = \frac{1}{2} \log d(P)$, où $d(P)$ est le dénominateur de $x(P)$, est presque quadratique, et la hauteur de Néron-Tate est l'unique forme bilinéaire symétrique $\langle \cdot, \cdot \rangle_\infty$ sur $E(\mathbf{Q})$ telle que $P \mapsto h(P) - \langle P, P \rangle_\infty$ soit bornée sur $E(\mathbf{Q})$. Tate a remarqué que l'on pouvait définir $\langle P, P \rangle_\infty$ comme la limite de la suite de terme général $4^{-n} h(2^n P)$. Par ailleurs, Néron [128, 109], a démontré que l'accouplement $\langle \cdot, \cdot \rangle_\infty$ pouvait s'exprimer comme une somme, sur toutes les places de \mathbf{Q} , de symboles locaux, les contributions aux places finies étant fournies par la théorie de l'intersection et celle à l'infini par la théorie du potentiel (fonctions de Green). Cette décomposition en somme de symboles locaux est fondamentale d'un point de vue théorique (elle est par exemple cruciale dans la démonstration du théorème de Gross-Zagier [85]), et sert de modèle pour la construction des hauteurs p -adiques (note 14). Si P_1, \dots, P_r , $r = r(E)$ sont des éléments de $E(\mathbf{Q})$ formant une base de $\mathbf{Q} \otimes_{\mathbf{Z}} E(\mathbf{Q})$ sur \mathbf{Q} , alors $R_\infty(E) = e^{-2} \det(\langle P_i, P_j \rangle_\infty)_{1 \leq i, j \leq r}$, où e est l'indice du sous-groupe engendré par P_1, \dots, P_r dans $E(\mathbf{Q})$. La non nullité de $R_\infty(E)$ suit de ce que l'accouplement $\langle \cdot, \cdot \rangle_\infty$ est non dégénéré car $\langle P, P \rangle_\infty > 0$ si P n'est pas de torsion comme on le constate en utilisant la formule de Tate.

⁵Les termes non encore définis dans cette formule sont :

- Le nombre m_v de composantes connexes de $E(\mathbf{Q}_v)$ si v est une place de \mathbf{Q} : si $v = \infty$, alors m_v est le nombre de composantes connexes de $E(\mathbf{R})$ au sens habituel, et si $v = p$ est un nombre premier, alors m_v est le nombre de composantes connexes sur \mathbf{F}_p de la réduction du modèle de Néron de E .
- La période réelle Ω_E^+ d'une différentielle de Néron (ou de Kähler?, [169, p.101]) ω_E ; on a donc $m_\infty \Omega_E^+ = |\int_{E(\mathbf{R})} \omega_E|$.

équivalente, on ne sait pas prouver que $L(E, 1) = 0$ entraîne l'existence d'un point d'ordre infini).

(ii) On ne dispose d'aucun résultat concernant le lien entre $r(E)$ et $r_\infty(E)$ ou la finitude de $\sqcup(E)$ dans le cas $r_\infty(E) \geq 2$.

(iii) Une des difficultés est que l'on ne connaît pas la valeur de $r(E)$ a priori ; on pense que si on prend une courbe E au hasard, alors $r(E) \leq 1$ avec une probabilité tendant vers 1 quand N_E tend vers $+\infty$, mais on connaît des courbes de rang ≥ 24 , et il y a tout lieu de croire que $r(E)$ n'est pas majoré.

(iv) Par contraste, on ne connaît pas de courbe elliptique E pour laquelle on peut prouver que⁶ $r_\infty(E) \geq 4$; le problème est qu'il est impossible de prouver qu'un réel est nul⁷ sauf si c'est un entier. C'est un peu dommage, car l'existence de telles courbes permettrait d'améliorer nettement les minoration effectives pour le nombre de classes des corps quadratiques imaginaires [81, 130].

0.3. La conjecture de Birch et Swinnerton-Dyer p -adique

En p -adique, le produit eulérien ci-dessus ne converge nulle part, mais on peut construire une fonction L p -adique⁸ à partir des valeurs en 1 de la fonction L complexe tordue⁹ par des caractères de Dirichlet de conducteur une puissance de p . Cette fonction L p -adique dépend d'un choix supplémentaire : on factorise le facteur d'Euler en p de $L(E, s)$ sous la forme $(1 - \alpha_1 p^{-s})(1 - \alpha_2 p^{-s})$ et on choisit $\alpha \in \{\alpha_1, \alpha_2\}$ vérifiant¹⁰ $v_p(\alpha) < 1$.

La construction de la fonction L p -adique $L_{p,\alpha}(E, s)$ de E associée à α repose sur la théorie des symboles modulaires¹¹ qui permet [117, 4, 111, 190] de démontrer :

⁶On peut, le cas échéant, vérifier que $r_\infty(E) \geq 3$ grâce au théorème de Gross-Zagier.

⁷En particulier, démontrer la conjecture 0.6 sous la forme faible « $r_\infty(E) = r(E)$ » ne fournit pas d'algorithme déterministe pour calculer $r(E)$ et $E(\mathbf{Q})$. Par contre, la conjecture 0.6 (même un peu affaiblie sous la forme $\lim_{s \rightarrow 1} (s-1)^{-r(E)} L(E, s) = n \cdot \Omega_E^+ \cdot R_\infty(E)$, avec n entier ≥ 1) fournit un tel algorithme, le point étant que plus $R_\infty(E)$ est petit et plus les générateurs de $E(\mathbf{Q})$ sont faciles à trouver.

⁸Il faut probablement supposer $p \neq 2$ ou $p \geq 5$ dans certains des énoncés qui suivent au niveau de ce qui est connu.

⁹Si χ est un tel caractère, on note $L(E, \chi, s)$ la fonction L de E tordue par χ ; elle est définie par la série de Dirichlet $L(E, \chi, s) = \sum_{n=1}^{+\infty} \chi(n) a_n n^{-s}$. Si le conducteur de χ n'est pas premier au conducteur de E , la fonction $L(E, \chi, s)$ n'est pas forcément primitive ; il peut manquer des facteurs d'Euler en les nombres premiers divisant N_E .

¹⁰Ce n'est pas toujours possible : si E a réduction additive en p , alors $\alpha_1 = \alpha_2 = 0$ et on ne sait pas construire de fonction L p -adique dans ce cas (sauf si la courbe acquiert bonne réduction sur une extension abélienne de \mathbf{Q} (cf. [62])). Si E a réduction multiplicative, alors $\{\alpha_1, \alpha_2\} = \{a_p, 0\}$, et on peut prendre $\alpha = a_p \in \{\pm 1\}$. Si E a bonne réduction, il y a deux cas de figure possibles : si $v_p(a_p) > 0$ (i.e. si E a bonne réduction supersingulière), alors il y a deux choix possibles pour α puisque $v_p(\alpha_{p,1}) = v_p(\alpha_{p,2}) = 1/2$, et si $v_p(a_p) = 0$ (bonne réduction ordinaire), alors une seule des deux racines $\alpha_{p,1}, \alpha_{p,2}$ est de valuation < 1 (de valuation 0), alors que l'autre est de valuation 1 (voir [149] et la rem. 4.12 pour ce dernier cas).

¹¹Panchishkin [132] a récemment trouvé une définition alternative de cette fonction qui colle nettement plus à la construction que l'on obtient en utilisant le système d'Euler de Kato.

THÉORÈME 0.8. — Soient Ω_E^+ et Ω_E^- les périodes réelles et imaginaires pures de ω_E . Alors

(i) si χ est un caractère de Dirichlet, $L(E, \chi, 1) \in \overline{\mathbf{Q}} \cdot \Omega_E^{\chi(-1)}$;

(ii) il existe une (unique) distribution $\mu_{E,\alpha}$ d'ordre $v_p(\alpha)$ sur \mathbf{Z}_p , telle que l'on ait

$$\int_{p\mathbf{Z}_p} f(x)\mu_{E,\alpha}(x) = \alpha^{-1} \int_{\mathbf{Z}_p} f(px)\mu_{E,\alpha}(x)$$

quelle que soit f localement analytique sur $p\mathbf{Z}_p$, et

$$\int_{\mathbf{Z}_p} \mu_{E,\alpha}(x) = (1 - \alpha^{-1})^b \frac{L(E, 1)}{\Omega_E^+} \quad \text{et} \quad \int_{\mathbf{Z}_p} \chi(x)\mu_{E,\alpha}(x) = p^n \alpha^{-n} \frac{L(E, \chi^{-1}, 1)}{G(\chi^{-1}) \cdot \Omega_E^{\chi(-1)}}$$

si $n \geq 1$ et χ est un caractère de Dirichlet¹² de conducteur p^n , et $b = 0$ (resp. $b = 1$) si E a mauvaise réduction multiplicative (resp. bonne réduction).

DÉFINITION 0.9. — La fonction L p -adique $s \mapsto L_{p,\alpha}(E, s)$ de E associée à α est la fonction définie, pour $s \in \mathbf{Z}_p^*$, par la formule

$$L_{p,\alpha}(E, s) = \int_{\mathbf{Z}_p^*} \langle x \rangle^{s-1} \mu_{E,\alpha}, \quad \text{avec} \quad \langle x \rangle^{s-1} = \exp((s-1) \log x).$$

REMARQUE 0.10. — On déduit du théorème 0.8 la formule $L_{p,\alpha}(E, 1) = (1 - \alpha^{-1})^{b+1} \frac{L(E, 1)}{\Omega_E^+}$.

En particulier, si $\alpha = 1$, alors la fonction $L_{p,\alpha}(E, s)$ a un zéro supplémentaire en $s = 1$.

La courbe E a alors réduction multiplicative déployée et le théorème d'uniformisation de Tate [184] nous fournit $q(E) \in \mathbf{Q}_p^*$, de valuation non nulle, tel que E soit isomorphe, en tant qu'espace analytique rigide, au quotient de \mathbf{G}_m par le groupe engendré par $q(E)$. Ceci nous permet de définir l'invariant \mathcal{L} de E par la formule $\mathcal{L}_p(E) = \frac{\log q(E)}{v_p(q(E))}$.

THÉORÈME 0.11. — Si $\alpha = 1$, alors¹³ $L'_{p,\alpha}(E, 1) = \mathcal{L}_p(E) \cdot \frac{L(E, 1)}{\Omega_E^+}$.

On sait définir un¹⁴ accouplement hauteur p -adique sur l'espace vectoriel $\mathbf{Q}_p \otimes E(\mathbf{Q})$, mais celui-ci dépend du choix d'un scindage de la filtration de Hodge sur $\mathbf{Q}_p \otimes_{\mathbf{Q}} H_{\text{dR}}^1(E)$

¹²On considère χ comme une fonction localement constante sur \mathbf{Z}_p , nulle sur $p\mathbf{Z}_p$, et on note $G(\chi^{-1}) = \sum_{a \in (\mathbf{Z}/p^n\mathbf{Z})^*} \zeta_p^a \chi^{-1}(a)$ la somme de Gauss de χ^{-1} .

¹³Ce théorème (cas particulier de la conjecture de Mazur-Tate-Teitelbaum) a été démontré par Greenberg et Stevens [84] en utilisant les familles de formes modulaires de Hida.

¹⁴On a trois définitions de cette hauteur p -adique : deux algébriques, l'une via la théorie d'Iwasawa [164, 138], l'autre via la cohomologie galoisienne [125, 138, 140], et une analytique utilisant l'intégration p -adique [46, 52, 195] ou les fonctions thêta p -adiques [118, 119, 129]. Il n'est pas évident a priori que ces définitions coïncident, mais c'est le cas ([138] pour les deux constructions algébriques et [164, 24, 94] pour la comparaison « algébrique-analytique »), ce qui permet de vérifier que les résultats que l'on obtient du côté de la théorie d'Iwasawa sont compatibles avec la conjecture de Birch et Swinnerton-Dyer via la conjecture principale. La définition analytique est obtenue en p -adifiant la définition de la hauteur de Néron-Tate (note 4) comme somme de symboles locaux (cf. [134] pour une définition à la Tate), la contribution en un premier $\ell \neq p$ restant la même, alors que la contribution à l'infini est transférée en p par l'intermédiaire d'une fonction de Green p -adique. Si $\partial = \frac{d}{\omega_E}$ est l'opérateur différentiel invariant par translation sur E , dual de $\omega_E = \frac{dx}{y}$, si v est une place de \mathbf{Q} , et si D est un diviseur de E , la fonction de Green de D en v s'obtient, par translation, à partir de la fonction de Green G_v de 0, qui est une fonction

(i.e. d'un supplémentaire de $\mathbf{Q}_p \cdot \omega_E$). On note $\langle \cdot, \cdot \rangle_{p,\alpha}$ l'accouplement correspondant à la droite de $\mathbf{Q}_p \otimes_{\mathbf{Q}} H_{\text{dR}}^1(E)$ sur laquelle Frobenius¹⁵ agit par multiplication par α , et $R_{p,\alpha}(E)$ le régulateur correspondant.

CONJECTURE 0.12 (Mazur-Tate-Teitelbaum). — (i) *L'ordre du zéro en $s = 1$ de la fonction $L_{p,\alpha}(E, s)$ est $\geq r(E)$ et on a*

$$\lim_{s \rightarrow 1} (s-1)^{-r(E)} L_{p,\alpha}(E, s) = (1 - \alpha^{-1})^{b+1} \cdot |\sqcup(E)| \cdot R_{p,\alpha}(E) \cdot \prod_v m_v.$$

(ii) *Si $\alpha = 1$, l'ordre du zéro en $s = 1$ de la fonction $L_{p,\alpha}(E, s)$ est $\geq r(E) + 1$ et on a*

$$\lim_{s \rightarrow 1} (s-1)^{-r(E)-1} L_{p,\alpha}(E, s) = \mathcal{L}_p(E) \cdot |\sqcup(E)| \cdot R_{p,\alpha}(E) \cdot \prod_v m_v.$$

REMARQUE 0.13. — (i) La non nullité de $\mathcal{L}_p(E)$ (conjecturée par Manin) a été démontrée par Barré, Diaz, Gramain et Philibert [5, 191] par des techniques de transcendance.

(ii) Contrairement au cas archimédien, on ne sait pas montrer que l'accouplement *hauteur p -adique* est non dégénéré¹⁶, et démontrer la non nullité de $R_{p,\alpha}(E)$ semble être du même ordre de difficulté que prouver la conjecture de Leopoldt¹⁷...

analytique réelle si $v = \infty$, localement analytique si v est finie, solution d'une équation différentielle du type $\partial^2 G_v = x + u$, avec $u \in \overline{\mathbf{Q}}_v$. Si $v = \infty$, le choix de u est imposé par la périodicité de G_v (vue comme fonction sur \mathbf{C}) par rapport au réseau des périodes de ω_E (u s'exprime alors en termes de la série d'Eisenstein non holomorphe E_2^*). Par contre, si $v = p$ est une place finie, on peut choisir u , ce qui revient à choisir un scindage de la filtration de Hodge, totalement arbitrairement. C'est pour cela que « la » hauteur p -adique dépend du choix d'un tel scindage.

¹⁵L'existence de α implique que E a bonne réduction ou est semi-stable, et donc que $\mathbf{Q}_p \otimes_{\mathbf{Q}} H_{\text{dR}}^1(E)$ est naturellement isomorphe à la cohomologie (log)-cristalline de E , ce qui permet de le munir d'une action du Frobenius φ .

¹⁶En fait, on ne sait même pas démontrer qu'il n'est pas identiquement nul sauf dans le cas de multiplication complexe [21].

¹⁷De manière générale, montrer qu'une fonction L p -adique ne s'annule pas en un entier est un problème difficile (alors que le problème analogue pour une fonction L complexe est en général très facile), et les seuls résultats généraux dont on dispose, pour le moment, reposent sur des techniques de transcendance. La conjecture de Leopoldt en est un exemple particulièrement frustrant. Si K est un corps de nombres, la formule analytique du nombre de classes est aussi équivalente à l'existence d'un pôle simple en $s = 1$ pour la fonction zêta de Dedekind ζ_K de K de résidu

$$\lim_{s \rightarrow 1} (s-1)\zeta_K(s) = \frac{2^{r_1} (2\pi)^{r_2} \cdot |\text{Pic}(\mathcal{O}_K)| \cdot R_{\infty}(K)}{\sqrt{|D_K|}},$$

où D_K est le discriminant de K . Si K est totalement réel (i.e. $r_2 = 0$ et $r_1 = [K : \mathbf{Q}]$), on sait lui associer [167, 6, 34, 66] une fonction zêta p -adique $\zeta_{K,p}$ et on dispose [50] de la formule analytique p -adique du nombre de classes :

$$\lim_{s \rightarrow 1} (s-1)\zeta_{K,p}(s) = \frac{2^{[K:\mathbf{Q}]} \cdot |\text{Pic}(\mathcal{O}_K)| \cdot R_p(K)}{\sqrt{|D_K|}} \cdot \prod_{\mathfrak{p}|p} (1 - (N\mathfrak{p})^{-1}),$$

où le régulateur p -adique $R_p(K)$ est défini par la même formule que $R_{\infty}(K)$, mais en remplaçant le logarithme complexe par le logarithme p -adique. Malheureusement, on ne sait prouver que $R_p(K) \neq 0$

En dehors du théorème 0.11 ci-dessus, on dispose d'un analogue p -adique du théorème de Gross-Zagier :

THÉORÈME 0.14. — Si $r_\infty(\mathbf{E}) = 1$ et \mathbf{E} a bonne réduction ordinaire¹⁸, alors

$$L'_{p,\alpha}(\mathbf{E}, 1) = (1 - \alpha^{-1})^2 \cdot R_{p,\alpha}(\mathbf{E}) \cdot \left(R_\infty(\mathbf{E})^{-1} \frac{L'(\mathbf{E}, 1)}{\Omega_{\mathbf{E}}^+} \right).$$

Finalement, le résultat le plus spectaculaire, pour lequel on n'a pas d'équivalent en complexe, est le théorème suivant de Kato¹⁹ [101] auquel le reste du texte va être consacré.

THÉORÈME 0.15. — L 'ordre du zéro de $L_{p,\alpha}(\mathbf{E}, s)$ en $s = 1$ est $\geq r(\mathbf{E})$ et même $\geq r(\mathbf{E}) + 1$ si $\alpha = 1$. De plus, si cette inégalité est une égalité, alors la p -partie $\sqcup_{p^\infty}(\mathbf{E})$ de $\sqcup(\mathbf{E})$ est finie et $R_{p,\alpha}(\mathbf{E}) \neq 0$.

REMARQUE 0.16. — (i) En partant de courbes de rang ≥ 24 , cela nous fournit des fonctions L p -adiques ayant un zéro d'ordre ≥ 24 en $s = 1$.

(ii) Vérifier la non nullité d'un nombre p -adique peut (contrairement à sa nullité) se faire en calculant ce nombre avec une précision suffisante. Le th. 0.15 nous fournit donc, en principe, un critère analytique pour vérifier que la p -partie du groupe de Tate-Shafarevitch est finie même si le rang est très grand²⁰.

0.4. Des courbes elliptiques aux formes modulaires

Il faut être bien conscient de ce que l'on serait totalement démuné pour étudier la fonction L d'une courbe elliptique si l'on ne disposait que de sa définition comme produit eulérien²¹. Tous les beaux théorèmes évoqués ci-dessus s'appuient de manière cruciale sur le fait que la conjecture de Taniyama-Weil est maintenant un théorème grâce aux travaux de Wiles [194, 131, 168] et ceux de Breuil, Conrad, Diamond et Taylor [29, 69]. Ceci se traduit de la manière suivante :

THÉORÈME 0.17. — Si \mathbf{E} est une courbe elliptique définie sur \mathbf{Q} , de conducteur $N_{\mathbf{E}}$, et si $L(\mathbf{E}, s) = \sum_{n=1}^{+\infty} a_n n^{-s}$, alors $f_{\mathbf{E}}(\tau) = \sum_{n=1}^{+\infty} a_n q^n$ (avec $q = e^{2i\pi\tau}$) est une forme modulaire primitive de poids 2 pour $\Gamma_0(N_{\mathbf{E}})$.

COROLLAIRE 0.18. — La fonction $L(\mathbf{E}, s)$ possède un prolongement analytique à tout le plan complexe. De plus, il existe $\varepsilon_{\mathbf{E}} \in \{\pm 1\}$ tel que la fonction $\Lambda(\mathbf{E}, s) = \frac{\Gamma(s)}{(2\pi)^s} N_{\mathbf{E}}^{s/2} L(\mathbf{E}, s)$ vérifie l'équation fonctionnelle $\Lambda(\mathbf{E}, s) = \varepsilon_{\mathbf{E}} \cdot \Lambda(\mathbf{E}, 2 - s)$.

(conjecture de Leopoldt) que si K est une extension abélienne de \mathbf{Q} , en utilisant le fait que le régulateur est alors un produit de formes linéaires de logarithmes de nombres algébriques que l'on peut étudier par une variante p -adique [30] de la méthode de Baker.

¹⁸Ce théorème est dû à Perrin-Riou [135] ; j'ignore s'il a été étendu entre-temps au cas supersingulier.

¹⁹Si \mathbf{E} est de type CM, ce résultat remonte à Rubin [155], au moins dans le cas ordinaire.

²⁰Cf. [146] pour une utilisation de ce point.

²¹On ne saurait pas que la fonction L est définie en $s = 1$ et on ne saurait pas construire la fonction L p -adique associée à \mathbf{E} .

REMARQUE 0.19. — (i) Si $\varepsilon_E = -1$, la fonction $L(E, s)$ s'annule en $s = 1$. La conjecture de Birch et Swinnerton-Dyer prédit alors que $E(\mathbf{Q})$ est un groupe infini. Nekovář [127] a démontré que, si $\varprojlim_p(E)$ est fini pour au moins un premier p en lequel E a bonne réduction ordinaire, alors $E(\mathbf{Q})$ est effectivement infini²².

(ii) Comme Γ a des pôles simples aux entiers négatifs, l'équation fonctionnelle ci-dessus et la non annulation de $L(E, s)$ pour $\operatorname{Re}(s) \geq 2$ impliquent que $L(E, s)$ a des zéros simples en tous les entiers négatifs.

THÉORÈME 0.20. — *Si E est une courbe elliptique définie sur \mathbf{Q} de conducteur N_E , il existe un morphisme surjectif $\pi_E : \mathcal{H} \rightarrow E$ de surfaces de Riemann, se factorisant à travers $\Gamma_0(N_E) \backslash \mathcal{H}$, tel que $\pi_E(i\infty) = 0$. De plus*

- (i) $\pi_E^* \omega_E = c_E f_E$, avec $c_E \in \mathbf{Z}$;
- (ii) si $P \in \mathbf{P}^1(\mathbf{Q})$ est une pointe de \mathcal{H} , alors $\pi_E(P) \in E(\mathbf{Q})_{\text{tors}}$.

REMARQUE 0.21. — (i) Le théorème 0.20 se déduit²³ du théorème 0.17, mais n'en est pas une conséquence formelle. Il résulte de l'existence d'une courbe elliptique E' quotient de $X_0(N_E)$ dont la fonction L est égale à $L(f_E, s)$ (travaux d'Eichler-Shimura [70, 170, 171, 172, 174] et Carayol [31]) et de ce que deux courbes elliptiques définies sur \mathbf{Q} ayant même fonction L sont isogènes. Ce dernier fait a été démontré par Serre [166] dans le cas où $j(E) \notin \mathbf{Z}$, et en toute généralité par Faltings [71, 65] comme cas particulier de la conjecture de Tate. Une autre démonstration utilisant des techniques de transcendance a été obtenue par Chudnovsky [38] (voir aussi [28] et [35]).

(ii) Le théorème 0.17 est un cas particulier de la correspondance (conjecturale) de Langlands entre motifs et formes automorphes arithmétiques. Il semble beaucoup plus difficile de généraliser le théorème 0.20 en une conjecture fournissant une description automorphe d'un motif quelconque. Le cas le plus simple posant problème est celui d'une courbe elliptique E , définie sur un corps quadratique réel, que l'on pourrait s'attendre à voir apparaître dans le H^1 d'une variété de Shimura convenable, mais Blasius a montré que ce n'est pas le cas si E a bonne réduction partout et n'est pas isogène (sur $\overline{\mathbf{Q}}$) à sa conjuguée.

(iii) Les motifs associés aux formes modulaires ont une structure extrêmement riche. On peut jouer sur les aspects « variété de modules » ou « quotient du demi-plan de Poincaré » de $X_0(N)$ pour construire de manière systématique des objets (symboles modulaires, points de Heegner, unités de Siegel...) permettant d'étudier leur arithmétique. Par contraste, on ne sait pas dire grand-chose directement de l'arithmétique d'une courbe elliptique. Le

²²Il a en fait démontré que la parité du corang du groupe de Selmer est celle prédite par la conjecture de Birch et Swinnerton-Dyer. Dans le cas CM, la non finitude du groupe de Selmer si $\varepsilon_E = -1$ remonte à Greenberg [83] dans le cas ordinaire et à Rubin [154] dans le cas supersingulier, et la parité du corang du groupe de Selmer à Guo [86].

²³Sauf le (ii), pour lequel on a en plus besoin du théorème de Manin-Drinfeld.

théorème 0.20 qui établit un isomorphisme entre la courbe elliptique E et le motif associé à f_E est donc crucial pour étudier l'arithmétique de E . De fait, les théorèmes 0.11 et 0.15 sont obtenus via le théorème 0.20, comme cas particuliers de théorèmes généraux (cf. th. 0.22, 0.23 et 4.17) sur les motifs associés aux formes modulaires de poids quelconque ; ils correspondent au cas des formes modulaires de poids 2 à coefficients rationnels.

0.5. Formes modulaires de poids quelconque

Soit $f = \sum_{n=1}^{+\infty} a_n q^n$ une forme primitive de niveau N , poids k et caractère ε . Soit $\mathbf{Q}(f) = \mathbf{Q}(a_1, \dots, a_n, \dots)$; c'est une extension finie de \mathbf{Q} contenue dans $\overline{\mathbf{Q}}$. Soit p un nombre premier et soit $\mathbf{Q}_p(f) = \mathbf{Q}_p(a_1, \dots, a_n, \dots)$; c'est une extension finie de \mathbf{Q}_p dont on note O_p l'anneau des entiers. D'après Deligne [63], on peut associer à f une $\mathbf{Q}_p(f)$ -représentation V_f de $\mathcal{G}_{\mathbf{Q}}$, de dimension 2, non ramifiée en dehors de Np , caractérisée par le fait que le déterminant de $1 - \text{Fr}_\ell^{-1}X$ agissant sur V_f est $1 - a_\ell X + \ell^{k-1}\varepsilon(\ell)X^2$, où $\text{Fr}_\ell \in \mathcal{G}_{\mathbf{Q}}$ est un frobenius arithmétique en ℓ si $\ell \nmid Np$.

Si T_f est un O_p -réseau de V_f stable²⁴ par $\mathcal{G}_{\mathbf{Q}}$, si $j \in \mathbf{Z}$, et si K est une extension finie de \mathbf{Q} , on définit²⁵ le groupe de Selmer $\text{Sel}(T_f(j)/K)$ qui est un O_p -module isomorphe à $(\mathbf{Q}_p(f)/O_p)^r \oplus X$, où X est fini et $r = \text{corg}_{O_p}(\text{Sel}(T_f(j))/K)$ est un entier.

Si f est la forme modulaire associée à une courbe elliptique E définie sur \mathbf{Q} et si $j = 1$, on peut prendre pour $T_f(1)$ le module de Tate $T_p(E)$ de E (i.e. le \mathbf{Z}_p -module des familles $(u_n)_{n \in \mathbf{N}}$ d'éléments de $E(\overline{\mathbf{Q}})$ vérifiant $u_0 = 0$ et $pu_{n+1} = u_n$). Si K est une extension finie de \mathbf{Q} , alors $\text{Sel}(T_p(E)/K)$ est la p -partie $\text{Sel}_{p^\infty}(E/K)$ du groupe de Selmer de E sur K , et on a la suite exacte

$$0 \longrightarrow E(K) \otimes (\mathbf{Q}_p/\mathbf{Z}_p) \longrightarrow \text{Sel}(T_p(E)/K) \longrightarrow \sqcup_{p^\infty}(E/K) \longrightarrow 0.$$

En particulier, $\text{corg}(\text{Sel}(T_p(E)/K)) = r(E/K) + \text{corg}(\sqcup_{p^\infty}(E/K))$ et donc si $\sqcup_{p^\infty}(E/K)$ est fini si et seulement si $\text{corg}(\text{Sel}(T_p(E)/K)) = r(E/K)$.

THÉORÈME 0.22. — *Soit f une forme primitive de poids $k \geq 2$, soit $j \in \{1, \dots, k-1\}$ un entier, et soit T_f un réseau de V_f stable par $\mathcal{G}_{\mathbf{Q}}$. Si K une extension finie abélienne*

²⁴Un tel réseau existe toujours car $\mathcal{G}_{\mathbf{Q}}$ est un groupe compact.

²⁵Si v est une place de K , on définit le sous-groupe $H_f^1(\mathcal{G}_{K_v}, V_f(j))$ de $H^1(\mathcal{G}_{K_v}, V_f(j))$ comme étant le noyau de l'application de $H^1(\mathcal{G}_{K_v}, V_f(j))$ dans $H^1(I_{K_v}, V_f(j))$ (resp. dans $H^1(\mathcal{G}_{K_v}, \mathbf{B}_{\text{cris}} \otimes_{\mathbf{Q}_p} V_f(j))$) si v est une place finie ne divisant pas p et I_{K_v} est le sous-groupe d'inertie de \mathcal{G}_{K_v} (resp. si v divise p et \mathbf{B}_{cris} est un des anneaux de Fontaine). On définit alors $H_f^1(\mathcal{G}_{K_v}, V_f(j)/T_f(j))$ comme le quotient de $H^1(\mathcal{G}_{K_v}, V_f(j)/T_f(j))$ par l'image de $H_f^1(\mathcal{G}_{K_v}, V_f(j))$ et on a

$$\text{Sel}(T_f(j)/K) = \text{Ker}(H^1(\mathcal{G}_K, V_f(j)/T_f(j)) \longrightarrow \oplus_v H_f^1(\mathcal{G}_{K_v}, V_f(j)/T_f(j))).$$

de \mathbf{Q} , si $\eta : \text{Gal}(\mathbf{K}/\mathbf{Q}) \rightarrow \overline{\mathbf{Q}}^*$ est un caractère de Dirichlet, et si $L(f, \eta, j) \neq 0$, le groupe $\text{Sel}(\text{T}_f(j)/\mathbf{K})^{(\eta)}$ est un groupe fini, nul si p est assez grand²⁶.

Soit α une racine du polynôme $X^2 - a_p X + \varepsilon(p)p^{k-1}$ avec $v_p(\alpha) < k - 1$. Comme pour les courbes elliptiques, on peut utiliser la théorie des symboles modulaires pour construire une fonction L p -adique $L_{p,\alpha}(f, \eta\chi_{\text{cycl}}^j, s)$, où η est un caractère d'ordre fini et j un entier, interpolant les valeurs de la fonction L complexe en les caractères critiques (i.e. les $L(f, \chi, j)$, où χ est un caractère de Dirichlet et j un entier, avec $1 \leq j \leq k - 1$). On a alors le théorème suivant [101] qui contient le théorème 0.15 comme cas particulier (à la non annulation du régulateur près).

THÉORÈME 0.23. — *L'ordre du zéro en $s = 0$ de la fonction $L_{p,\alpha}(f, \eta\chi_{\text{cycl}}^j, s)$ est $\geq \text{corg}(\text{Sel}(\text{T}_f(j)/\mathbf{K})^{(\eta)})$ et même $\geq \text{corg}(\text{Sel}(\text{T}_f(j)/\mathbf{K})^{(\eta)}) + 1$, si $j = \frac{k}{2}$ et $\alpha = p^{j-1}\eta(p)$.*

REMARQUE 0.24. — L'ordre d'annulation d'une fonction L complexe peut s'aborder sous l'angle de la théorie analytique des nombres. De ce point de vue, on dispose des résultats suivants :

(i) La condition $L(f, \eta, j) \neq 0$ est automatique si $j \neq \frac{k}{2}, \frac{k-1}{2}$: si $j > \frac{k+1}{2}$ cela suit de ce que le produit eulérien définissant $L(f, \eta, j)$ est absolument convergent grâce à la conjecture de Ramanujan-Peterson ; le cas $j = \frac{k+1}{2}$ (bord de la bande critique) est un théorème de Jacquet et Shalika [96] et les autres cas s'en déduisent en utilisant l'équation fonctionnelle²⁷.

(ii) Si S est un ensemble fini de nombres premiers et \mathbf{Q}_S^{ab} est l'extension abélienne maximale de \mathbf{Q} non ramifiée en dehors de S , Rohrlich [153] a démontré que

$$\{\eta : \text{Gal}(\mathbf{Q}_S^{\text{ab}}/\mathbf{Q}) \rightarrow \mathbf{C}^*, L(f, \eta, k/2) = 0\}$$

est fini. Le (i) du théorème 0.22 permet donc de montrer que le groupe $E(\mathbf{Q}_S^{\text{ab}})$ est de rang fini ; plus généralement, le groupe $A(\mathbf{Q}_S^{\text{ab}})$ est de rang fini si A est une variété abélienne quotient de la jacobienne d'une courbe modulaire.

(iii) En comparant l'inégalité du théorème 0.22 avec ce que prédisent les conjectures de Beilinson, on obtient l'inégalité *conjecturale*

$$\text{ord}_{s=j} L(f, \eta, s) \leq \text{ord}_{s=0} L_{p,\alpha}(f, \eta\chi_{\text{cycl}}^j, s)$$

qui ne fait intervenir que des objets définis analytiquement à partir de la forme modulaire. Peut-on démontrer cette inégalité par des techniques de théorie analytique des nombres ? (Le problème ne se pose que pour $j = \frac{k}{2}$ d'après ce qui précède.)

²⁶Il faut que l'image de $\mathcal{G}_{\mathbf{Q}}$ dans la représentation modulo p soit assez grosse, ce qui est le cas pour p grand, si f n'est pas de type CM, d'après un théorème de Ribet [150, 152]. Le cas CM se déduit des résultats de Rubin [157].

²⁷Les zéros potentiels pour $j = \frac{k-1}{2}$ viennent de ce que $L(f, \eta, s)$ n'étant pas forcément primitive, certains des facteurs d'Euler manquants peuvent avoir un zéro en $s = \frac{k-1}{2}$.

(iv) En utilisant les « formules explicites », on peut borner [123] l'ordre du zéro en $s = \frac{k}{2}$ de la fonction $L(f, \eta, s)$ en fonction de k et N . Peut-on obtenir des bornes similaires pour l'ordre du zéro de $L_{p,\alpha}$? On pourrait peut-être en tirer la finitude du groupe de Tate-Shafarevitch (ou au moins la finitude de sa p -partie²⁸ pour tout p) pour des courbes elliptiques de grand rang et « petit » conducteur comme celles obtenues [123] en maximisant le nombre de points modulo p pour beaucoup de petits p .

0.6. Survol de la démonstration

0.6.1. *Systèmes d'Euler et bornes pour les groupes de Selmer.* Grâce aux travaux de Kolyvagin [106], étendus par Kato [99], Perrin-Riou [142] et Rubin [158] (voir aussi [115]), on dispose d'une machine extrêmement puissante (dérivées de Kolyvagin) pour borner le groupe de Selmer d'une représentation p -adique V de $\mathcal{G}_{\mathbf{Q}}$. Pour fonctionner, la machine nécessite la construction d'un système d'Euler pour $V^*(1)$, c'est-à-dire la construction d'une famille de classes de cohomologies $c_M \in H^1(\mathcal{G}_{\mathbf{Q}(\zeta_M)}, T^*(1))$, T^* réseau de V^* et M parcourant²⁹ les entiers ≥ 1 , vérifiant les relations

$$\text{cor}_{\mathbf{Q}(\zeta_{M\ell})/\mathbf{Q}(\zeta_M)} c_{M\ell} = \begin{cases} c_M & \text{si } \ell | M, \\ c_M \star P_{\ell}(\text{Fr}_{\ell}^{-1}) & \text{si } \ell \nmid M \text{ et } V \text{ est non ramifiée en } \ell, \end{cases}$$

où $P_{\ell}(\ell^{-s})$ est le facteur d'Euler³⁰ en ℓ de la fonction L attachée à V . On obtient alors une borne pour l'ordre du groupe de Selmer de V en termes de l'indice de c_1 dans $H^1(\mathcal{G}_{\mathbf{Q}}, T^*(1))$. Cette borne ne représente qu'une partie du travail pour obtenir des énoncés du type de la conjecture de Birch et Swinnerton-Dyer car il reste à relier l'indice de c_1 dans $H^1(\mathcal{G}_{\mathbf{Q}}, T^*(1))$ à une valeur spéciale de fonction L .

0.6.2. *Conjectures de Beilinson et systèmes d'Euler.* La construction de systèmes d'Euler est une activité totalement artisanale. Si V est la réalisation p -adique d'un motif \mathbb{M} , on peut essayer de construire « géométriquement » des éléments des groupes $H^1(\mathbf{Q}(\zeta_M), \mathbb{M}^*(1))$ d'extensions du motif trivial par le motif $\mathbb{M}^*(1)$ et considérer leurs réalisations p -adiques. L'existence ou la non existence de telles extensions est prédite par la conjecture de Beilinson [7]. D'après cette conjecture, si $L(\mathbb{M}, \eta, s)$ a un zéro simple en $s = 0$ pour tout caractère de Dirichlet η , il existe une famille $c_M \in H^1(\mathbf{Q}(\zeta_M), \mathbb{M}^*(1))$, $M \geq 1$ et une

²⁸Profitons de l'occasion pour signaler que, contrairement à ce que prétend une rumeur persistante, on ne semble pas savoir démontrer que $\varprojlim_{p \infty} (E/K)$ est divisible pour presque tout p . Cette rumeur provient d'une note de bas de page [32, p. 240] dans laquelle Cassels écrit que Shafarevitch lui a dit savoir démontrer le résultat en question, mais Colliot-Thélène m'a signalé que, dans [33, p. 277], Cassels écrit que Shafarevich lui a dit que la preuve prévue s'était heurtée à des difficultés imprévues...

²⁹Plus généralement, on peut remplacer \mathbf{Q} par une extension finie F et la tour des corps cyclotomiques par les extensions finies K de F contenues dans une extension abélienne suffisamment grande F_{∞} de F .

³⁰i.e. $P_{\ell}(X) = \det(1 - X \cdot \text{Fr}_{\ell}^{-1}|_V)$.

constante C , telle que l'on ait

$$L'(\mathbb{M}, \eta, 0) = C \cdot \sum_{\sigma \in \text{Gal}(\mathbf{Q}(\zeta_M)/\mathbf{Q})} \eta(\sigma) \cdot \text{reg}_\infty(c_M \star \sigma),$$

pour tout $M \geq 1$ et tout caractère η de conducteur M , où reg_∞ est l'application régulateur. De tels éléments sont (conjecturalement) uniquement déterminés à addition près d'éléments de torsion et, dans les cas favorables³¹, leurs réalisations p -adiques forment un système d'Euler³².

0.6.3. *Torsion à la Soulé et lois de réciprocité explicites.* On dispose d'une méthode (torsion à la Soulé) pour obtenir un système d'Euler pour $V^*(k)$, $k \in \mathbf{Z}$, à partir d'un système d'Euler pour $V^*(1)$. En d'autres termes, la construction d'un système d'Euler pour n'importe quel tordu de $V^*(1)$ permet de borner le groupe de Selmer de V .

Comme nous l'avons mentionné plus haut, la borne obtenue ne fait absolument pas intervenir de valeur spéciale de fonction L . Cependant, l'expérience montre que, si on est parti d'un système d'Euler relié aux valeurs spéciales de fonctions L pour un tordu de $V^*(1)$, alors les tordus du système d'Euler continuent³³ à être reliés aux valeurs spéciales de fonctions L en les points considérés. C'est le royaume des lois de réciprocité explicites [95, 193, 44, 27, 97, 98, 51, 102, 11, 36, 100, 161, 163, 78, 79, ...] qui constituent en général le point le plus technique des démonstrations.

0.6.4. *Le système d'Euler de Kato* ([101, §§ 1-5 et § 8]). Si E est une courbe elliptique définie sur \mathbf{Q} et \mathbb{M} est le motif $h_1(E)$ (dont la réalisation p -adique est $\mathbf{Q}_p \otimes T_p(E)$, où $T_p(E)$ est le module de Tate de E), alors $H^1(\mathbf{Q}(\zeta_M), \mathbb{M}^*(1)) = \mathbf{Q} \otimes E(\mathbf{Q}(\zeta_M))$, et on

³¹C'est le cas pour les motifs attachés aux formes modulaires. Dans le cas général, il y a une obstruction de nature locale à l'existence de systèmes compatibles de classes de cohomologies « géométriques » à valeurs dans un réseau (cf. [143, 144] pour les cas de « bonne réduction » et de « réduction semi-stable » et [13] pour le cas général). Les éléments prédits par la conjecture de Beilinson doivent donc faire intervenir des dénominateurs, et il semble intéressant d'essayer de comprendre ce que l'on peut en tirer (par exemple pour un motif dont tous les nombres de Hodge sont différents car, dans ce cas, la conjecture de Beilinson a l'air de suggérer l'apparition d'un nouveau système compatible « de rang 1 » en chacun des twists par un des poids de Hodge-Tate).

³²On peut voir le système c_M , $M \geq 1$, comme une incarnation algébrique de la fonction L du motif \mathbb{M} . Si $L(\mathbb{M}, \eta, s) \neq 0$, en $s = 0$, un zéro d'ordre $r \geq 2$, alors le groupe $H^1(\mathbf{Q}(\zeta_M), \mathbb{M}^*(1))$ est un $\mathbf{Q}[\text{Gal}(\mathbf{Q}(\zeta_M)/\mathbf{Q})]$ -module de rang r ; c'est le déterminant de ce groupe qui est relié à la fonction L via l'application régulateur, et l'existence d'un système d'Euler ne semble pas automatique; en tout cas, on n'a aucun exemple auquel se rattacher pour essayer de deviner ce qui est vrai.

³³Il y a de quoi rester un peu perplexe : on part d'un système compatible d'éléments motiviques pour l'extension cyclotomique, on choisit un nombre premier p , on prend la réalisation étale p -adique du système, on la tord par une représentation « indépendante de p », on fait entrer le résultat du côté galoisien dans le labyrinthe des anneaux de Fontaine, et il ressort du côté de Rham sous la forme d'un système compatible pour l'extension cyclotomique quasiment *indépendant de p* ! C'est d'autant plus remarquable qu'il n'y a rien qui puisse permettre de prédire, dans l'état actuel de notre compréhension, que les éléments que l'on va récupérer vont être définis sur $\overline{\mathbf{Q}}$...

est ramené à construire de manière systématique des points de E rationnels sur $\mathbf{Q}(\zeta_M)$. Malheureusement, la conjecture de Birch et Swinnerton-Dyer et le théorème de Rohrlich (cf. (ii) de la rem. 0.24) montrent que ceci n'est pas possible. Par contre, si $n \geq 1$, le groupe $H^1(\mathbf{Q}(\zeta_M), h_1(E)(n))$ est un $\mathbf{Q}[\text{Gal}(\mathbf{Q}(\zeta_M)/\mathbf{Q})]$ -module de rang 1 d'après la conjecture de Beilinson car $L(E, \eta, s)$ a un zéro simple³⁴ en $s = 1 - n$, quel que soit le caractère de Dirichlet η . On est donc ramené à construire explicitement les éléments de $H^1(\mathbf{Q}(\zeta_M), h_1(E)(n))$ dont l'existence est prédite par la conjecture de Beilinson et à les normaliser correctement pour en faire un système d'Euler.

Pour $n = 1$, le projet a été mené à bien³⁵ par Beilinson [7] qui a construit, en partant d'unités de Siegel³⁶ sur la courbe modulaire $Y(N)$ (avec $N =$ multiple du conducteur de E), des éléments dans le K_2 de la courbe³⁷ elliptique E dont les images par l'application régulateur font intervenir les dérivées premières en $s = 0$ des tordues de la fonction L de E par les caractères de Dirichlet.

La théorie de Kummer permet d'associer à une unité sur $Y(M)$ (et donc à une unité de Siegel) une classe de cohomologie dans $H_{\text{ét}}^1(Y(M), \mathbf{Z}_p(1))$. En faisant le cup-produit de deux de ces classes, on obtient une classe dans $H_{\text{ét}}^2(Y(M), \mathbf{Z}_p(2))$ qui n'est autre que l'image, au signe près, de l'élément de Beilinson par l'application classe de Chern. Kato a trouvé un moyen de normaliser³⁸ les unités de Siegel et les éléments de Beilinson pour

³⁴Ce zéro peut être d'ordre ≥ 1 si $n = 0$ et E a potentiellement réduction multiplicative en une place divisant le conducteur de η , la fonction $L(E, \eta, s)$ n'étant pas forcément primitive.

³⁵ Beilinson ne s'est pas arrêté en si bon chemin : comme confirmation de ses conjectures générales sur les valeurs spéciales de fonctions L de motifs, il a [8]

- construit, si $k \geq 2$, des éléments « d'Eisenstein » dans la K -théorie de la variété de Kuga-Sato $Y^{(k-2)}(M)$ (rappelons que $Y(M)$ étant un espace de module de courbes elliptiques, on dispose d'une courbe elliptique universelle \mathcal{E} au-dessus de $Y(M)$; la variété $Y^{(k-2)}(M)$ est alors obtenue à partir du produit de $k - 2$ copies de \mathcal{E} au-dessus de $Y(M)$) dans la cohomologie de laquelle sont découpés [159] les motifs associés aux formes modulaires de poids k ; pour $k = 2$, on récupère les unités de Siegel ;

- déduit de ces éléments d'Eisenstein, par cup-produit et image directe, des éléments dans la cohomologie motivique de la courbe modulaire ;

- calculé l'image de ces éléments par l'application régulateur et vérifié que le résultat faisait intervenir les dérivées des fonctions L de formes modulaires de poids 2 en $s = 2 - k$.

Par ailleurs, Scholl [67, 162] a construit, par cup-produit à partir des éléments d'Eisenstein de Beilinson, des éléments dans la cohomologie motivique de la variété de Kuga-Sato $Y^{(k-2)}(M)$ dont l'image par l'application régulateur fait intervenir les dérivées des fonctions L de formes modulaires de poids k en $s = -n$, $n \in \mathbf{N}$.

³⁶Ces unités, appelées aussi *unités modulaires*, sont définies comme des produits infinis (cf. [107] par exemple) ; on obtient de la sorte des fonctions modulaires sans zéro ni pôle sur le demi-plan de Poincaré dont le q -développement est à coefficients dans $\mathbf{Q}(\zeta_N)$, ce qui nous fournit des fonctions inversibles sur la courbe $Y(N)$ qui est une courbe algébrique affine définie sur $\mathbf{Q}(\zeta_N)$.

³⁷On commence par construire des éléments dans le K_2 de $Y(N)$ que l'on projette dans $K_2(E)$ en utilisant le morphisme $Y(N) \rightarrow E$ fourni par le théorème 0.20.

³⁸Scholl [161] en a fait de même avec les éléments de Beilinson dans la K -théorie de la variété de Kuga-Sato.

obtenir de la sorte un système cohérent de classes $c_M \in H_{\text{ét}}^2(Y(M), \mathbf{Z}_p(2))$, $M \geq 1$. En utilisant la technique de torsion à la Soulé (pour une extension de type \mathbf{GL}_2 au lieu de l'extension cyclotomique qui est de type \mathbf{GL}_1), cela lui permet de construire, si $k \geq 2$ et $j \in \mathbf{Z}$, un système cohérent de classes $c_M(k, j) \in H_{\text{ét}}^2(Y(M), W_k(j))$, $M \geq 1$, où W_k est le système local sur $Y(M)$ dans la cohomologie duquel on découpe les représentations p -adiques associées aux formes modulaires de poids k et niveau M .

Maintenant, la courbe $Y(M)$ étant affine, on a $H_{\text{ét}}^2(Y(M)_{\overline{\mathbf{Q}}}, W_k(j)) = 0$, ce qui fournit une application naturelle³⁹ de $H_{\text{ét}}^2(Y(M), W_k(j))$ dans $H^1(\mathcal{G}_{\mathbf{Q}(\zeta_M)}, H_{\text{ét}}^1(Y(M)_{\overline{\mathbf{Q}}}, W_k(j)))$. Si f est une forme primitive de poids k et de niveau divisant M , il n'y a plus qu'à projeter l'image de $c_M(k, j)$ sur la composante de $H_{\text{ét}}^1(Y(M)_{\overline{\mathbf{Q}}}, W_k(j))$ correspondant à f , pour obtenir le système d'Euler $c_M(f, j) \in H^1(\mathcal{G}_{\mathbf{Q}(\zeta_M)}, V_f(j))$, $M \geq 1$ que l'on cherchait à construire⁴⁰ (ce système d'Euler permet de borner le groupe de Selmer de $V_f(j)^*(1) = V_{f^*}(k - j)$).

0.6.5. *La loi de réciprocité explicite de Kato* ([100] et [101, §§ 9-11]). Le point le plus délicat est de relier les classes de cohomologie construites dans le numéro précédent aux valeurs spéciales de fonctions L. Pour ce faire, on utilise deux applications *exponentielles duales* utilisant l'anneau de Fontaine \mathbb{B}_{dR}^+ pour deux corps différents : le corps $\overline{\mathbf{Q}}_p$, et le corps $\overline{\mathcal{K}}$, clôture algébrique du corps des fractions⁴¹ \mathcal{K} du complété de $\mathbf{Z}_p[[q]][[q^{-1}]$ pour la topologie p -adique. Si V est une représentation de de Rham de $\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}$, le cup-produit avec $\log \chi_{\text{cycl}}$ fournit un isomorphisme

$$H^0(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes V) \cong H^1(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes V),$$

et on définit \exp^* comme l'inverse de cet isomorphisme. Ce qui précède s'applique en particulier à la représentation $W_{k,j} = H_{\text{ét}}^1(Y(M)_{\overline{\mathbf{Q}}_p}, W_k(j))$. Par ailleurs, si $M \geq 1$, et $\mathcal{K}_M = \mathcal{K}[q^{1/M}, \zeta_M]$, on dispose, si $1 \leq j \leq k - 1$, d'un isomorphisme naturel

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes W_k(j)) \cong \mathcal{K}_M.$$

³⁹Gealy [80] a vérifié que, si $j \geq k$, les éléments de Kato et les éléments (cf. note 35) de Beilinson ($j = k$) et Scholl ($j \geq k$) ont même image dans $H^1(\mathcal{G}_{\mathbf{Q}(\zeta_M)}, H_{\text{ét}}^1(Y(M)_{\overline{\mathbf{Q}}}, W_k(j)))$, ce qui est une illustration du phénomène mentionné au n° 0.6.3. Scholl a démontré (cf. note 43) un résultat du même type en ce qui concerne les éléments de Beilinson.

⁴⁰Cette description « du » système d'Euler associé à f est un peu idéalisée. On est en fait forcé de faire un certain nombre de choix au cours de la construction, ce qui nous fournit toute une famille de systèmes d'Euler, mais aucun d'eux n'est optimal (i.e. les valeurs spéciales de fonctions L que l'on obtient via la loi de réciprocité explicite sont multipliées par des facteurs parasites). On peut construire un système d'Euler optimal [101, § 12] à partir de cette famille de systèmes d'Euler, mais c'est au prix de pas mal de complications techniques... En particulier, la normalisation de ce système d'Euler passe par la loi de réciprocité explicite.

⁴¹Ce corps apparaît naturellement comme complété du corps des fonctions de $Y(1)$ en la pointe $i\infty$. En associant son q -développement à une forme modulaire, cela permet de voir les formes modulaires comme des éléments de $\overline{\mathcal{K}}$.

En utilisant ces deux isomorphismes, l'injection naturelle de $\mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p)$ dans $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$, la localisation de $Y(\mathbf{M})_{\mathbf{Q}_p(\zeta_M)}$ à \mathcal{K}_M et l'isomorphisme⁴²

$$H^1(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes H_{\text{et}}^1(Y(\mathbf{M})_{\overline{\mathbf{Q}}_p}, W_k(j))) \cong H_{\text{et}}^2(Y(\mathbf{M})_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes W_k(j))$$

fourni par la suite spectrale de Hochschild-Serre, on obtient le diagramme suivant :

$$\begin{array}{ccc} & H^1(\mathcal{G}_{\mathbf{Q}(\zeta_M)}, W_{k,j}) & \\ & \downarrow & \\ H^0(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes W_{k,j}) & \xleftarrow[\sim]{\text{exp}^*} H^1(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes W_{k,j}) & \xrightarrow{\text{exp}^*} \mathcal{K}_M. \end{array}$$

Le résultat crucial sur lequel tout repose est une formule explicite pour l'image de $c_M(k, j)$, $1 \leq j \leq k-1$, dans \mathcal{K}_M . Cette formule fait intervenir le q -développement du produit de deux séries d'Eisenstein⁴³. En particulier, cela permet d'identifier $H^0(\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}, \mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes W_{k,j})$ à un espace de formes modulaires et d'obtenir du même coup une version d'une partie du théorème de comparaison⁴⁴ entre la cohomologie étale p -adique et la cohomologie de de Rham.

Il reste à projeter tous les objets sur la composante correspondant à f pour terminer le calcul. Cette dernière étape demande de calculer le produit scalaire de Petersson de f avec un produit de deux séries d'Eisenstein, ce qui se fait au moyen de la méthode de Rankin, et le résultat fait intervenir les valeurs spéciales des fonctions L attachées à f et à ses tordues par des caractères de Dirichlet⁴⁵.

0.6.6. *La machine de Perrin-Riou* [140, 141, 54] *et la conjecture principale*. Une comparaison des formules ainsi obtenues avec celles fournies par la loi de réciprocité explicite [36] pour les représentations de de Rham de $\mathcal{G}_{\mathbf{Q}_p(\zeta_M)}$ permet de montrer que la fonction L p -adique attachée à une forme modulaire est l'image du système d'Euler de

⁴²Comme la cohomologie étale à coefficients dans quelque chose d'aussi gros que $\mathbb{B}_{\text{dR}}^+(\overline{\mathbf{Q}}_p) \otimes W_k(j)$ n'est pas vraiment définie, on utilise le fait que $Y(\mathbf{M})$ est une courbe affine sur un corps et donc que sa cohomologie étale est aussi la cohomologie continue de son groupe fondamental.

⁴³Scholl [163] a suivi le même chemin pour étudier le système d'Euler qu'il a construit en partant des éléments K-théoriques de Beilinson. En particulier, en utilisant une variante d'une loi de réciprocité explicite de Kato [100], il a calculé l'image par l'application exp^* du tordu à la Soulé de ce système d'Euler par $\mathbf{Q}_p(-1)$. Comme on tombe sur le même produit de séries d'Eisenstein que Kato pour $c_M(k, k-1)$, cela permet, en utilisant le fait que les deux systèmes d'Euler vivent dans un même module de rang 1 sur l'algèbre d'Iwasawa (ce calcul de rang est un des résultats non triviaux que l'on obtient par la méthode des systèmes d'Euler), de montrer qu'ils coïncident.

⁴⁴On peut se demander si l'identification ainsi obtenue est compatible avec celle déduite des théorèmes de comparaison de Faltings [72] ou Tsuji [187, 188]; Kato a vérifié [101, §11] la compatibilité avec le théorème de comparaison de Tsuji, et la compatibilité avec celui de Faltings devrait être plus ou moins automatique car la définition des applications exp^* repose sur la notion d'extension presque étale qui est au cœur de l'approche de Faltings. D'un autre côté, cette compatibilité ne semble pas nécessaire pour les applications aux fonctions L de formes modulaires.

⁴⁵Ce type de résultats remonte aux travaux de Shimura [175].

Kato par la machine de Perrin-Riou, ce qui permet d'utiliser les résultats généraux de Perrin-Riou [141, 54] concernant les fonctions L p -adiques attachées aux représentations p -adiques. En utilisant cette machine, Perrin-Riou a construit (de manière purement algébrique) une fonction L d'Iwasawa attachée à une représentation p -adique de $\mathcal{G}_{\mathbf{Q}}$ et minoré l'ordre du zéro d'une telle fonction en tous les entiers en termes de groupes de Selmer⁴⁶. Comme par ailleurs, la méthode de Kolyvagin [99, 142, 158] permet de démontrer que la fonction L d'Iwasawa divise la fonction L p -adique, cela nous fournit la minoration cherchée pour l'ordre du zéro de la fonction L p -adique, ainsi qu'une moitié de la « conjecture principale » [141, 54] selon laquelle les fonctions L p -adique et d'Iwasawa sont égales, à multiplication près par une unité de l'algèbre d'Iwasawa.

0.7. Remords

Les points suivants mériteraient qu'on leur consacre plus de place.

0.7.1. *Formes modulaires à multiplication complexe.* Les formes modulaires correspondent (au moins conjecturalement) aux motifs de rang 2 sur \mathbf{Q} . Parmi ceux-ci se trouvent les motifs de type CM qui sont de rang 1 sur une extension quadratique imaginaire K de \mathbf{Q} , mais sont vus comme motifs de rang 2 sur \mathbf{Q} . Ces motifs de type CM sont plus faciles à manier, et tous les énoncés des n° 0.3-0.5 ont été démontrés dans le cas de type CM avant le cas général. Par exemple, pour une courbe elliptique E de type CM, le prolongement analytique de la fonction $L(E, s)$ remonte à Deuring [68] : cette fonction s'exprime en termes de fonction L attachées à des caractères de Hecke de K ; le théorème 0.20 a, quant à lui été démontré par Shimura [173], la finitude de $E(\mathbf{Q})$ si $L(E, 1) \neq 0$ est due à Coates et Wiles [41, 42, 110], et la finitude de $\sqcup(E)$ sous la même hypothèse, à Rubin [156] dont la démonstration avait été inspirée par un résultat de Thaine [185] que l'on peut voir comme une première approximation de la méthode de Kolyvagin. Finalement, l'inégalité entre l'ordre du zéro de la fonction L p -adique et le corang du groupe de Selmer d'un motif à multiplication complexe est une conséquence de la conjecture principale démontrée par Rubin [157].

En fait, on ne peut pas retrouver les résultats de Rubin via le système d'Euler de Kato car l'image de $\mathcal{G}_{\mathbf{Q}}$ dans la représentation p -adique V associée à un motif de type CM est trop petite pour que la méthode des systèmes d'Euler puisse fonctionner. La démonstration de Rubin utilise à la place le système d'Euler des unités elliptiques (cf. aussi [1]).

0.7.2. *Courbes elliptiques supersingulières.* Si E est une courbe elliptique ordinaire, définie sur \mathbf{Q} , le dual de Pontryagin de $\text{Sel}_p(E/\mathbf{Q}(\zeta_{p^\infty}))$ est de torsion comme module sur l'algèbre d'Iwasawa (ex-conjecture de Mazur, maintenant un théorème grâce aux travaux de Kato).

⁴⁶Elle a de plus vérifié, pour cette fonction L d'Iwasawa, la conjecture de Bloch-Kato à une unité près dans le cas où la minoration précédente est une égalité, ce qui devrait toujours être le cas sauf si on est en présence de zéros « triviaux » ou « supplémentaires ».

On en déduit le fait [112] que $E(\mathbf{Q}(\zeta_{p^\infty}))$ est de rang fini et qu'il existe $n_0 \in \mathbf{N}$, $\lambda, \mu \in \mathbf{N}$ et $\nu \in \mathbf{Z}$ tels que, si $n \geq n_0$, on ait

$$|\sqcup(E/\mathbf{Q}(\zeta_{p^n}))/\sqcup(E/\mathbf{Q}(\zeta_{p^{n_0}}))| = p^{\lambda n + \mu p^n + \nu}.$$

Si E est supersingulière, le dual de Pontryagin de $\text{Sel}_p(E/\mathbf{Q}(\zeta_{p^\infty}))$ n'est plus de torsion⁴⁷ et la situation est plus compliquée, mais on peut quand même, en utilisant les résultats de Kato, montrer [139, 146] que $E(\mathbf{Q}(\zeta_{p^\infty}))$ est de rang fini, et donner des formules [108, 146] pour la croissance du groupe de Tate-Shafarevitch dans la tour cyclotomique. Si $p \geq 5$ ou, plus généralement, si $a_p = 0$, une remarque⁴⁸ de Pollack [147] a donné naissance à une série de jolis travaux [105, 92, 148, 61] précisant et étendant les résultats mentionnés ci-dessus.

0.7.3. *La direction anticyclotomique.* Soit E une courbe elliptique définie sur \mathbf{Q} ou, plus généralement, soit f une forme primitive de poids $k \geq 2$ pair, dont les coefficients du q -développement sont totalement réels. Soit K une extension quadratique de \mathbf{Q} . Si on regarde l'équation fonctionnelle de $L(f \otimes \eta, s)$ pour $\eta : \text{Gal}(K^{\text{ab}}/K) \rightarrow \mathbf{C}^*$ d'ordre fini, on s'aperçoit qu'il existe une grosse sous-extension L_∞ de l'extension anticyclotomique⁴⁹ K^{anti} de K telle que l'on ait $L(f \otimes \eta, \frac{k}{2}) = 0$, quel que soit $\eta : \text{Gal}(L_\infty/K) \rightarrow \mathbf{C}^*$ d'ordre fini. Ceci laisse entrevoir la possibilité de construire de manière systématique des éléments motiviques dans $H^1(\mathcal{G}_L, V_f(\frac{k}{2})^*(1)) = H^1(\mathcal{G}_L, V_f(\frac{k}{2}))$, L parcourant les extensions finies de K contenues dans L_∞ . De fait, si K est imaginaire, on sait construire ces éléments : ce sont les points de Heegner⁵⁰, ou, si $k > 2$, les cycles de Heegner.

Ces points et cycles de Heegner jouent un rôle fondamental dans la démonstration du théorème de Gross-Zagier [85] et de son analogue p -adique [135, 126], et Kolyvagin a développé ses techniques de dérivation pour démontrer l'égalité $r(E) = r_\infty(E)$ et la finitude de $\sqcup(E)$ pour une courbe elliptique ayant un zéro d'ordre ≤ 1 en $s = 1$, à partir du système d'Euler des points de Heegner [106] (cf. [124] pour une extension en poids supérieur à 2). Ils jouent aussi un rôle fondamental dans le résultat de Nekovář [127] sur la parité du corang des groupes de Selmer via la démonstration d'une conjecture de Mazur [113], sur la non trivialité « asymptotique » des points de Heegner, par Vatsal [189] et Cornut [49] (voir aussi [116]).

⁴⁷Par exemple, si $(\Omega_E^+)^{-1}L(E, 1)$ est une unité p -adique, Kurihara [108] a montré, en utilisant les résultats de Kato, que ce module est en fait libre de rang 1, ce qui constitue une vérification de la conjecture principale. Perrin-Riou [146] a généralisé le résultat de Kurihara, ce qui lui permet de vérifier, grâce à des calculs sur ordinateur, la conjecture principale dans des cas particuliers, sans supposer que $(\Omega_E^+)^{-1}L(E, 1)$ est une unité p -adique ou même que $L(E, 1) \neq 0$.

⁴⁸Cette remarque part de l'observation que, si $a_p = 0$, et si $1 - a_p p^{-s} + p^{1-2s} = (1 - \alpha_1 p^{-s})(1 - \alpha_2 p^{-s})$, alors $\alpha_1 = -\alpha_2$, et les distributions $\mu_{E, \alpha_1} \pm \mu_{E, \alpha_2}$ du th. 0.8 ont beaucoup de zéros triviaux.

⁴⁹ K^{anti} est la plus grande extension abélienne L de K telle que l'élément non trivial de $\text{Gal}(K/\mathbf{Q})$ agisse par multiplication par -1 sur $\text{Gal}(L/K)$.

⁵⁰Si $\pi_E : \mathcal{H} \rightarrow E$ est une uniformisation de E par le demi-plan de Poincaré se factorisant à travers $Y_0(N_E)$, les points de Heegner sont les images de $K \cap \mathcal{H}$.

Par ailleurs, ces points et cycles de Heegner, et leurs analogues p -adiques obtenus en utilisant l'uniformisation de courbes de Shimura par le demi-plan de Drinfeld (demi-plan de Poincaré p -adique) permettent de démontrer [14, 15, 16, 17, 20, 18, 19, 61, 93] des versions anticyclotomiques de la plupart des résultats des n° 0.3 et 0.5. Nous renvoyons à [18] pour une introduction à ce cercle d'idées. Signalons juste que la situation anticyclotomique présente quelques particularités intéressantes. C'est le seul cas, à ma connaissance, où l'on sache construire une fonction L p -adique comme transformée de Mellin d'une forme modulaire p -adique (i.e. une construction purement p -adique de la fonction L p -adique). En outre, Darmon [60] a proposé une construction conjecturale de points rationnels sur des extensions abéliennes d'un corps quadratique réel qui mélange le complexe et le p -adique de manière particulièrement alléchante.

0.7.4. *La méthode de Ribet.* On ne peut espérer démontrer, à partir d'un système d'Euler, qu'une divisibilité du type « ordre d'un groupe de Selmer » divise « valeur spéciale de fonction L ». Dans le cas des motifs de rang 1 sur \mathbf{Q} , cela suffit pour montrer une égalité car on dispose de la formule analytique du nombre de classes⁵¹.

Dans le cas général, il semble qu'il va falloir construire explicitement des éléments dans les groupes de Selmer pour démontrer une divisibilité dans l'autre sens. Pour les motifs de rang 1 sur \mathbf{Q} , cela a été fait par Ribet [151] en utilisant le fait que la divisibilité d'une fonction L se traduit par l'existence d'une congruence entre une série d'Eisenstein (i.e. une forme modulaire « provenant de \mathbf{GL}_1 ») et une forme parabolique f ; on peut alors utiliser la représentation galoisienne attachée à f pour construire un élément du groupe de Selmer du motif de rang 1 (qui n'est rien d'autre qu'une partie du groupe des classes d'une extension cyclotomique). La méthode de Ribet a été étendue par Mazur et Wiles [122, 39], ce qui leur a permis de démontrer la conjecture principale dans le cas cyclotomique (en fait ils ne démontrent qu'une divisibilité et concluent en utilisant la formule analytique du nombre de classes).

Ce n'est que récemment que la méthode de Ribet a été étendue aux motifs de rang 2 (les ingrédients automorphes entrant dans les constructions, ainsi que les démonstrations, se sont nettement sophistiqués⁵² en passant de \mathbf{GL}_1 à \mathbf{GL}_2 ...). Si f est une forme primitive

⁵¹On tord pour se ramener aux caractères d'ordre fini et on fait le produit sur tous les caractères de conducteur divisant un entier N fixé ; la formule analytique du nombre de classes pour $\mathbf{Q}(\zeta_N)$ montrant alors que le produit de ces divisibilités est une égalité, ce qui permet de conclure.

⁵² En particulier, la construction ne fournit pas automatiquement, contrairement au cas considéré par Ribet, les extensions que l'on cherche : les semi-simplifiées des représentations que l'on construit par voie automorphe comportent 3 facteurs au lieu de 2, et il faut éliminer les extensions parasites éventuelles pour démontrer que l'on a bien construit l'extension voulue. Cette problématique apparaît dans [9], et la non existence d'extensions parasites correspond à un cas particulier des conjectures de Bloch-Kato [27] qui peut se déduire, dans le cas considéré par Bellaïche, des résultats de Rubin [157]. Pour résoudre le même type de problèmes dans leur cas [176], Skinner et Urban s'appuient à la place sur le théorème 0.22 (de Kato). Par ailleurs, un résultat récent de Kisin [104], concernant la variation du D_{cris} d'une représentation

de poids $2k$ dont la fonction L vérifie l'équation fonctionnelle $\Lambda(f, s) = -\Lambda(f, 2k - s)$, où l'on a posé $\Lambda(f, s) = \frac{\Gamma(s)}{(2\pi)^s} N^{s/2} L(f, s)$, la fonction $L(f, s)$ a un zéro d'ordre ≥ 1 (car impair) en $s = k$, et les conjectures de Beilinson et de Bloch-Kato prédisent l'existence d'extensions non triviales, ayant « bonne réduction partout », de la représentation triviale par $V_f(k)$. En utilisant des congruences entre formes automorphes pour $\mathbf{U}(2, 1)$, Bellaïche [9] a démontré l'existence de telles extensions modulo p , pour « beaucoup de p », si f est de type CM. En utilisant des familles de représentations galoisiennes associées à des formes automorphes pour \mathbf{GSp}_4 , Skinner et Urban [176] ont construit (au moins en niveau 1) de telles extensions dans le cas où f est ordinaire en p , et Bellaïche et Chenevier [10] ont combiné⁵³ les méthodes de [9, 176], pour traiter le cas CM en utilisant des familles de représentations galoisiennes associées à des formes automorphes pour $\mathbf{U}(3)$. Il est à noter que, dans le cas CM, on peut déduire l'existence de telles extensions des travaux de Rubin [157] sur la conjecture principale (en utilisant encore une fois la formule analytique du nombre de classes pour transformer une divisibilité dans le mauvais sens en égalité), et que, dans le cas général, cette existence peut aussi se déduire du résultat de Nekovář [127] concernant la parité du corang des groupes de Selmer, mais que les constructions automorphes de Bellaïche, Chenevier, Skinner et Urban sont pleines de promesses pour l'avenir. Par ailleurs, en ce qui concerne la conjecture principale pour une forme modulaire ordinaire en p , Skinner et Urban [177] ont démontré, modulo l'existence de représentations galoisiennes attachées aux formes automorphes sur le groupe $\mathbf{U}(2, 2)$, la divisibilité opposée à celle démontrée par Kato, ce qui fournit une démonstration de cette conjecture principale (malheureusement conditionnelle pour le moment).

Signalons encore, dans ce cercle d'idées, dans le cas des motifs de rang 2 de type CM, la méthode de Mazur et Tilouine [121, 186] et son extension [89, 90, 88] à un corps CM quelconque.

0.8. Organisation de l'article

La démonstration de Kato des th. 0.15, 0.22 et 0.23 est découpée en quatre gros articles [99, 100, 101, 102] ; le texte principal ne couvre pas la totalité de cette démonstration. Il contient en particulier une construction du système d'Euler de Kato (chap. 1), une esquisse de démonstration de la loi de réciprocité explicite de Kato (chap. 2), et une analyse assez précise du lien entre le système d'Euler de Kato et les fonctions L p -adiques de

dans une famille, joue un grand rôle pour contrôler la bonne réduction en p des extensions construites via les familles de formes automorphes [176, 10].

⁵³Ils ont aussi réussi à éliminer le recours aux résultats profonds de Rubin et Kato pour se débarrasser des extensions parasites éventuelles mentionnées dans la note 52, et ont ramené leur non existence à celle d'extensions de \mathbf{Q}_p par $\mathbf{Q}_p(1)$ (sur un corps quadratique imaginaire) ayant bonne réduction partout, ce qui suit de la théorie de Kummer et de ce que l'anneau des entiers d'un corps quadratique imaginaire n'a pas d'unité d'ordre infini.

formes modulaires (chap. 4). Il manque un traitement de la méthode des systèmes d'Euler (correspondant à [99], et utilisée de manière cruciale pour démontrer les résultats du chap. 3), ainsi que des rappels un peu conséquents sur la machine de Perrin-Riou et les résultats qui en sortent (correspondant aux §§ 17-18 de [101]), mais ces techniques ont déjà fait l'objet d'exposés [137, 54] à ce séminaire. J'ai assez sensiblement modifié le point de vue par rapport à celui de Kato; j'espère que cela n'a pas introduit d'erreur fatale (je n'ai pas vérifié tous les détails, mais ça a une bonne tête).

1. LE SYSTÈME D'EULER DE KATO

Ce chapitre est consacré à la construction du (ou plutôt des) système d'Euler de Kato. Par rapport à la présentation qu'en donnent Kato [101] ou Scholl [161], il y a deux différences sensibles : la situation a été complètement adélisée, ce qui rend les calculs de corestriction quasi automatiques (mais ne rend pas les notations plus digestes...) et la géométrie algébrique a disparu (par exemple, la cohomologie étale des courbes modulaires a été remplacée par la cohomologie continue de leur groupe fondamental, ou plutôt d'un groupe $\Pi_{\mathbf{Q}}$ qui leur est commensurable, et qui est défini de manière purement modulaire).

1.1. Notations et préliminaires

1.1.1. *Adèles.* On note \mathcal{P} l'ensemble des nombres premiers et $\widehat{\mathbf{Z}}$ le complété profini de \mathbf{Z} . On a donc $\widehat{\mathbf{Z}} = \prod_{p \in \mathcal{P}} \mathbf{Z}_p$, et $\mathbf{Q} \otimes \widehat{\mathbf{Z}}$ est l'anneau des adèles finis de \mathbf{Q} ; c'est le produit restreint des \mathbf{Q}_p , $p \in \mathcal{P}$.

Si x est un objet adélique, on note x_p (resp. $x^{[p]}$) la composante de x en p (resp. en dehors de p). Soit $\widehat{\mathbf{Z}}^{[p]} = \prod_{\ell \neq p} \mathbf{Z}_\ell$. La décomposition $\widehat{\mathbf{Z}} = \mathbf{Z}_p \times \widehat{\mathbf{Z}}^{[p]}$ induit, pour $d \geq 1$, des décompositions

$$\mathbf{M}_d(\mathbf{Q} \otimes \widehat{\mathbf{Z}}) = \mathbf{M}_d(\mathbf{Q}_p) \times \mathbf{M}_d(\mathbf{Q} \otimes \widehat{\mathbf{Z}}^{[p]}) \quad \text{et} \quad \mathbf{GL}_d(\mathbf{Q} \otimes \widehat{\mathbf{Z}}) = \mathbf{GL}_d(\mathbf{Q}_p) \times \mathbf{GL}_d(\mathbf{Q} \otimes \widehat{\mathbf{Z}}^{[p]}).$$

Ceci nous permet de définir les sous-ensembles suivants de $\mathbf{Q} \otimes \widehat{\mathbf{Z}}$ et $\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$:

$$\begin{aligned} \widehat{\mathbf{Z}}^{(p)} &= \mathbf{Z}_p^* \times \widehat{\mathbf{Z}}^{[p]} & \text{et} & & \mathbf{M}_2(\widehat{\mathbf{Z}})^{(p)} &= \mathbf{GL}_2(\mathbf{Z}_p) \times \mathbf{M}_2(\widehat{\mathbf{Z}}^{[p]}), \\ (\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)} &= \mathbf{Z}_p^* \times (\mathbf{Q} \otimes \widehat{\mathbf{Z}}^{[p]}) & \text{et} & & \mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)} &= \mathbf{GL}_2(\mathbf{Z}_p) \times \mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}}^{[p]}). \end{aligned}$$

1.1.2. *Cohomologie des groupes (localement) profinis.* Si X est un espace topologique localement profini (comme tous les espaces ci-dessus) et V est un \mathbf{Z} -module, on note $\text{LC}_c(X, V)$ le module des fonctions localement constantes à valeurs dans V dont le support est compact dans X . On note $\mathcal{D}_{\text{alg}}(X, V)$ l'ensemble des applications \mathbf{Z} -linéaires de $\text{LC}_c(X, \mathbf{Z})$ dans V ; un élément de $\mathcal{D}_{\text{alg}}(X, V)$ est une *distribution algébrique sur X à valeurs dans V* . On note $\int_X \phi \mu$ la valeur de μ sur ϕ .

Soit G un groupe localement profini (comme $\mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$) agissant continûment à droite sur X et V (l'action de $g \in G$ sur x est notée $x \star g$, et on a $(x \star g_1) \star g_2 = x \star (g_1 g_2)$). On munit $\mathrm{LC}_c(X, \mathbf{Z})$ et $\mathcal{D}_{\mathrm{alg}}(X, V)$ d'actions de G à droite $\phi \mapsto \phi \star g$ et $\mu \mapsto \mu \star g$, avec

$$(\phi \star g)(x) = \phi(xg^{-1}) \quad \text{et} \quad \int_X \phi(\mu \star g) = \left(\int_X (\phi \star g^{-1}) \mu \right) \star g.$$

Si M est un G -module topologique muni d'une action à droite de G , on note $H^i(G, M)$ le i -ième groupe de cohomologie continue de G à valeurs dans M . Si X est de plus muni d'une action à gauche de G (notée $(g, x) \mapsto g \cdot x$) commutant à l'action à droite de G (i.e. $g_1 \cdot (x \star g_2) = (g_1 \cdot x) \star g_2$), les modules $H^i(G, \mathcal{D}_{\mathrm{alg}}(X, M))$ sont naturellement des G -modules à gauche. Ce qui précède s'applique en particulier à $X = \mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ et $G = \mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$, les actions à gauche et à droite étant simplement données par la multiplication des matrices.

Si H est un sous-groupe fermé de G , si $\phi \in H^0(H, \mathrm{LC}_c(X, \mathbf{Z}))$ et si $\mu \in H^i(G, \mathcal{D}_{\mathrm{alg}}(X, V))$, le cup-produit nous définit un élément $\int_X \phi \mu$ de $H^i(H, V)$. Par ailleurs, si H est d'indice fini dans G , l'application de corestriction $H^i(H, V) \rightarrow H^i(G, V)$ envoie $\int_X \phi \mu$ sur $\mathrm{cor}(\int_X \phi \mu) = \int_X \mathrm{cor} \phi \mu$, où $\mathrm{cor} \phi \in H^0(G, \mathrm{LC}_c(X, \mathbf{Z}))$ est définie par la formule

$$(\mathrm{cor} \phi)(x) = \sum_{g \in H \backslash G} \phi(xg).$$

1.2. Formes modulaires

1.2.1. *Définition.* Soit $\mathcal{H} = \{x + iy, y > 0\}$ le demi-plan de Poincaré. Si A est un sous-anneau de \mathbf{R} , on note $\mathbf{GL}_2(A)_+$, le sous-groupe de $\mathbf{GL}_2(A)$ des éléments de déterminant > 0 . Si $k \in \mathbf{N}$, on définit une action à droite $f \rightarrow f|_k \gamma$ de $\mathbf{GL}_2(\mathbf{R})_+$ sur les fonctions de classe C^∞ de \mathcal{H} dans \mathbf{C} , par la formule

$$(f|_k \gamma)(\tau) = \frac{(\det \gamma)^{k-1}}{(c\tau + d)^k} \cdot f\left(\frac{a\tau + b}{c\tau + d}\right), \quad \text{si } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Si Γ est un sous-groupe d'indice fini de $\mathbf{SL}_2(\mathbf{Z})$, on note $\mathcal{M}_k(\Gamma, \mathbf{C})$, le \mathbf{C} -espace vectoriel des formes modulaires de poids k pour Γ , c'est-à-dire l'ensemble des $f : \mathcal{H} \rightarrow \mathbf{C}$, holomorphes sur \mathbf{C} , à croissance lente à l'infini, vérifiant $f|_k \gamma = f$ quel que soit $\gamma \in \Gamma$. Si $f \in \mathcal{M}_k(\Gamma, \mathbf{C})$, alors f est périodique de période N pour un certain entier $N \geq 1$, et f est somme de sa série de Fourier

$$f(\tau) = \sum_{n=0}^{+\infty} a_{n/N} e^{2i\pi n\tau/N} = \sum_{n=0}^{+\infty} a_{n/N} q^{n/N}, \quad \text{avec } q = e^{2i\pi\tau}.$$

La série $\sum_{n \in \mathbf{Q}_+} a_n q^n$ s'appelle le q -développement de f . Si A est un sous-anneau de \mathbf{C} , on note $\mathcal{M}_k(\Gamma, A)$ le sous- A -module de $\mathcal{M}_k(\Gamma, \mathbf{C})$ des formes dont le q -développement est à coefficients dans A et $\mathcal{M}(\Gamma, A)$ la A -algèbre des formes modulaires pour Γ à coefficients dans A , somme directe des $\mathcal{M}_k(\Gamma, A)$, pour $k \geq 0$. Finalement, on note $\mathcal{M}_k(A)$

(resp. $\mathcal{M}(A)$) la réunion des $\mathcal{M}_k(\Gamma, A)$ (resp. $\mathcal{M}(\Gamma, A)$), où Γ décrit l'ensemble des sous-groupes d'indice fini de $\mathbf{SL}_2(\mathbf{Z})$. L'algèbre $\mathcal{M}(\mathbf{C})$ est munie d'une action de $\mathbf{GL}_2(\mathbf{Q})_+$ notée $f \rightarrow f \star \gamma$, avec $f \star \gamma = (\det \gamma)^{1-k} f|_k \gamma$, si $f \in \mathcal{M}_k(\mathbf{C})$ et $\gamma \in \mathbf{GL}_2(\mathbf{Q})_+$.

1.2.2. *Sous-groupes de congruence.* Si N est un entier ≥ 1 , soit

$$\Gamma_N = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbf{Z}), a-1, b, c, d-1 \equiv 0 \pmod{N} \right\}.$$

C'est un sous-groupe d'indice fini de $\mathbf{SL}_2(\mathbf{Z})$ et on dit qu'un sous-groupe de $\mathbf{SL}_2(\mathbf{Z})$ est *de congruence* s'il contient Γ_N , pour un certain $N \geq 1$. Si A est un sous-anneau de \mathbf{C} , on note $\mathcal{M}_k^{\text{cong}}(A)$ (resp. $\mathcal{M}^{\text{cong}}(A)$) la réunion des $\mathcal{M}_k(\Gamma, A)$ (resp. $\mathcal{M}(\Gamma, A)$), où Γ décrit l'ensemble des sous-groupes de congruence de $\mathbf{SL}_2(\mathbf{Z})$.

1.2.3. *Les groupes Π_K et Π'_K .* Si K est un sous-corps de \mathbf{C} , on note Π_K le groupe des automorphismes de $\mathcal{M}(\bar{K})$ au-dessus de $\mathcal{M}(\mathbf{SL}_2(\mathbf{Z}), K)$; c'est un groupe profini. Si K est algébriquement clos, alors Π_K est le complété profini de $\mathbf{SL}_2(\mathbf{Z})$ (qui est beaucoup plus gros que $\mathbf{SL}_2(\hat{\mathbf{Z}})$). Dans le cas général, on dispose de la suite exacte

$$1 \longrightarrow \Pi_{\bar{K}} \longrightarrow \Pi_K \longrightarrow \mathcal{G}_K \longrightarrow 1,$$

qui est scindée, \mathcal{G}_K agissant sur les coefficients du q -développement des formes modulaires. Par ailleurs, l'algèbre $\mathcal{M}(\bar{K})$ est stable sous l'action de $\mathbf{GL}_2(\mathbf{Q})_+$ définie ci-dessus, et on note Π'_K le sous-groupe des automorphismes de $\mathcal{M}(\bar{K})$ engendré par Π_K et $\mathbf{GL}_2(\mathbf{Q})_+$. Plus généralement, si $S \subset \mathcal{P}$ est fini, on note $\Pi_K^{(S)}$ le sous-groupe de Π'_K engendré par Π_K et $\mathbf{GL}_2(\mathbf{Z}^{(S)})$, où $\mathbf{Z}^{(S)}$ est le sous-anneau de \mathbf{Q} obtenu en inversant tous les nombres premiers qui n'appartiennent pas à S .

Soit \mathbf{Q}^{cycl} l'extension cyclotomique de \mathbf{Q} réunion des $\mathbf{Q}(\zeta_N)$, $N \geq 1$. La sous-algèbre $\mathcal{M}^{\text{cong}}(\mathbf{Q}^{\text{cycl}})$ est stable par $\Pi_{\mathbf{Q}}$ et $\Pi'_{\mathbf{Q}}$ qui agissent à travers $\mathbf{GL}_2(\hat{\mathbf{Z}})$ et $\mathbf{GL}_2(\mathbf{Q} \otimes \hat{\mathbf{Z}})$ respectivement. On a le diagramme commutatif de groupes suivant :

$$\begin{array}{ccccccc} 1 & \longrightarrow & \Pi_{\bar{K}} & \longrightarrow & \Pi_K & \longrightarrow & \mathcal{G}_K \longrightarrow 1 \\ & & \downarrow & & \downarrow & & \downarrow \chi_{\text{cycl}} \\ 1 & \longrightarrow & \mathbf{SL}_2(\hat{\mathbf{Z}}) & \longrightarrow & \mathbf{GL}_2(\hat{\mathbf{Z}}) & \xrightarrow{\det} & \hat{\mathbf{Z}}^* \longrightarrow 1. \end{array}$$

La section de $\mathcal{G}_{\mathbf{Q}}$ dans $\Pi_{\mathbf{Q}}$ décrite ci-dessus induit une section de l'application déterminant $\mathbf{GL}_2(\hat{\mathbf{Z}}) \rightarrow \hat{\mathbf{Z}}^*$; c'est celle qui envoie u sur la matrice $\begin{pmatrix} 1 & 0 \\ 0 & u \end{pmatrix}$.

1.3. Séries d'Eisenstein-Kronecker

1.3.1. *Définition.* Si $(\tau, z) \in \mathcal{H} \times \mathbf{C}$, on pose $q = e^{2i\pi\tau}$, $q_z = e^{2i\pi z}$ et on note ∂_z l'opérateur $\frac{1}{2i\pi} \frac{\partial}{\partial z} = q_z \frac{\partial}{\partial q_z}$. On pose aussi $\mathbf{e}(a) = e^{2i\pi a}$. Si $k \in \mathbf{N}$, $\tau \in \mathcal{H}$, $z, u \in \mathbf{C}$, la série d'Eisenstein-Kronecker (cf. [192] par exemple),

$$H_k(s, \tau, z, u) = \frac{\Gamma(s)}{(-2i\pi)^k} \left(\frac{\tau - \bar{\tau}}{2i\pi} \right)^{s-k} \sum_{\omega \in \mathbf{Z} + \mathbf{Z}\tau} \frac{\overline{\omega + z}^k}{|\omega + z|^{2s}} \mathbf{e} \left(\frac{\omega \bar{u} - u \bar{\omega}}{\tau - \bar{\tau}} \right),$$

qui converge⁵⁴ si $\operatorname{Re}(s) > 1 + \frac{k}{2}$, possède un prolongement méromorphe à tout le plan complexe, holomorphe en dehors de pôles simples en $s = 1$ (si $k = 0$ et $u \in \mathbf{Z} + \mathbf{Z}\tau$) et $s = 0$ (si $k = 0$ et $z \in \mathbf{Z} + \mathbf{Z}\tau$) et vérifie l'équation fonctionnelle

$$\mathbf{H}_k(s, \tau, z, u) = \mathbf{e}\left(\frac{\omega\bar{u} - u\bar{\omega}}{\tau - \bar{\tau}}\right) \cdot \mathbf{H}_k(k - s, \tau, u, z).$$

On définit les fonctions \mathbf{E}_k et \mathbf{F}_k par les formules

$$\mathbf{E}_k(\tau, z) = \mathbf{H}_k(k, \tau, z, 0) \quad \text{et} \quad \mathbf{F}_k(\tau, z) = \mathbf{H}_k(k, \tau, 0, z).$$

On a

$$\mathbf{E}_{k+1}(\tau, z) = \partial_z \mathbf{E}_k(\tau, z), \quad \text{si } k \in \mathbf{N} \quad \text{et} \quad \mathbf{E}_0(\tau, z) = \log |\theta(\tau, z)|, \quad \text{si } z \notin \mathbf{Z}\tau + \mathbf{Z},$$

où $\theta(\tau, z)$ est donnée par le produit infini

$$\theta(\tau, z) = q^{1/12} (q_z^{1/2} - q_z^{-1/2}) \prod_{n \geq 1} ((1 - q^n q_z)(1 - q^n q_z^{-1})).$$

On note $\Delta = (\partial_z \theta(\tau, z)|_{z=0})^{12} = q \prod_{n \geq 1} (1 - q^n)^{24}$ la forme de poids 12 habituelle.

1.3.2. *Les formes modulaires* $\mathbf{E}_{\alpha, \beta}^{(k)}$ et $\mathbf{F}_{\alpha, \beta}^{(k)}$. Les fonctions $\mathbf{E}_k(\tau, z)$ et $\mathbf{F}_k(\tau, z)$ sont périodiques en z de période $\mathbf{Z}\tau + \mathbf{Z}$. Si $(\alpha, \beta) \in (\mathbf{Q}/\mathbf{Z})^2$ et si $(a, b) \in \mathbf{Q}^2$ a pour image (α, β) dans $(\mathbf{Q}/\mathbf{Z})^2$, on pose

$$\mathbf{E}_{\alpha, \beta}^{(k)} = \mathbf{E}_k(\tau, a\tau + b) \quad \text{et} \quad \mathbf{F}_{\alpha, \beta}^{(k)} = \mathbf{F}_k(\tau, a\tau + b).$$

On a les relations de distribution suivantes si e est un entier ≥ 1 :

$$\begin{aligned} \sum_{e\alpha' = \alpha, e\beta' = \beta} \mathbf{E}_{\alpha', \beta'}^{(k)} &= e^k \mathbf{E}_{\alpha, \beta}^{(k)} & \text{et} & & \sum_{e\alpha' = \alpha, e\beta' = \beta} \mathbf{F}_{\alpha', \beta'}^{(k)} &= e^{2-k} \mathbf{F}_{\alpha, \beta}^{(k)} \\ \sum_{e\beta' = \beta} \mathbf{E}_{\alpha, \beta'}^{(k)}\left(\frac{\tau}{e}\right) &= e^k \mathbf{E}_{\alpha, \beta}^{(k)} & \text{et} & & \sum_{e\beta' = \beta} \mathbf{F}_{\alpha, \beta'}^{(k)}\left(\frac{\tau}{e}\right) &= e \mathbf{F}_{\alpha, \beta}^{(k)}. \end{aligned}$$

PROPOSITION 1.1. — (i) $\mathbf{E}_{0,0}^{(2)} = \mathbf{F}_{0,0}^{(2)} = \frac{-1}{24} \mathbf{E}_2^*$, où

$$\mathbf{E}_2^* = \frac{6}{i\pi(\tau - \bar{\tau})} + 1 - 24 \sum_{n=1}^{+\infty} \sigma_1(n) q^n$$

est la série d'Eisenstein non holomorphe de poids 2 habituelle.

(ii) Si $N\alpha = N\beta = 0$, alors

- (a) $\mathbf{E}_{\alpha, \beta}^{(2)} - \mathbf{E}_{0,0}^{(2)} \in \mathcal{M}_2(\Gamma_N, \mathbf{Q}(\zeta_N))$ et $\mathbf{E}_{\alpha, \beta}^{(k)} \in \mathcal{M}_k(\Gamma_N, \mathbf{Q}(\zeta_N))$ si $k \geq 1$, $k \neq 2$.
- (b) $\mathbf{F}_{\alpha, \beta}^{(k)} \in \mathcal{M}_k(\Gamma_N, \mathbf{Q}(\zeta_N))$ si $k \geq 1$, $k \neq 2$ ou si $k = 2$ et $(\alpha, \beta) \neq (0, 0)$.

Par ailleurs, en faisant agir $\Pi_{\mathbf{Q}}$ trivialement sur $\mathbf{E}_{0,0}^{(2)}$, on obtient le résultat suivant.

PROPOSITION 1.2. — Si $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{GL}_2(\widehat{\mathbf{Z}})$, si $k \geq 1$, et si $(\alpha, \beta) \in (\mathbf{Q}/\mathbf{Z})^2$, alors

$$\mathbf{E}_{\alpha, \beta}^{(k)} \star \gamma = \mathbf{E}_{a\alpha+c\beta, b\alpha+d\beta}^{(k)} \quad \text{et} \quad \mathbf{F}_{\alpha, \beta}^{(k)} \star \gamma = \mathbf{F}_{a\alpha+c\beta, b\alpha+d\beta}^{(k)}.$$

⁵⁴Si $z \in \mathbf{Z} + \mathbf{Z}\tau$, on supprime le terme correspondant à $\omega = -z$ de la somme.

1.3.3. *Quelques q -développements.* Si $\alpha \in \mathbf{Q}/\mathbf{Z}$, posons

$$\zeta(\alpha, s) = \sum_{n \in \mathbf{Q}_+^*, n \equiv \alpha \pmod{\mathbf{Z}}} n^{-s} \quad \text{et} \quad \zeta^*(\alpha, s) = \sum_{n=1}^{+\infty} e^{2i\pi n\alpha} n^{-s}.$$

PROPOSITION 1.3. — Si $k \geq 1$, et $\alpha, \beta \in \mathbf{Q}/\mathbf{Z}$, alors le q -développement $\sum_{n \in \mathbf{Q}_+} a_n q^n$ de $F_{\alpha, \beta}^{(k)}$ est donné par

$$\sum_{n \in \mathbf{Q}_+^*} \frac{a_n}{n^s} = \zeta(\alpha, s - k + 1) \zeta^*(\beta, s) + (-1)^k \zeta(-\alpha, s - k + 1) \zeta^*(-\beta, s)$$

et $a_0 = \zeta(\alpha, 1 - k)$ [resp. $a_0 = \frac{1}{2}(\zeta^*(\beta, 0) - \zeta^*(-\beta, 0))$] si $k \neq 1$ ou $\alpha \neq 0$ (resp. si $k = 1$ et $\alpha = 0$).

REMARQUE 1.4. — Il y a des formules similaires pour le q -développement de $E_{\alpha, \beta}^{(k)}$, mais nous n'en aurons pas besoin.

1.3.4. *Les distributions $\mathbf{z}_{\text{Eis}}(k)$ et $\mathbf{z}'_{\text{Eis}}(k)$.* Les relations de distributions et la proposition 1.2 peuvent se condenser agréablement en l'énoncé suivant⁵⁵ :

THÉORÈME 1.5. — Si $k \geq 1$, il existe une distribution algébrique $\mathbf{z}_{\text{Eis}}(k)$ (resp. $\mathbf{z}'_{\text{Eis}}(k)$) sur $(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2$, à valeurs dans $\mathcal{M}_k(\overline{\mathbf{Q}})$ telle que, quels que soient $r \in \mathbf{Q}^*$ et $(a, b) \in \mathbf{Q}^2$, on ait

$$\int_{(a+r\widehat{\mathbf{Z}}) \times (b+r\widehat{\mathbf{Z}})} \mathbf{z}_{\text{Eis}}(k) = r^{-k} E_{r^{-1}a, r^{-1}b}^{(k)} \quad \left(\text{resp.} \quad \int_{(a+r\widehat{\mathbf{Z}}) \times (b+r\widehat{\mathbf{Z}})} \mathbf{z}'_{\text{Eis}}(k) = r^{k-2} F_{r^{-1}a, r^{-1}b}^{(k)} \right).$$

De plus, si $\gamma \in \mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$, alors

$$\mathbf{z}_{\text{Eis}}(k)|_k \gamma = |\det \gamma|^{k-1} \mathbf{z}_{\text{Eis}}(k) \quad \text{et} \quad \mathbf{z}'_{\text{Eis}}(k)|_k \gamma = \mathbf{z}'_{\text{Eis}}(k).$$

1.3.5. *La distribution $\mathbf{z}_{\text{Eis}}(k, j)$.* On identifie $(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2 \times (\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2$ à $\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ en envoyant $((a, b), (c, d))$ sur la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. En utilisant le fait que le produit de deux formes modulaires de poids i et j est une forme modulaire de poids $i + j$, cela nous fournit une application naturelle

$$\mathcal{D}_{\text{alg}}((\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2, \mathcal{M}_i(\overline{\mathbf{Q}})) \otimes \mathcal{D}_{\text{alg}}((\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2, \mathcal{M}_j(\overline{\mathbf{Q}})) \longrightarrow \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}}), \mathcal{M}_{i+j}(\overline{\mathbf{Q}})).$$

Si $k \geq 2$ et $1 \leq j \leq k - 1$, soit

$$\mathbf{z}_{\text{Eis}}(k, j) = \frac{(-1)^j}{(j-1)!} \mathbf{z}'_{\text{Eis}}(k-j) \otimes \mathbf{z}_{\text{Eis}}(j) \in \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}}), \mathcal{M}_k(\overline{\mathbf{Q}})).$$

⁵⁵L'existence des distributions $\mathbf{z}_{\text{Eis}}(k)$ et $\mathbf{z}'_{\text{Eis}}(k)$ est une conséquence de la première relation de distribution; la loi de transformation sous $\mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ se démontre en utilisant la proposition 1.2, la seconde relation de distribution, et le fait que tout élément de $\mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ peut s'écrire sous la forme $g_1 \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e \end{pmatrix} g_2$, avec $g_1, g_2 \in \mathbf{GL}_2(\widehat{\mathbf{Z}})$, $r \in \mathbf{Q}^*$ et e entier ≥ 1 .

La valeur de cette distribution sur une fonction localement constante à support compact est une combinaison linéaire de produits de séries d'Eisenstein⁵⁶. D'autre part, il résulte du théorème 1.5 que :

PROPOSITION 1.6. — Si $\gamma \in \mathbf{GL}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$, alors

$$\mathbf{z}_{\text{Eis}}(k, j)|_k \gamma = |\det \gamma|^{j-1} \mathbf{z}_{\text{Eis}}(k, j).$$

1.4. Unités de Siegel et éléments de Beilinson

1.4.1. *Unités de Siegel.* La fonction $\theta(\tau, z)$ n'est pas périodique en z de période $\mathbf{Z}\tau + \mathbf{Z}$ mais, si $c \geq 2$ est un entier premier à 6, alors la fonction $\theta(\tau, z)^{c^2} \theta(\tau, cz)^{-1}$ est périodique. Si $(\alpha, \beta) \in (\mathbf{Q}/\mathbf{Z})^2 - (0, 0)$, et si $(a, b) \in \mathbf{Q}^2$ a pour image $(\alpha, \beta) \in (\mathbf{Q}/\mathbf{Z})^2$, posons

$$g_{c,\alpha,\beta} = \theta(\tau, a\tau + b)^{c^2} \theta(\tau, ca\tau + cb)^{-1}.$$

PROPOSITION 1.7. — Soient $\alpha, \beta \in \frac{1}{\mathbf{N}}\mathbf{Z}/\mathbf{Z}$.

(i) Si $c \in \mathbf{N}$ est premier à 6 et si $(c\alpha, c\beta) \neq (0, 0)$, alors $g_{c,\alpha,\beta}$ est une unité de $\mathcal{M}(\Gamma_{\mathbf{N}}, \mathbf{Q}(\zeta_{\mathbf{N}}))[\frac{1}{\Delta}]$.

(ii) L'élément $g_{\alpha,\beta} = g_{c,\alpha,\beta}^{1/(c^2-1)}$ de $\mathbf{Q} \otimes (\mathcal{M}(\overline{\mathbf{Q}})[\frac{1}{\Delta}])^*$ ne dépend pas du choix de c congru à 1 modulo \mathbf{N} . De plus, quel que soit c premier à 6, on a $g_{c,\alpha,\beta} = g_{\alpha,\beta}^{c^2} g_{c\alpha,c\beta}^{-1}$.

Le théorème suivant se démontre de la même manière que le théorème 1.5

THÉORÈME 1.8. — Il existe une distribution algébrique $\mathbf{z}_{\text{Siegel}}$ sur $(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2 - (0, 0)$, à valeurs dans $\mathbf{Q} \otimes (\mathcal{M}(\overline{\mathbf{Q}})[\frac{1}{\Delta}])^*$, telle que, quels que soient $r \in \mathbf{Q}^*$ et $(a, b) \in \mathbf{Q}^2 - r\widehat{\mathbf{Z}}^2$, on ait

$$\int_{(a+r\widehat{\mathbf{Z}}) \times (b+r\widehat{\mathbf{Z}})} \mathbf{z}_{\text{Siegel}} = g_{r^{-1}a, r^{-1}b}.$$

De plus, $\mathbf{z}_{\text{Siegel}}$ est invariante sous l'action de $\Pi'_{\mathbf{Q}}$.

1.4.2. *Éléments de Beilinson.* Si A est un anneau, on dispose d'une application « symbole de Steinberg » $x \otimes y \mapsto \{x, y\}$ de $A^* \otimes A^*$ dans le groupe de K-théorie $K_2(A)$. Ceci permet de construire une distribution algébrique $\mathbf{z}_{\text{Bei}} = \{\mathbf{z}_{\text{Siegel}}, \mathbf{z}_{\text{Siegel}}\}$ sur⁵⁷ $\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ à valeurs dans $K_2(\mathcal{M}(\overline{\mathbf{Q}})[\frac{1}{\Delta}])$. Cette distribution est invariante sous l'action à droite de $\Pi'_{\mathbf{Q}}$.

⁵⁶Si M, N sont des entiers ≥ 1 , soit $O_{M,N} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a-1, b \in M\widehat{\mathbf{Z}}, c, d-1 \in N\widehat{\mathbf{Z}} \right\}$ et soit $\phi_{M,N}$ la fonction caractéristique de $O_{M,N}$. Par construction, on a

$$\int \phi_{M,N} \mathbf{z}_{\text{Eis}}(k, j) = \frac{(-1)^j}{(j-1)!} M^{k-j-2} N^{-j} F_{\frac{1}{M}, 0}^{(k-j)} E_{0, \frac{1}{N}}^{(j)}.$$

⁵⁷Elle n'est définie que sur l'ouvert $\mathbf{M}'_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$ des matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ avec $(a, b) \neq (0, 0)$ et $(c, d) \neq (0, 0)$. Si M, N sont des entiers ≥ 1 , et si $\Gamma_{M,N} = \mathbf{SL}_2(\mathbf{Z}) \cap O_{M,N}$, alors

$$\int \phi_{M,N} \mathbf{z}_{\text{Bei}} = \{g_{\frac{1}{M}, 0}, g_{0, \frac{1}{N}}\} \in \mathbf{Q} \otimes K_2\left(\mathcal{M}(\Gamma_{M,N}, \mathbf{Q}(\zeta_{(M,N)}))\left[\frac{1}{\Delta}\right]\right)$$

est l'élément construit par Beilinson.

1.5. La théorie p -adique

1.5.1. *Théorie de Kummer et classe de Chern p -adique.* D'après le th. 1.8, la distribution $\mathbf{z}_{\text{Siegel}}$ est invariante par $\Pi'_{\mathbf{Q}}$ et on note $\mathbf{z}_{\text{Siegel}}^{(p)} \in H^1(\Pi'_{\mathbf{Q}}, \mathcal{D}_{\text{alg}}((\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2 - (0, 0), \mathbf{Q}_p(1)))$ son image par l'application de Kummer⁵⁸. On note

$$\mathbf{z}_{\text{Kato}} \in H^2(\Pi_{\mathbf{Q}}^{(p)}, \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, \mathbf{Q}_p(2)))$$

la restriction de $\mathbf{z}_{\text{Siegel}}^{(p)} \otimes \mathbf{z}_{\text{Siegel}}^{(p)}$. C'est aussi, au signe près, l'image de \mathbf{z}_{Bei} par l'application classe de Chern, mais nous n'utiliserons pas ce fait.

Pour aller plus loin, il faut se débarrasser des dénominateurs dans la distribution \mathbf{z}_{Kato} . Si $x \in \mathbf{Z}_p^*$, on note $\langle x \rangle$ son image dans $\widehat{\mathbf{Z}}^*$ et si $x, y \in \mathbf{Z}_p^*$, on note $\langle x, y \rangle$ l'image de $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$ dans $\mathbf{GL}_2(\widehat{\mathbf{Z}})$. Si $c \in \mathbf{Z}_p^*$, on déduit du (i) de la proposition 1.7 l'appartenance de $(c^2 - \langle c^{-1} \rangle) \cdot \mathbf{z}_{\text{Siegel}}^{(p)}$ à $H^1(\Pi'_{\mathbf{Q}}, \mathcal{D}_{\text{alg}}((\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2 - (0, 0), \mathbf{Z}_p(1)))$ et donc

$$\mathbf{z}_{\text{Kato},c,d} = (c^2 - \langle c^{-1}, 1 \rangle) \cdot (d^2 - \langle 1, d^{-1} \rangle) \cdot \mathbf{z}_{\text{Kato}} \in H^2(\Pi_{\mathbf{Q}}^{(p)}, \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, \mathbf{Z}_p(2))).$$

1.5.2. *Torsion à la Soulé.* L'intérêt d'avoir supprimé les dénominateurs est de pouvoir intégrer des fonctions continues à support compact et pas seulement des fonctions localement constantes (en d'autres termes, $\mathcal{D}_{\text{alg}}(\mathbf{M}'_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}}), \mathbf{Z}_p(2))$ est l'espace des mesures $\mathcal{D}_0(\mathbf{M}'_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}}), \mathbf{Z}_p(2))$ sur $\mathbf{M}'_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})$, à valeurs dans $\mathbf{Z}_p(2)$).

On note $t = (\zeta_{p^n})_{n \in \mathbf{N}}$ le générateur canonique de $\mathbf{Z}_p(1)$ et on fait agir $\gamma \in \mathbf{GL}_2(\mathbf{Z}_p)$ sur $\mathbf{Z}_p(1)$ par multiplication par $\det \gamma$. On note $V_p = \mathbf{Q}_p e_1 \oplus \mathbf{Q}_p e_2$ la représentation de dimension 2 de $\mathbf{GL}_2(\mathbf{Z}_p)$ donnée par $e_1 \star \gamma = a e_1 + b e_2$ et $e_2 \star \gamma = c e_1 + d e_2$ si $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{GL}_2(\mathbf{Z}_p)$. Si $k \geq 2$ et $j \in \mathbf{Z}$, soit $V_{k,j} = \text{Sym}^{k-2} V_p \otimes \mathbf{Q}_p(2-j)$. Multipliant la mesure $\mathbf{z}_{\text{Kato},c,d}$ par la fonction $x \mapsto (e_1^{k-2} t^{-j}) \star x_p$ qui est continue sur $\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}$, on définit, pour $j \in \mathbf{Z}$,

$$\mathbf{z}_{\text{Kato},c,d}(k, j) = ((e_1^{k-2} t^{-j}) \star x_p) \mathbf{z}_{\text{Kato},c,d} \in H^2(\Pi_{\mathbf{Q}}^{(p)}, \mathcal{D}_0(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, V_{k,j})),$$

$\Pi_{\mathbf{Q}}^{(p)}$ agissant sur $V_{k,j}$ à travers son quotient $\mathbf{GL}_2(\mathbf{Z}_p)$.

1.6. Du système d'Euler de Kato aux séries d'Eisenstein

Soit \mathcal{K} le corps des fractions du séparé complété de $\mathbf{Z}_p[[q]][q^{-1}]$ pour la topologie p -adique. C'est un corps local de dimension 2 de corps résiduel $\mathbf{F}_p((q))$. On se fixe une clôture algébrique $\overline{\mathcal{K}}$ de \mathcal{K} , un plongement de $\overline{\mathbf{Q}}_p$ dans $\overline{\mathcal{K}}$ et un morphisme $r \mapsto q^r$ de \mathbf{Q} dans $\overline{\mathcal{K}}^*$ vérifiant $q^1 = q$.

⁵⁸ Soit $Z^0 = \{(x_n)_{n \in \mathbf{N}}, x_n \in (\mathcal{M}(\overline{\mathbf{Q}})[\frac{1}{\Delta}])^*, x_{n+1}^p = x_n \text{ si } n \in \mathbf{N}\}$. Soit $Z = \mathbf{Q} \otimes Z^0$. Alors Z est muni d'une action de $\Pi'_{\mathbf{Q}}$ et la suite $0 \rightarrow \mathbf{Q}_p(1) \rightarrow Z \rightarrow (\mathcal{M}(\overline{\mathbf{Q}})[\frac{1}{\Delta}])^* \otimes \mathbf{Q} \rightarrow 0$ est une suite exacte de $\Pi'_{\mathbf{Q}}$ -modules. Posons $X = (\mathbf{Q} \otimes \widehat{\mathbf{Z}})^2 - (0, 0)$ et soit $(\phi_i)_{i \in I}$ une base de $\text{LC}_c(X, \mathbf{Z})$ sur \mathbf{Z} . On peut fabriquer une distribution algébrique μ sur X , à valeurs dans Z , en prenant pour $\int_X \phi_i \mu$ n'importe quel relèvement dans Z de $\int_X \phi_i \mathbf{z}_{\text{Siegel}}^{(p)}$ et alors $\mathbf{z}_{\text{Siegel}}^{(p)}$ est l'image du cocycle $\sigma \mapsto \mu \star \sigma - \mu$.

Si M est un entier ≥ 1 , on note q_M la racine M -ième $q^{1/M}$ de q . Soit $\mathcal{K}_M = \mathcal{K}[q_M, \zeta_M]$; c'est une extension galoisienne de \mathcal{K} de groupe de Galois $P_M = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G_M = \mathbf{GL}_2(\mathbf{Z}/M\mathbf{Z}), a = 1, c = 0 \right\}$. On note \mathcal{K}_∞ la réunion des \mathcal{K}_M , $M \geq 1$.

Dans le chapitre suivant, nous allons définir une application exponentielle duale

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}}, \mathbf{Z}_p[P_M] \otimes V_{k,j}) = H^2(\mathcal{G}_{\mathcal{K}_M}, V_{k,j}) \longrightarrow \mathcal{K}_M = H^0(\mathcal{G}_{\mathcal{K}}, \mathcal{K}_\infty \otimes \mathbf{Z}_p[P_M]).$$

Plus généralement, si W est une représentation de $\mathcal{G}_{\mathcal{K}}$ sur laquelle $\mathcal{G}_{\mathcal{K}}$ agit à travers un quotient fini de $\text{Gal}(\mathcal{K}_\infty/\mathcal{K})$, on définit une application exponentielle duale

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}}, W \otimes V_{k,j}) \longrightarrow H^0(\mathcal{G}_{\mathcal{K}}, \mathcal{K}_\infty \otimes W).$$

Ces applications se recollent pour donner naissance à une application

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}}, \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, V_{k,j})) \longrightarrow H^0(\mathcal{G}_{\mathcal{K}}, \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, \mathcal{K}_\infty)).$$

L'application qui à une forme modulaire associe son q -développement nous fournit une injection de $\mathcal{M}(\overline{\mathbf{Q}})$ dans \mathcal{K}_∞ et un morphisme de $\mathcal{G}_{\mathcal{K}}$ dans $\Pi_{\mathbf{Q}}$. Ce morphisme induit un morphisme « de localisation » $H^i(\Pi_{\mathbf{Q}}, W) \rightarrow H^i(\mathcal{G}_{\mathcal{K}}, W)$ pour tout $\Pi_{\mathbf{Q}}$ -module W et tout $i \in \mathbf{N}$. On note $\mathbf{z}_{\text{Eis}}^{(p)}(k, j)$ l'élément de $H^0(\mathcal{G}_{\mathcal{K}}, \mathcal{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}, \mathcal{K}_\infty))$ obtenu en localisant la restriction de $\mathbf{z}_{\text{Eis}}(k, j)$ à $\mathbf{M}_2(\mathbf{Q} \otimes \widehat{\mathbf{Z}})^{(p)}$, et en utilisant l'injection de $\mathcal{M}(\overline{\mathbf{Q}})$ dans \mathcal{K}_∞ .

Le résultat suivant permet de faire le lien entre le système d'Euler de Kato et les fonctions L ; c'est la clef de voûte de tout l'édifice. Sa démonstration est un long calcul délicat dans les anneaux de Fontaine auquel le chapitre suivant est consacré.

THÉORÈME 1.9. — *Si $k \geq 2$, et $1 \leq j \leq k - 1$, et si $c, d \in \mathbf{Z}_p^*$, alors*

$$\exp^*(\mathbf{z}_{\text{Kato},c,d}(k, j)) = (c^2 - \langle c^{-1}, 1 \rangle) \cdot (d^2 - \langle 1, d^{-1} \rangle) \cdot \mathbf{z}_{\text{Eis}}^{(p)}(k, j).$$

1.7. Opérateurs de Hecke

1.7.1. Définition. Si Γ est un sous-groupe ouvert de $\mathbf{GL}_2(\widehat{\mathbf{Z}})$, on note $\tilde{\Gamma}$ l'image inverse de Γ dans $\Pi_{\mathbf{Q}}$. Si $g \in \mathbf{GL}_2(\mathbf{Q})$ est tel que $g^{-1}\Gamma g \subset \mathbf{GL}_2(\widehat{\mathbf{Z}})$, alors le sous-groupe $g^{-1}\tilde{\Gamma}g$ de $\Pi'_{\mathbf{Q}}$ est inclus dans $\Pi_{\mathbf{Q}}$ et est l'image inverse de $g^{-1}\Gamma g$ dans $\Pi_{\mathbf{Q}}$. On en déduit, si V est un $\Pi'_{\mathbf{Q}}$ -module, un isomorphisme⁵⁹ $c \mapsto c \star g$ de $H^i(\tilde{\Gamma}, V)$ sur $H^i(g^{-1}\tilde{\Gamma}g, V)$.

Si ℓ est un nombre premier, et M, N sont des entiers ≥ 1 , on définit le sous-groupe $\widehat{\Gamma}_{M(\ell), N}$ (resp. $\widehat{\Gamma}_{M, N(\ell)}$) comme l'intersection de $\mathbf{GL}_2(\widehat{\mathbf{Z}})$ avec l'ensemble des matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de $\mathbf{M}_2(\widehat{\mathbf{Z}})$ vérifiant $a - 1 \in M\widehat{\mathbf{Z}}$, $b \in M\ell\widehat{\mathbf{Z}}$ (resp. $b \in M\widehat{\mathbf{Z}}$), $c \in N\widehat{\mathbf{Z}}$ (resp. $c \in N\ell\widehat{\mathbf{Z}}$),

⁵⁹Le cocycle $(\sigma_1, \dots, \sigma_i) \mapsto c_{\sigma_1, \dots, \sigma_i}$ est envoyé sur $(\sigma_1, \dots, \sigma_i) \mapsto (c_{g\sigma_1g^{-1}, \dots, g\sigma_ig^{-1}}) \star g$.

$d - 1 \in \widehat{\mathbf{N}}\widehat{\mathbf{Z}}$. On définit⁶⁰ les opérateurs de Hecke $T(\ell)$ et $T'(\ell)$ comme les composés :

$$T(\ell) : H^i(\widetilde{\Gamma}_{M,N}, V) \xrightarrow{\text{res}} H^i(\widetilde{\Gamma}_{M(\ell),N}, V) \xrightarrow{\star(1,\ell)} H^i(\widetilde{\Gamma}_{M,N(\ell)}, V) \xrightarrow{\text{cor}} H^i(\widetilde{\Gamma}_{M,N}, V),$$

$$T'(\ell) : H^i(\widetilde{\Gamma}_{M,N}, V) \xrightarrow{\text{res}} H^i(\widetilde{\Gamma}_{M,N(\ell)}, V) \xrightarrow{\star(\ell,1)} H^i(\widetilde{\Gamma}_{M(\ell),N}, V) \xrightarrow{\text{cor}} H^i(\widetilde{\Gamma}_{M,N}, V).$$

Si $(\ell, MN) = 1$, on a $T'(\ell)(\ell^{-1}, 1)^{|\ell|} = T(\ell)(1, \ell^{-1})^{|\ell|}$.

1.7.2. *Corestriction et opérateurs de Hecke.* Soit $S \subset \mathscr{P}$ un ensemble fini, soit V un $\Pi_{\mathbf{Q}}^{(S)}$ -module (cf. n° 1.2.3), soit $i \in \mathbf{N}$ et soit $\mu \in H^i(\Pi_{\mathbf{Q}}^{(S)}, \mathscr{D}_{\text{alg}}(\mathbf{M}_2(\mathbf{Z}^{(S)} \otimes \widehat{\mathbf{Z}}), V))$. Soient A, M, N des entiers ≥ 1 , et soit ϕ une fonction sur $\mathbf{M}_2(\widehat{\mathbf{Z}})$, constante modulo A , invariante par $\widetilde{\Gamma}_{M,N}$. Si ℓ est un nombre premier ne divisant pas AMN , soit ϕ_{ℓ} la fonction définie par

$$\phi_{\ell}(x) = \phi(x) \cdot \mathbf{1}_{1+\ell\mathbf{M}_2(\mathbf{Z}_{\ell})}(x_{\ell});$$

cette fonction est invariante par $\widetilde{\Gamma}_{M\ell, N\ell}$.

PROPOSITION 1.10. — *Si $\ell \notin S$, alors*

- (i) $\int \phi \mu \in H^i(\widetilde{\Gamma}_{M,N}, V)$ et $\int \phi_{\ell} \mu \in H^i(\widetilde{\Gamma}_{M\ell, N\ell}, V)$;
- (ii) $\text{cor}(\int \phi_{\ell} \mu) = (\int \phi \mu) \star (1 - T'(\ell) \star (\ell^{-1}, 1)^{|\ell|} + \ell(\ell, \ell) \star (\ell^{-1}, \ell^{-1})^{|\ell|})$.

Démonstration. — Le (i) est immédiat. Pour démontrer le (ii), on part des formules

$$\text{cor}\left(\int \phi_{\ell} \mu\right) = \int \text{cor}(\phi_{\ell}) \mu \quad \text{et} \quad (\text{cor}(\phi_{\ell}))(x) = \phi(x) \cdot \mathbf{1}_{\mathbf{GL}_2(\mathbf{Z}_{\ell})}.$$

On est donc ramené à relier les fonctions caractéristiques de $\mathbf{GL}_2(\mathbf{Z}_{\ell})$ et $\mathbf{M}_2(\mathbf{Z}_{\ell})$. Plus généralement⁶¹, soient $0 \leq r \leq s$ des entiers, et soient $O_{r,s}$ et $G_{r,s}$ les sous-ensembles de $\mathbf{M}_2(\mathbf{Z}_{\ell})$ définis par

$$O_{r,s} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a - 1 \equiv b \equiv 0 \pmod{\ell^r}, c \equiv d - 1 \equiv 0 \pmod{\ell^s} \right\} \quad \text{et} \quad G_{r,s} = \mathbf{GL}_2(\mathbf{Z}_{\ell}) \cap O_{r,s}.$$

Soient $\alpha_{s,i} = \begin{pmatrix} 1 & 0 \\ i\ell^s & 1 \end{pmatrix}$ si $0 \leq i \leq \ell - 1$ et $\alpha_{0,\ell} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Les fonctions caractéristiques de $O_{r,s}$ et $G_{r,s}$ sont reliées de la manière suivante :

$$\mathbf{1}_{G_{r,s}} = \begin{cases} \mathbf{1}_{O_{r,s}} & \text{si } r \geq 1 \text{ et } s \geq 1, \\ \mathbf{1}_{O_{r,s}} \star \left(1 - \sum_{i=0}^{\ell-1} ((\ell, 1)_{\ell} \alpha_{s,i})\right) & \text{si } r = 0 \text{ et } s \geq 1, \\ \mathbf{1}_{O_{r,s}} \star \left(1 - \sum_{i=0}^{\ell} ((\ell, 1)_{\ell} \alpha_{0,i}) + \ell(\ell, \ell)_{\ell}\right) & \text{si } r = s = 0. \end{cases}$$

La proposition se déduit alors de la troisième de ces relations, du fait que, si $(\ell, M) = 1$, les $\alpha_{s,i}$ forment un système de représentants de $\widetilde{\Gamma}_{M(\ell),N} \backslash \widetilde{\Gamma}_{M,N}$, et de la formule $\int (\phi \star g) \mu = (\int \phi \mu) \star g$ qui se démontre à partir du cas $i = 0$ par « décalage » de dimension.

⁶⁰Cette définition est la traduction, en termes de cohomologie des groupes, de la définition usuelle via les correspondances sur les courbes modulaires.

⁶¹Cela peut être utile pour calculer des corestrictions de $\widetilde{\Gamma}_{M\ell, N\ell}$ à $\widetilde{\Gamma}_{M,N}$ dans le cas où ℓ divise N par exemple.

2. LA LOI DE RÉCIPROCITÉ EXPLICITE DE KATO

Ce chapitre est consacré à la définition de l'application \exp^* et au calcul de l'image du système d'Euler de Kato par cette application. Les techniques sont celles introduites par Tate [183] pour calculer la cohomologie galoisienne de \mathbf{C}_p et étendues par Hyodo [91] dans le cas de corps locaux de corps résiduel non parfait.

2.1. L'anneau \mathbb{B}_{dR}

2.1.1. *Définition.* Soit L un corps de caractéristique 0 muni d'une extension v_p de la valuation p -adique normalisée par $v_p(p) = 1$. On note $\mathbb{C}(L)$ le complété de L pour la valuation v_p .

Soit

$$\mathbb{R}(L) = \{(x_n)_{n \in \mathbf{N}}, x_n \in \mathcal{O}_L/p\mathcal{O}_L \text{ et } x_{n+1}^p = x_n \text{ si } n \in \mathbf{N}\}.$$

Si $x = (x_n)_{n \in \mathbf{N}} \in \mathbb{R}(L)$ et si $\hat{x}_n \in \mathcal{O}_L$ a pour image x_n modulo p , alors la suite $\hat{x}_{n+k}^{p^k}$ converge, quand k tend vers $+\infty$, vers un élément $x^{(n)}$ de $\mathcal{O}_{\mathbb{C}(L)}$ qui ne dépend que de x . Ceci permet de mettre $\mathbb{R}(L)$ en bijection avec l'ensemble des suites $x = (x^{(n)})_{n \in \mathbf{N}}$, avec $x^{(n)} \in \mathcal{O}_{\mathbb{C}(L)}$ et $(x^{(n+1)})^p = x^{(n)}$ si $n \in \mathbf{N}$.

L'anneau $\mathbb{R}(L)$ est un anneau parfait de caractéristique p et on note $\mathbb{A}_{\text{inf}}(L) = W(\mathbb{R}(L))$ l'anneau des vecteurs de Witt à coefficients dans $\mathbb{R}(L)$. Si $x \in \mathbb{R}(L)$, on note $[x] \in \mathbb{A}_{\text{inf}}(L)$ son représentant de Teichmüller, et tout élément de $\mathbb{A}_{\text{inf}}(L)$ peut s'écrire de manière unique sous la forme $\sum_{k=0}^{+\infty} p^k [x_k]$, où $(x_k)_{k \in \mathbf{N}}$ est une suite d'éléments de $\mathbb{R}(L)$. On définit un morphisme d'anneaux $\theta : \mathbb{A}_{\text{inf}}(L) \rightarrow \mathcal{O}_{\mathbb{C}(L)}$ par la formule $\theta(\sum_{k=0}^{+\infty} p^k [x_k]) = \sum_{k=0}^{+\infty} p^k x_k^{(0)}$. On note $\mathbb{B}_{\text{inf}}(L)$ l'anneau $\mathbb{A}_{\text{inf}}(L)[\frac{1}{p}]$ et on étend θ en un morphisme de $\mathbb{B}_{\text{inf}}(L)$ dans $\mathbb{C}(L)$. Si $m \geq 1$ est un entier, on note $\mathbb{B}_m(L)$ l'anneau $\mathbb{B}_{\text{inf}}(L)/(\ker \theta)^m$. On fait de $\mathbb{B}_m(L)$ un anneau de Banach en prenant l'image de $\mathbb{A}_{\text{inf}}(L)$ comme anneau d'entiers.

On définit l'anneau $\mathbb{B}_{\text{dR}}^+(L)$ comme la limite projective des $\mathbb{B}_m(L)$ que l'on munit de la topologie de la limite projective, ce qui en fait un anneau de Fréchet. Par construction, θ s'étend en un morphisme continu d'anneaux topologiques de $\mathbb{B}_{\text{dR}}^+(L)$ dans $\mathbb{C}(L)$ et on a $\mathbb{B}_m(L) = \mathbb{B}_{\text{dR}}^+(L)/(\ker \theta)^m$ quel que soit $m \in \mathbf{N}$.

2.1.2. *L'anneau $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$.* Le groupe $\mathcal{G}_{\mathcal{K}}$ agit continûment sur les anneaux $\mathbb{C}(\overline{\mathcal{K}})$ et $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$ et la définition de l'application \exp^* repose sur l'étude de la cohomologie de $\mathcal{G}_{\mathcal{K}}$ à valeurs dans ces anneaux. Si M est un entier, on note \mathcal{K}_{Mp^∞} la réunion des \mathcal{K}_{Mp^n} , $n \in \mathbf{N}$. L'extension $\overline{\mathcal{K}}/\mathcal{K}_{Mp^\infty}$ est presque étale, ce qui se traduit par la nullité de $H^i(\mathcal{G}_{\mathcal{K}_{Mp^\infty}}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V)$ quelle que soit la \mathbf{Q}_p -représentation V de $\mathcal{G}_{\mathcal{K}_M}$ et quel que soit $i \geq 1$. On est donc ramené à étudier de près l'extension $\mathcal{K}_{Mp^\infty}/\mathcal{K}_M$ dont le groupe de Galois $P_{\mathcal{K}_M}$ est un groupe de Lie de dimension 2 isomorphe au groupe P_m [$m = v_p(M)$] des matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de $\mathbf{GL}_2(\mathbf{Z}_p)$ vérifiant $a = 1, c = 0, b \in p^m \mathbf{Z}_p, d \in 1 + p^m \mathbf{Z}_p$. Cette étude repose sur la construction (prop. 2.1) de « traces de Tate normalisées » et des résultats généraux concernant la cohomologie des groupes de Lie p -adiques.

2.1.3. *Traces de Tate normalisées.* Soit $\tilde{\zeta}$ [resp. $\tilde{\zeta}_M$, si M est un entier ≥ 1] le représentant de Teichmüller dans $\mathbb{A}_{\text{inf}}(\overline{\mathcal{K}})$ de $(1, \zeta_p, \dots, \zeta_{p^n}, \dots)$ [resp. $(\zeta_M, \dots, \zeta_{Mp^n}, \dots)$]. Si $M|N$, on a $\tilde{\zeta}_N^{N/M} = \tilde{\zeta}_M$.

Soit $t = \log \tilde{\zeta} = -\sum_{n=1}^{+\infty} \frac{(1-\tilde{\zeta})^n}{n} \in \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$ le $2i\pi$ p -adique de Fontaine. C'est un générateur du noyau de $\theta : \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \rightarrow \mathbb{C}(\overline{\mathcal{K}})$ sur lequel $\mathcal{G}_{\mathcal{K}}$ agit par multiplication par χ_{cycl} .

Soit \tilde{q} [resp. \tilde{q}_M , si M est un entier ≥ 1] le représentant de Teichmüller dans $\mathbb{A}_{\text{inf}}(\overline{\mathcal{K}})$ de $(q, q_p, \dots, q_{p^n}, \dots)$ [resp. $(q_M, \dots, q_{Mp^n}, \dots)$]. Si $M|N$, on a $\tilde{q}_N^{N/M} = \tilde{q}_M$. L'application $f(q) \mapsto f(\tilde{q})$ permet d'identifier \mathcal{K} à un sous-anneau de $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$, mais il faut faire attention au fait que \mathcal{K} n'est pas stable par $\mathcal{G}_{\mathcal{K}}$ car $\tilde{q} \star \sigma = \tilde{q}^{\tilde{\zeta}^{c_q(\sigma)}}$ si $\sigma \in \mathcal{G}_{\mathcal{K}}$, où $\sigma \mapsto c_q(\sigma)$ est le cocycle à valeurs dans $\mathbf{Z}_p(1) = \mathbf{Z}_p(\chi_{\text{cycl}})$ associé à q par la théorie de Kummer. Par contre l'anneau $\widetilde{\mathcal{K}} = \mathcal{K}[[t]]$ est stable par $\mathcal{G}_{\mathcal{K}}$.

Si M est un entier ≥ 1 , on note $\widetilde{\mathcal{K}}_M$ l'anneau $\widetilde{\mathcal{K}}[\tilde{q}_M, \tilde{\zeta}_M]$; on a aussi $\widetilde{\mathcal{K}}_M = \mathcal{K}_M[[t]]$. On définit, si $v_p(M) \geq v_p(2p)$, une application R_M de la sous- \mathbf{Q}_p -algèbre $\widetilde{\mathcal{K}}_{Mp^\infty}$ de $\mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty})$ engendrée par les $\tilde{\zeta}_{Mp^n}$ et les \tilde{q}_{Mp^n} , $n \in \mathbf{N}$, dans $\widetilde{\mathcal{K}}_M$ en envoyant $\tilde{\zeta}_{Mp^n}^a \tilde{q}_{Mp^n}^b$ sur $\tilde{\zeta}_{Mp^n}^a \tilde{q}_{Mp^n}^b$ (resp. sur 0) si p^n divise a et b (resp. p^n ne divise pas a ou b).

PROPOSITION 2.1. — *Si M est un entier ≥ 1 , alors*

- (i) $H^0(\mathcal{G}_{\mathcal{K}_{Mp^\infty}}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})) = \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty})$.
- (ii) $\widetilde{\mathcal{K}}_{Mp^\infty}$ est dense dans $\mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty})$.
- (iii) Si $v_p(M) \geq v_p(2p)$, alors R_M s'étend par continuité en une application \mathcal{K} -linéaire $R_M : \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty}) \rightarrow \widetilde{\mathcal{K}}_M$ qui commute à l'action de $\mathcal{G}_{\mathcal{K}}$.

PROPOSITION 2.2. — *Si $v_p(M) \geq v_p(2p)$, si V est une \mathbf{Q}_p -représentation de $P_{\mathcal{K}_M}$ possédant un \mathbf{Z}_p -réseau Γ tel que $P_{\mathcal{K}_M}$ agisse trivialement sur $\Gamma/2p\Gamma$, et si $i \in \mathbf{N}$, alors R_M induit un isomorphisme*

$$R_M : H^i(P_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty}) \otimes V) \cong H^i(P_{\mathcal{K}_M}, \widetilde{\mathcal{K}}_M \otimes V).$$

Le gros intérêt de descendre de $\mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty})$ à $\widetilde{\mathcal{K}}_M$ est que l'action de $P_{\mathcal{K}_M}$ devient analytique, ce qui permet d'utiliser les techniques du § 2.2.

2.2. Cohomologie de P_m

Si $m \geq v_p(2p)$, soit $P_m = \left\{ \begin{pmatrix} 1 & b \\ 0 & d \end{pmatrix}, b \in p^m \mathbf{Z}_p, d \in 1 + p^m \mathbf{Z}_p \right\}$. C'est un groupe analytique p -adique compact et, si $u, v \in p^m \mathbf{Z}_p$, on note (u, v) l'élément $\begin{pmatrix} 1 & u \\ 0 & e^v \end{pmatrix}$ de P_m . La loi de groupe s'écrit alors sous la forme

$$(u_1, v_1)(u_2, v_2) = (e^{v_2}u_1 + u_2, v_1 + v_2).$$

Soient U_m et Γ_m les sous-groupes de P_m topologiquement engendrés par $(p^m, 0)$ et $(0, p^m)$ respectivement. Ces deux sous-groupes sont isomorphes à \mathbf{Z}_p , U_m est distingué dans P_m

et on a $P_m/U_m \cong \Gamma_m$. En particulier, U_m et Γ_m n'ayant pas de H^2 , la suite spectrale de Hochschild-Serre nous fournit, si V est une représentation⁶² de P_m , un isomorphisme⁶³

$$H^2(P_m, V) \cong H^1(\Gamma_m, H^1(U_m, V)) \cong V/((p^m, 0) - 1, (0, p^m) - e^{p^m}).$$

Soit V une représentation analytique⁶⁴ de P_m . Soient $\partial_i : V \rightarrow V$, $i = 1, 2$ les opérateurs définis par

$$x \star (u, v) = x + u\partial_1 x + v\partial_2 x + O((u, v)^2).$$

Ces opérateurs se comportent comme des dérivations : si $x_1 \in V_1$ et $x_2 \in V_2$ et si $i = 1, 2$, alors $\partial_i(x_1 \otimes x_2) = (\partial_i x_1) \otimes x_2 + x_1 \otimes \partial_i x_2$. Comme $\partial_1 = \log(p^m, 0)$ et $\partial_2 - 1 = \log(e^{-p^m}(0, p^m))$ sont divisibles par $(p^m, 0) - 1$ et $(0, p^m) - e^{p^m}$ respectivement, on dispose d'une application naturelle de $V/(\partial_1, \partial_2 - 1)$ dans $V/((p^m, 0) - 1, (0, p^m) - e^{p^m})$ qui est un isomorphisme car V est supposée analytique. On en déduit un isomorphisme

$$H^2(P_m, V) \cong V/(\partial_1, \partial_2 - 1).$$

PROPOSITION 2.3. — *Si V est une représentation analytique de P_m , alors*

- (i) *tout élément de $H^2(P_m, V)$ est représentable par un 2-cocycle analytique ;*
- (ii) *l'image, dans $H^2(P_m, V) \cong V/(\partial_1, \partial_2 - 1)$, d'un 2-cocycle analytique*

$$((u, v), (x, y)) \mapsto c_{(u,v),(x,y)} = \sum_{i+j+k+l \geq 2} c_{i,j,k,l} \cdot u^i v^j x^k y^l$$

est aussi celle de $\delta^{(2)}(c_{(u,v),(x,y)}) = c_{1,0,0,1} - c_{0,1,1,0}$ dans $V/(\partial_1, \partial_2 - 1)$.

2.3. L'application \exp^* pour la représentation $V_{k,j}$

La construction⁶⁵ de l'application \exp^* repose sur le résultat suivant qui se démontre en constatant que $e_1^i e_2^{k-2-i} t^\ell f(\tilde{q})$ est vecteur propre de $\partial_2 - 1$ pour la valeur propre $(k-3-i+\ell)$ et en calculant son image par ∂_1 .

PROPOSITION 2.4. — *Si $v_p(M) \geq v_p(2p)$, et si $1 \leq j \leq k-1$, l'application $f \mapsto te_1^{k-2} f$ induit un isomorphisme de \mathcal{K}_M sur $(\mathcal{K}_M \otimes V_{k,j})/(\partial_1, \partial_2 - 1)$.*

⁶²i.e. un espace vectoriel de dimension finie sur \mathbf{Q}_p ou \mathcal{K} muni d'une action continue de P_m agissant de manière linéaire.

⁶³Le groupe $H^1(U_m, V)$ est isomorphe, en tant que groupe, à $V/((p^m, 0) - 1)$ mais, comme $(0, v)(u, 0)(0, v)^{-1} = (e^{-v}u, 0)$, l'action de $(0, v) \in \Gamma_m$ sur $H^1(U_m, V)$ est celle sur $V/((p^m, 0) - 1)$ multipliée par e^{-v} , d'où l'apparition de e^{p^m} dans l'explicitation du groupe $H^1(\Gamma_m, H^1(U_m, V))$.

⁶⁴i.e. une représentation de P_m telle que les coordonnées de la matrice de (u, v) dans une base de V soient les restrictions à $p^m \mathbf{Z}_p \times p^m \mathbf{Z}_p$ de fonctions analytiques sur $\{v_p(u), v_p(v) > s\}$, avec $s < m$.

⁶⁵La construction qui suit est un peu ad hoc, mais elle a le mérite de coller aux calculs du paragraphe suivant. On peut définir une application \exp^* pour toute représentation de de Rham de \mathcal{G}_L , L extension finie de \mathcal{K} ; la définition de cette application ainsi que celle de représentation de de Rham utilise l'anneau $\mathbb{B}_{\text{dR}}(\overline{\mathcal{K}}/\mathcal{K})$ obtenu en complétant $\mathcal{K} \otimes_{\mathbf{Q}_p} \mathbb{B}_{\text{dR}}(\overline{\mathcal{K}})$. (On a $\mathbb{B}_{\text{dR}}(\overline{\mathcal{K}}/\mathcal{K})^{\mathcal{G}_\mathcal{K}} = \mathcal{K}$, alors que $\mathbb{B}_{\text{dR}}(\overline{\mathcal{K}})^{\mathcal{G}_\mathcal{K}} = \mathbf{Q}_p$.)

On note $\text{res}_{k,j} : \mathcal{K}_M \otimes V_{k,j} \rightarrow \mathcal{K}_M$ l'application obtenue en composant la projection de $\mathcal{K}_M \otimes V_{k,j}$ sur $(\mathcal{K}_M \otimes V_{k,j})/(\partial_1, \partial_2 - 1)$ avec l'inverse de l'isomorphisme précédent.

Maintenant, soit $M \geq 1$ avec $v_p(M) \geq v_p(2p)$, soit $c \in H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j})$ et soit $(\sigma, \tau) \mapsto c_{\sigma,\tau}$ un 2-cocycle continu sur $\mathcal{G}_{\mathcal{K}_M}$, à valeurs dans $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j}$, représentant c . L'extension $\overline{\mathcal{K}}/\mathcal{K}_{Mp^\infty}$ étant presque étale, l'application d'inflation induit un isomorphisme

$$H^2(P_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty}) \otimes V_{k,j}) \cong H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j}),$$

et il existe un 2-cocycle continu $(\sigma, \tau) \mapsto c'_{\sigma,\tau}$ sur $P_{\mathcal{K}_M}$, à valeurs dans $\mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty})$ dont l'inflation a pour image c dans $H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j})$.

Comme $v_p(M) \geq v_p(2p)$, on est dans les conditions d'application de la prop. 2.2, et R_M induit un isomorphisme

$$R_M : H^2(P_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty}) \otimes V_{k,j}) \cong H^2(P_{\mathcal{K}_M}, \widetilde{\mathcal{K}}_M \otimes V_{k,j}).$$

Il existe donc (cf. ⁶⁶ prop. 2.3 (i)), un 2-cocycle analytique $(\sigma, \tau) \mapsto c''_{\sigma,\tau}$ sur $P_{\mathcal{K}_M}$, à valeurs dans $\widetilde{\mathcal{K}}_M \otimes V_{k,j}$, ayant même image que $(\sigma, \tau) \mapsto c'_{\sigma,\tau}$ dans $H^2(P_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\mathcal{K}_{Mp^\infty}) \otimes V_{k,j})$.

L'élément $\text{res}_{k,j}(\delta^{(2)}(c''_{\sigma,\tau}))$ de \mathcal{K}_M ne dépend d'aucun des choix que l'on a faits; on le note $\exp^*(c)$, et l'application \exp^* induit un isomorphisme

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j}) \rightarrow \mathcal{K}_M.$$

Plus généralement, si W est une représentation de $\text{Gal}(\mathcal{K}_\infty/\mathcal{K})$ d'image finie, on définit un isomorphisme

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes W \otimes V_{k,j}) \rightarrow H^0(\mathcal{G}_{\mathcal{K}}, \mathcal{K}_\infty \otimes W)$$

en choisissant un entier $M \geq 1$, $v_p(M) \geq v_p(2p)$, tel que $\mathcal{G}_{\mathcal{K}_M}$ agisse trivialement sur W , et en restreignant l'application \exp^*

$$\exp^* : H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes W \otimes V_{k,j}) \rightarrow \mathcal{K}_M \otimes W$$

précédemment définie à la partie fixe par $\text{Gal}(\mathcal{K}_M/\mathcal{K})$.

2.4. Image du système d'Euler de Kato par l'exponentielle duale

2.4.1. *Préliminaires.* Si M est un entier ≥ 1 , et $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, avec $\alpha, \beta, \gamma, \delta \in \{1, \dots, M\}$, on note $\psi_{M,A}$ la fonction caractéristique de $A + MM_2(\widehat{\mathbf{Z}})$. C'est une fonction invariante sous l'action de $\mathcal{G}_{\mathcal{K}_M}$ et $\int \psi_{M,A} \mathbf{z}_{\text{Kato}}(k, j)$ est ⁶⁷ un élément de $H^2(\mathcal{G}_{\mathcal{K}_M}, V_{k,j})$ dont on note

⁶⁶ $\widetilde{\mathcal{K}}_M \otimes V_{k,j}$ n'est pas une représentation analytique de $P_{\mathcal{K}_M}$, mais c'est la limite projective des $(\widetilde{\mathcal{K}}_M/t^n) \otimes V_{k,j}$ qui sont des représentations analytiques de $P_{\mathcal{K}_M}$.

⁶⁷ Nous allons faire comme s'il existait une distribution $\mathbf{z}_{\text{Kato}}(k, j)$ telle que l'on ait (cf. th. 1.9)

$$\mathbf{z}_{\text{Kato},c,d}(k, j) = (c^2 - \langle c^{-1}, 1 \rangle) \cdot (d^2 - \langle 1, d^{-1} \rangle) \cdot \mathbf{z}_{\text{Kato}}(k, j).$$

Les calculs montrent que l'existence d'une telle distribution est plus que probable, et pour les transformer en une « vraie » démonstration, il suffit d'appliquer l'opérateur $(c^2 - \langle c^{-1}, 1 \rangle) \cdot (d^2 - \langle 1, d^{-1} \rangle)$ à tous les objets en présence, ce qui ne fait que compliquer les formules sans modifier les arguments.

$z_{M,A}$ l'image dans $H^2(\mathcal{G}_{\mathcal{K}_M}, \mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}}) \otimes V_{k,j})$. Pour démontrer le théorème 1.9, il suffit⁶⁸ de démontrer le résultat suivant

PROPOSITION 2.5. — *Pour tout couple M, A comme ci-dessus avec $v_p(M) \geq v_p(2p)$ et $\det A \in \mathbf{Z}_p^*$, on a*

$$\exp^*(z_{M,A}) = \frac{(-1)^j}{(j-1)!} M^{k-2-2j} F_{\alpha/M, \beta/M}^{(k-j)} E_{\gamma/M, \delta/M}^{(j)}.$$

Pour démontrer ceci, nous allons avoir besoin d'écrire un 2-cocycle explicite représentant $z_{M,A}$ et le suivre à travers les étapes de la construction de l'application \exp^* . Il y a deux petits miracles qui permettent de mener le calcul à bien (cf. lemme 2.9 et apparition de $ad - bc$ dans le lemme 2.12).

Comme $z_{M,A}$ est fabriqué à partir d'unités de Siegel que l'on peut voir comme éléments de $\overline{\mathcal{K}}$ et même de $\mathcal{K}_\infty \subset \overline{\mathcal{K}}$, nous allons choisir des relèvements privilégiés de ces unités dans $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$. Si N est un entier ≥ 1 , si $a, b \in \mathbf{Z}^2 - N\mathbf{Z}^2$, soit $[\theta(q, q_N^a \zeta_N^b)] \in \mathbb{A}_{\text{inf}}(\overline{\mathcal{K}})$ un représentant de Teichmüller dont l'image dans $\mathbb{C}(\overline{\mathcal{K}})$ est $\theta(q, q_N^a \zeta_N^b)$ (on note $\theta(q, q_z)$, ce qui est noté $\theta(\tau, z)$ dans le n° 1.3.1; même chose pour les séries d'Eisenstein). Ceci détermine $[\theta(q, q_N^a \zeta_N^b)]$ à multiplication près par $\tilde{\zeta}^u$, $u \in \mathbf{Z}_p$.

2.4.2. *Construction d'un 2-cocycle.* Si $a, b, c, d \in \{1, \dots, p^n M\}$ vérifient $a \equiv \alpha$, $b \equiv \beta$, $c \equiv \gamma$ et $d \equiv \delta$ modulo M , soient $\psi_{a,b}^{(n)}$, $\psi_{c,d}^{(n)}$ et $\psi_{a,b,c,d}^{(n)}$ les fonctions caractéristiques respectives de $(a + Mp^n \mathbf{Z}_p) \times (b + Mp^n \mathbf{Z}_p)$, $(c + Mp^n \mathbf{Z}_p) \times (d + Mp^n \mathbf{Z}_p)$ et $\begin{pmatrix} a & b \\ c & d \end{pmatrix} + Mp^n \mathbf{M}_2(\mathbf{Z}_p)$. Notons U_1 et U_2 respectivement les ouverts $(\alpha + M\mathbf{Z}_p) \times (\beta + M\mathbf{Z}_p)$ et $(\gamma + M\mathbf{Z}_p) \times (\delta + M\mathbf{Z}_p)$ de \mathbf{Z}_p^2 , et $U = U_1 \times U_2$ que l'on voit comme un ouvert de $\mathbf{M}_2(\mathbf{Z}_p)$ et même de $\mathbf{GL}_2(\mathbf{Z}_p)$ puisque $\det A \in \mathbf{Z}_p^*$ et $p|M$.

Si $i = 1, 2$, soit $\mu_i \in H^1(\mathcal{G}_{\mathcal{K}_M}, \mathcal{D}_0(U_i, \mathbf{Z}_p(1)))$ la mesure définie par

$$\int \psi_{a,b}^{(n)} \mu_1 = \int_{(a+Mp^n \widehat{\mathbf{Z}}) \times (b+Mp^n \widehat{\mathbf{Z}})} \mathbf{z}_{\text{Siegel}}^{(p)} \quad \text{et} \quad \int \psi_{c,d}^{(n)} \mu_2 = \int_{(c+Mp^n \widehat{\mathbf{Z}}) \times (d+Mp^n \widehat{\mathbf{Z}})} \mathbf{z}_{\text{Siegel}}^{(p)}.$$

Soit $\nu = \mu_1 \otimes \mu_2 \in H^2(\mathcal{G}_{\mathcal{K}_M}, \mathcal{D}_0(U, \mathbf{Z}_p(2)))$.

Si $i = 1, 2$, soit Ψ_i une base du \mathbf{Z} -module des fonctions localement constantes sur U_i constituée de fonctions du type $\psi_{a,b}^{(n)}$ (resp. $\psi_{c,d}^{(n)}$), avec $n \in \mathbf{N}$ et a, b (resp. c, d) comme

⁶⁸Par linéarité, on en déduit le résultat pour toute fonction à support dans $\mathbf{M}_2(\widehat{\mathbf{Z}})^{(p)}$ et on en déduit le cas général en utilisant l'action des homothéties.

ci-dessus. Soit μ_{1,Ψ_1} (resp. μ_{2,Ψ_2}) la distribution algébrique sur U_1 (resp. U_2) définie⁶⁹ par

$$\int \psi_{a,b}^{(n)} \mu_{1,\Psi_1} = \log[\theta(q, q_{M^{p^n}}^a \zeta_{M^{p^n}}^b)] \quad \text{et} \quad \int \psi_{c,d}^{(n)} \mu_{2,\Psi_2} = \log[\theta(q, q_{M^{p^n}}^c \zeta_{M^{p^n}}^d)],$$

si $\psi_{a,b}^{(n)}$ (resp. $\psi_{c,d}^{(n)}$) appartient à Ψ_1 (resp. Ψ_2).

LEMME 2.6. — Si $i = 1, 2$, alors μ_i est représenté par le cocycle $\sigma \mapsto \mu_{i,\Psi_i} \star (\sigma - 1)$ pour tout choix de Ψ_i .

Démonstration. — Il suffit de revenir à la définition de l'application de Kummer (note 58).

2.4.3. *Descente de $\overline{\mathcal{K}}$ à $\mathcal{K}_{M^{p^\infty}}$.* Si $i = 1, 2$, soit $\tilde{\mu}_{i,\Psi_i}$ la distribution algébrique sur U_i définie par

$$\int \psi_{a,b}^{(n)} \tilde{\mu}_{1,\Psi_1} = \log \theta(\tilde{q}, \tilde{q}_{M^{p^n}}^a \tilde{\zeta}_{M^{p^n}}^b) \quad \text{et} \quad \int \psi_{c,d}^{(n)} \tilde{\mu}_{2,\Psi_2} = \log \theta(\tilde{q}, \tilde{q}_{M^{p^n}}^c \tilde{\zeta}_{M^{p^n}}^d),$$

si $\psi_{a,b}^{(n)}$ (resp. $\psi_{c,d}^{(n)}$) appartient à Ψ_1 (resp. Ψ_2).

LEMME 2.7. — Si $i = 1, 2$, alors

- (i) $\tilde{\mu}_{i,\Psi_i} - \mu_{i,\Psi_i}$ est une mesure à valeurs dans $t\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$.
- (ii) L'image de μ_i dans $H^1(\mathcal{G}_{\mathcal{K}_M}, \mathcal{D}_0(U_i, t\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})))$ est représentée par le cocycle $\sigma \mapsto \tilde{\mu}_{i,\Psi_i} \star (\sigma - 1)$ qui est l'inflation d'un cocycle sur $P_{\mathcal{K}_M}$ à valeurs dans $\mathcal{D}_0(U_i, t\mathbb{B}_{\text{dR}}^+(\mathcal{K}_{M^{p^\infty}}))$.

Démonstration. — Le (ii) est une conséquence immédiate du (i) qui, quant à lui, résulte de ce que $[\theta(q, q_{M^{p^n}}^a \zeta_{M^{p^n}}^b)]^{-1} \theta(\tilde{q}, \tilde{q}_{M^{p^n}}^a \tilde{\zeta}_{M^{p^n}}^b)$ appartient à $1 + \tilde{q}^{-1} \mathbb{A}_{\text{inf}}(\overline{\mathcal{K}})$ et a pour image 1 dans $\mathbb{C}(\overline{\mathcal{K}})$, et donc que l'ensemble des $\log \theta(\tilde{q}, \tilde{q}_{M^{p^n}}^a \tilde{\zeta}_{M^{p^n}}^b) - \log[\theta(q, q_{M^{p^n}}^a \zeta_{M^{p^n}}^b)]$ est borné dans $t\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$ quand n décrit \mathbf{N} et a, b décrivent \mathbf{Z} .

Si Λ_1 et Λ_2 sont deux G -modules à droite, si $x_1 \in \Lambda_1$ et $x_2 \in \Lambda_2$, et si $\sigma, \tau \in G$, on note $\{x_1 \otimes x_2\}_{\sigma, \tau}$ l'élément de $\Lambda_1 \otimes \Lambda_2$ défini par

$$\{x_1 \otimes x_2\}_{\sigma, \tau} = (x_1 \star (\tau\sigma - \sigma)) \otimes (x_2 \star (\sigma - 1)).$$

COROLLAIRE 2.8. — L'image de $\nu = \mu_1 \otimes \mu_2$ dans $H^2(\mathcal{G}_{\mathcal{K}_M}, \mathcal{D}_0(U, t^2\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})))$ peut se représenter par l'inflation de $P_{\mathcal{K}_M}$ à $\mathcal{G}_{\mathcal{K}_M}$ du 2-cocycle $(\sigma, \tau) \mapsto \{\mu_{1,\Psi_1} \otimes \mu_{2,\Psi_2}\}_{\sigma, \tau}$.

⁶⁹ $\log[\theta(q, q_M^a, \zeta_M^b)]$ et $\log \theta(\tilde{q}, \tilde{q}_M^a, \tilde{\zeta}_M^b)$ sont des notations commodes, mais ne correspondent pas à des éléments de $\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$; le principal problème est que le q -développement de $\log \theta(q, q_M^a, \zeta_M^b)$ admet des dénominateurs non bornés et donc ne définit pas un élément de \mathcal{K}_M . Par contre,

- $\log[\theta(q, q_M^a, \zeta_M^b)] \star (\sigma - 1) = \log \frac{[\theta(q, q_M^a, \zeta_M^b)] \star \sigma}{[\theta(q, q_M^a, \zeta_M^b)]} \in \mathbf{Q}_p \cdot t$, si $\sigma \in \mathcal{G}_{\mathcal{K}_M}$,
- $\log \theta(\tilde{q}, \tilde{q}_M^a, \tilde{\zeta}_M^b) - \log[\theta(q, q_M^a, \zeta_M^b)] = \log \frac{\theta(\tilde{q}, \tilde{q}_M^a, \tilde{\zeta}_M^b)}{[\theta(q, q_M^a, \zeta_M^b)]} \in t\mathbb{B}_{\text{dR}}^+(\overline{\mathcal{K}})$,
- $\log \theta(\tilde{q} \zeta^x, \tilde{q}_M^a, \tilde{\zeta}_M^{b+y}) - \log \theta(\tilde{q}, \tilde{q}_M^a, \tilde{\zeta}_M^b) \in t\mathcal{O}_{\mathcal{K}_M}[[\frac{t}{p}]]$, si $x, y \in \mathbf{Z}_p$,

sont bien définis; c'est toujours ce type d'expression qui intervient dans les calculs.

2.4.4. *Descente de \mathcal{K}_{Mp^∞} à \mathcal{K}_M .* Par définition de $\mathbf{z}_{\text{Kato},c,d}(k, j)$ et de ν , on a $z_{M,A} = \int_U ((e_1^{k-2}t^{-j}) \star g) \nu$ et, d'après ce qui précède, $z_{M,A}$ est la classe du 2-cocycle

$$(\sigma, \tau) \mapsto \int_U ((e_1^{k-2}t^{-j}) \star g) \{\mu_{1,\Psi_1} \otimes \mu_{2,\Psi_2}\}_{\sigma,\tau},$$

qui est aussi, d'après la proposition 2.2, la classe du 2-cocycle

$$(\sigma, \tau) \mapsto R_M \left(\int_U ((e_1^{k-2}t^{-j}) \star g) \{\mu_{1,\Psi_1} \otimes \mu_{2,\Psi_2}\}_{\sigma,\tau} \right) = \int_U ((e_1^{k-2}t^{-j}) \star g) R_M(\{\mu_{1,\Psi_1} \otimes \mu_{2,\Psi_2}\}_{\sigma,\tau}).$$

LEMME 2.9. — Si $ad - bc \in \mathbf{Z}_p^*$, alors

$$R_M(\log \theta(\tilde{q}, \tilde{q}_{Mp^n}^a \tilde{\zeta}_{Mp^n}^b) \cdot \log \theta(\tilde{q}, \tilde{q}_{Mp^n}^c \tilde{\zeta}_{Mp^n}^d)) = p^{-2n} \log \theta(\tilde{q}^{p^n}, \tilde{q}_M^a \tilde{\zeta}_M^b) \cdot \log \theta(\tilde{q}^{p^n}, \tilde{q}_M^c \tilde{\zeta}_M^d).$$

Démonstration. — Ce lemme se démontre en regardant les q -développements des logarithmes des unités de Siegel et en utilisant le fait que

$$R_M((\tilde{q}_{Mp^n}^a \tilde{\zeta}_{Mp^n}^b)^i (\tilde{q}_{Mp^n}^c \tilde{\zeta}_{Mp^n}^d)^j) = \begin{cases} (\tilde{q}_{Mp^n}^a \tilde{\zeta}_{Mp^n}^b)^i (\tilde{q}_{Mp^n}^c \tilde{\zeta}_{Mp^n}^d)^j & \text{si } p^n | ai + cj \text{ et } p^n | bi + dj, \\ 0 & \text{sinon,} \end{cases}$$

la condition $ad - bc \in \mathbf{Z}_p^*$ entraînant miraculeusement que $ai + cj$ et $bi + dj$ sont divisibles par p^n si et seulement si i et j le sont.

COROLLAIRE 2.10. — *On peut représenter $z_{M,A}$ par le 2-cocycle*

$$(\sigma, \tau) \mapsto \lim_{n \rightarrow +\infty} p^{-2n} \sum \frac{(ae_1 + be_2)^{k-2}}{(ad - bc)^{jt}} \{ \log \theta(\tilde{q}^{p^n}, \tilde{q}_M^a \tilde{\zeta}_M^b) \otimes \log \theta(\tilde{q}^{p^n}, \tilde{q}_M^c \tilde{\zeta}_M^d) \}_{\sigma,\tau},$$

la somme portant sur les quadruplets a, b, c, d d'entiers $\in \{1, \dots, Mp^n\}$, avec $a \equiv \alpha, b \equiv \beta, c \equiv \gamma, d \equiv \delta$ modulo M .

2.4.5. *Passage à l'algèbre de Lie.* Le cocycle ci-dessus est un 2-cocycle analytique, à valeurs dans $\widetilde{\mathcal{K}}_M \otimes V_{k,j}$; son image dans \mathcal{K}_M par l'application \exp^* peut donc se calculer en utilisant les techniques différentielles du § 2.2.

Remarquons que les rôles de a, b, c et d ne sont pas tout à fait symétriques car $\tilde{\zeta}^x$ est analytique en x , alors que \tilde{q}^x n'est défini que pour les valeurs entières de x . On dit que $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto F \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est p^n -négligeable si, à a et c fixés, la fonction $(x, y) \mapsto F \begin{pmatrix} a & \beta + Mx \\ c & \beta + My \end{pmatrix}$ est élément de $p^n \mathcal{O}_{\mathcal{K}_M}[[Mx, My, \frac{tx}{p}, \frac{ty}{p}]]$. On a alors

$$p^{-2n} \sum_{\substack{b \equiv \beta [M], \quad d \equiv \delta [M] \\ 1 \leq b, d \leq Mp^n}} F \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in p^n \mathcal{O}_{\mathcal{K}_M}[[p^{-1}t]].$$

Si $f(x_1, x_2)$ est une fonction de deux variables, on note D_1 et D_2 respectivement les opérateurs $x_1 \frac{d}{dx_1}$ et $x_2 \frac{d}{dx_2}$. Si $n \in \mathbf{N}$ et si a, b sont deux entiers, on pose $f_{a,b}^{(n)} = f(\tilde{q}^{p^n}, \tilde{q}_M^a \tilde{\zeta}_M^b)$.

LEMME 2.11. — A addition près d'un élément p^n -négligeable près, on a

$$\delta^{(2)} \left(\{ \log \theta_{a,b}^{(n)} \otimes \log \theta_{c,d}^{(n)} \}_{\sigma,\tau} \right) = \frac{(ad - bc)t^2}{M^2} \cdot D_2 \log \theta_{a,b}^{(n)} \cdot D_2 \log \theta_{c,d}^{(n)}.$$

Démonstration. — Cela résulte du développement limité

$$\log \theta_{a,b}^{(n)} \star \left(\begin{pmatrix} 1 & u \\ 0 & e^v \end{pmatrix} - 1 \right) = p^n t \cdot D_1 \log \theta_{a,b}^{(n)} + \frac{(au + bv)t}{M} \cdot D_2 \log \theta_{a,b}^{(n)} + O((u, v)^2),$$

et de ce que $D_1 \log \theta$ et $D_2 \log \theta$ ont des développements à coefficients entiers.

LEMME 2.12. — Si f et g sont à coefficients entiers, si $s \geq 2 - j$ et $a, b, c, d \in \mathbf{Z}$, alors

$$(ae_1 + be_2)^{k-2} t^s f_{a,b}^{(n)} \left(a \cdot g_{c,d}^{(n)} + \frac{1}{1-s} \cdot (ad - bc) \cdot \frac{t}{M} \cdot D_2 g_{c,d}^{(n)} \right)$$

est p^n -négligeable dans $(\widetilde{\mathcal{K}}_M \otimes V_{k,j})/(\partial_1, 1 - \partial_2)$.

Démonstration. — Cela résulte du calcul de $(a(1 - \partial_2) + b\partial_1)((ae_1 + be_2)^{k-2} t^s f_{a,b}^{(n)} g_{c,d}^{(n)})$ qui peut se faire en utilisant les formules

$$\begin{aligned} \partial_1 t &= 0, \quad \partial_2 t = t & \partial_1 e_1 &= e_2, \quad \partial_2 e_1 = 0 & \partial_1 e_2 &= 0, \quad \partial_2 e_2 = e_2 \\ \partial_1 f_{a,b}^{(n)} &= p^n t \cdot D_1 f_{a,b}^{(n)} + \frac{at}{M} \cdot D_2 f_{a,b}^{(n)}, & \partial_2 f_{a,b}^{(n)} &= \frac{bt}{M} \cdot D_2 f_{a,b}^{(n)}, \end{aligned}$$

en modifiant de manière évidente la dernière ligne pour calculer $\partial_1 g_{c,d}^{(n)}$ et $\partial_2 g_{c,d}^{(n)}$.

COROLLAIRE 2.13. — A addition d'un élément p^n -négligeable près, on a

$$\begin{aligned} \text{res}_{k,j} \left(\frac{(ae_1 + be_2)^{k-2}}{(ad - bc)^j t^j} \cdot \delta^{(2)}(\{\log \theta_{a,b}^{(n)} \otimes \log \theta_{c,d}^{(n)}\}_{\sigma,\tau}) \right) \\ = M^{-1-j} \cdot \frac{(-1)^{j-1}}{(j-1)!} \cdot a^{k-1-j} \cdot D_2 \log \theta_{a,b}^{(n)} \cdot D_2^j \log \theta_{c,d}^{(n)}. \end{aligned}$$

Démonstration. — Cela résulte du lemme 2.11, d'une récurrence utilisant le lemme 2.12, de la définition de l'application $\text{res}_{k,j}$ (cf. prop. 2.4), et de ce que $\text{res}_{k,j}(t^s e_1^i e_2^{k-2-i} f(\tilde{q})) = 0$ si $s \geq 2$.

Comme $D_2^r \log \theta(x_1, x_2) = E_r(x_1, x_2)$, et comme $\zeta_M^b = \zeta_M^\beta$ si $b \equiv \beta [M]$ et $\zeta_M^d = \zeta_M^\delta$ si $d \equiv \delta [M]$, on en déduit la formule

$$\exp^*(z_{M,A}) = M^{-3-j} \frac{(-1)^{j-1}}{(j-1)!} \lim_{n \rightarrow +\infty} \sum_{\substack{a \equiv \alpha [M], c \equiv \gamma [M] \\ 1 \leq a, c \leq Mp^n}} a^{k-1-j} E_1(q^{p^n}, q_M^a \zeta_M^\beta) E_j(q^{p^n}, q_M^c \zeta_M^\delta).$$

On termine la démonstration de la proposition 2.5, en utilisant le lemme suivant :

LEMME 2.14. — (i) $\sum_{c \equiv \gamma [M], 1 \leq c \leq Mp^n} E_j(q^{p^n} q_M^c \zeta_M^\delta) = E_j(q, q_M^\gamma \zeta_M^\delta) = E_{\gamma/M, \delta/M}^{(j)}$

(ii) Si $r \in \mathbf{N}$, alors

$$\lim_{n \rightarrow \infty} \sum_{a \equiv \alpha [M], 1 \leq a \leq Mp^n} a^r E_1(q^{p^n}, q_M^a \zeta_M^\beta) = M^r F_{r+1}(q, q_M^\alpha, \zeta_M^\beta) = M^r F_{\alpha/M, \beta/M}^{(r+1)}.$$

Démonstration. — Le (i) se démontre en revenant à la définition. Pour démontrer le (ii), on part de la formule

$$E_1(q^{p^n}, q_M^a \zeta_M^\beta) = F_1(q^{p^n}, q_M^a \zeta_M^\beta) = F_{a/Mp^n, \beta/M}^{(1)}(q^{p^n}),$$

et on utilise la formule pour le q -développement de $F_{\alpha,\beta}^{(k)}$ donnée dans la prop. 1.3.

3. FORMES PROPRES POUR LES OPÉRATEURS DE HECKE

Ce chapitre est consacré à la construction (th. 3.1) d'un système d'Euler pour la représentation p -adique associée à une forme primitive, et aux résultats de théorie d'Iwasawa que l'on peut en tirer (th. 3.2). Nous renvoyons au (v) de la remarque 3.3 pour des commentaires sur l'articulation des arguments.

3.1. Le système d'Euler de Kato associé à une forme primitive

3.1.1. *Projection sur un espace propre.* Soient $N \geq 1$ et $k \geq 2$ des entiers, et ε un caractère de Dirichlet modulo N (pas nécessairement primitif). Soit $f = \sum_{n=1}^{+\infty} a_n q^n \in S_k(\Gamma_0(N), \varepsilon)$ une forme primitive. En particulier, $a_1 = 1$ et $\mathbf{Q}(f) = \mathbf{Q}(a_2, \dots, a_n, \dots)$ est une extension finie de \mathbf{Q} . De plus, on a $\overline{a_n} = \varepsilon^{-1}(n)a_n$ quel que soit $n \in \mathbf{N}$ premier à N , et

$$f \star T(\ell) = a_\ell f, \quad f \star T'(\ell) = \overline{a_\ell} f, \quad f \star \begin{pmatrix} u^{-1} & 0 \\ 0 & u \end{pmatrix} = \varepsilon(u)f,$$

si $\ell \nmid N$ est un nombre premier et $u \in \widehat{\mathbf{Z}}^*$.

Si S est un sous-ensemble fini de \mathcal{P} , et si M est un $\mathbf{Q}(f)$ -espace vectoriel muni d'actions des $T(\ell), T'(\ell), \ell \notin S$ et $\begin{pmatrix} u^{-1} & 0 \\ 0 & u \end{pmatrix}$, pour $u \in U$ sous-groupe ouvert de $\widehat{\mathbf{Z}}^*$, on note M_{π_f} le quotient de M par le sous- $\mathbf{Q}(f)$ -espace vectoriel engendré par les $x \star T(\ell) - a_\ell x, x \in M$ et $\ell \notin S$. Dans tous les cas que nous aurons à considérer, le théorème de multiplicité 1 fort entraîne que l'on a $x \star T'(\ell) = \overline{a_\ell} x$ et $x \star \begin{pmatrix} u^{-1} & 0 \\ 0 & u \end{pmatrix} = \varepsilon(u)x$, si $x \in M$, et si $\ell \notin S$ et $u \in U$.

3.1.2. *La représentation p -adique associée à une forme primitive.* Un exemple d'utilisation de la construction précédente est la définition de la représentation p -adique associée à f . Notons $\widehat{\Gamma}_{1,N}$ le sous-groupe de $\mathbf{SL}_2(\widehat{\mathbf{Z}})$ des matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ avec $c \equiv d - 1 \equiv 0$ modulo N . Soit $\widetilde{\Gamma}_{1,N}$ l'image inverse de $\widehat{\Gamma}_{1,N}$ dans $\Pi_{\mathbf{Q}}$ et soit $\overline{\Gamma}_{1,N}$ l'intersection⁷⁰ de $\widetilde{\Gamma}_{1,N}$ et $\Pi_{\overline{\mathbf{Q}}}$. On définit⁷¹ alors la représentation V_f associée à f par

$$V_f = (\mathrm{H}^1(\overline{\Gamma}_{1,N}, \mathrm{Sym}^{k-2} V_p) \otimes_{\mathbf{Q}_p} \mathbf{Q}_p(f))_{\pi_f} \otimes_{\mathbf{Q}_p} \mathbf{Q}_p(2 - k).$$

C'est une $\mathbf{Q}_p(f)$ -représentation de $\mathcal{G}_{\mathbf{Q}}$ de dimension 2, non ramifiée en dehors de Np . Si $\ell \nmid pN$, le déterminant de $1 - \mathrm{Frob}_\ell^{-1} X$ agissant sur V_f est $1 - a_\ell X + \varepsilon(\ell) \ell^{k-1} X^2$. La restriction de V_f à $\mathcal{G}_{\mathbf{Q}_p}$ est de de Rham, de poids de Hodge-Tate 0 et $1 - k$; elle est cristalline si $p \nmid N$ et le polynôme caractéristique de φ sur $\mathbf{D}_{\mathrm{cris}}(V_f)$ est $X^2 - a_p X + \varepsilon(p) p^{k-1}$; elle est semi-stable si $p|N$ et $a_p \neq 0$ (auquel cas, a_p est la valeur propre de φ sur $\mathbf{D}_{\mathrm{cris}}(V_f)$) et seulement potentiellement semi-stable si $p|N$ et $a_p = 0$.

⁷⁰Si $N \geq 5$, c'est le complété profini du groupe fondamental de la surface de Riemann $Y_1(N)$.

⁷¹ V_p est la représentation apparaissant au n° 1.5.2; c'est le module de Tate de la courbe elliptique universelle.

3.1.3. *La fonction L de f et ses tordues.* Si ϕ est une fonction localement constante sur $\widehat{\mathbf{Z}}$ à valeurs dans $\mathbf{Q}(f)$, on définit la fonction $L(f, \phi, s)$ par la formule

$$L(f, \phi, s) = \sum_{n \geq 1} \phi(n) a_n n^{-s}.$$

Cette fonction possède un prolongement analytique à tout le plan complexe et il existe des nombres complexes non nuls Ω_f^+ et Ω_f^- , tels que, si ϕ est constante modulo $N\widehat{\mathbf{Z}}$, alors

$$\frac{\Gamma(j)}{(2i\pi)^j} L(f, \phi, j) \in \begin{cases} \mathbf{Q}(f, \zeta_N) \cdot \Omega_f^+ & \text{si } 1 \leq j \leq k-1 \text{ et } \phi(-x) = (-1)^j \phi(x), \\ \mathbf{Q}(f, \zeta_N) \cdot \Omega_f^- & \text{si } 1 \leq j \leq k-1 \text{ et } \phi(-x) = (-1)^{j+1} \phi(x). \end{cases}$$

Si $\phi \in \text{LC}(\widehat{\mathbf{Z}}, \mathbf{Q}(f))$, on définit la fonction $\phi \circ (-1)$ par $(\phi \circ (-1))(x) = \phi(-x)$ et on modifie $L(f, \phi, j)$ de manière à le rendre algébrique en posant

$$\widetilde{L}(f, \phi, j) = \frac{1}{2} \cdot \frac{\Gamma(j)}{(2i\pi)^j} \left(\frac{L(f, \phi + (-1)^j \phi \circ (-1), j)}{\Omega_f^+} + \frac{L(f, \phi - (-1)^j \phi \circ (-1), j)}{\Omega_f^-} \right),$$

ce qui permet d'étendre $\phi \mapsto \widetilde{L}(f, \phi, j)$ par linéarité à $\text{LC}(\widehat{\mathbf{Z}}, \mathbf{A})$ pour toute $\mathbf{Q}(f)$ -algèbre \mathbf{A} .

3.1.4. *Caractérisation du système d'Euler de Kato et applications.* On note f^* la conjuguée complexe de f (i.e. $f^*(z) = \overline{f(\bar{z})} = \sum \overline{a_n} q^n$).

THÉORÈME 3.1. — *Il existe un unique élément $\mathbf{z}_{\text{Kato}}(f)$ de $H^1(\mathcal{G}_{\mathbf{Q}}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_f))$ tel que, quels que soient $j \in \{1, \dots, k-1\}$ et $\phi \in \text{LC}(\widehat{\mathbf{Z}}^{(p)}, \mathbf{Q}(f))$, on ait⁷²*

$$\exp^* \left(\int_{\widehat{\mathbf{Z}}^{(p)}} \phi(t \cdot x_p)^j \mathbf{z}_{\text{Kato}}(f) \right) = \widetilde{L}(f^*, \phi, k-j) \cdot f.$$

Soit Λ l'algèbre de groupe complétée de $\text{Gal}(\mathbf{Q}(\zeta_{p^\infty})/\mathbf{Q})$ identifié à \mathbf{Z}_p^* . Si E est une extension finie de \mathbf{Q}_p , alors $E \otimes_{\mathbf{Z}_p} \Lambda = \mathcal{D}_0(\mathbf{Z}_p^*, E)$ est un produit⁷³ d'anneaux principaux, ce qui permet d'associer à un module M de torsion sur $E \otimes_{\mathbf{Z}_p} \Lambda$ sa mesure caractéristique $|M|_\Lambda$ définie, composante par composante, à multiplication près par une unité de $E \otimes_{\mathbf{Z}_p} \Lambda$, comme le produit des diviseurs élémentaires de M .

Si W est une E -représentation de $\mathcal{G}_{\mathbf{Q}}$, soient $H_{\text{Iw}}^i(W) = H^i(\mathcal{G}_{\mathbf{Q}}, \mathcal{D}_0(\mathbf{Z}_p^*, W))$, $i \in \mathbf{N}$, les modules d'Iwasawa associés à W ; ce sont des $E \otimes_{\mathbf{Z}_p} \Lambda$ -modules de type fini, nuls si $i = 0$ ou si $i \geq 3$. Si η est un caractère de Dirichlet de conducteur M premier à p , on note $\mathbf{z}_{\text{Kato}}(f, \eta)$ l'élément de $H_{\text{Iw}}^1(V_{f \otimes \eta})$ défini⁷⁴ par

$$\int_{\mathbf{Z}_p^*} \phi \mathbf{z}_{\text{Kato}}(f, \eta) = \int_{\widehat{\mathbf{Z}}^{(p)}} \eta \phi \mathbf{z}_{\text{Kato}}(f).$$

⁷²Dans cet énoncé, $t = 2i\pi$ désigne le générateur habituel de $\mathbf{Z}_p(1)$, $x_p \in \mathbf{Z}_p^*$ désigne la composante de x sur \mathbf{Z}_p , ce qui fait que, si ϕ est constante modulo $N\widehat{\mathbf{Z}}$, alors $\int_{\widehat{\mathbf{Z}}^{(p)}} \phi(t \cdot x_p)^j \mathbf{z}_{\text{Kato}}(f) \in H^1(\mathcal{G}_{\mathbf{Q}(\zeta_N)}, V_f(j))$. Son image par \exp^* appartient donc à $\mathbf{D}_{\text{dR}}^0(V_f(j)) \otimes_{\mathbf{Q}_p(f)} \mathbf{Q}_p(f, \zeta_N) = \mathbf{Q}_p(f, \zeta_N) \cdot f$, ce qui donne un sens à l'égalité du théorème, l'identification de $\mathbf{D}_{\text{dR}}^0(V_f)$ à $\mathbf{Q}_p(f) \cdot f$ se faisant via les deux applications \exp^* comme expliqué dans l'introduction (cf. n° 0.6.5).

⁷³Sur les caractères du sous-groupe de torsion de \mathbf{Z}_p^* .

⁷⁴Dans cette formule, on voit η comme une fonction sur $\widehat{\mathbf{Z}}$ localement constante modulo M .

THÉORÈME 3.2. — (i) $H_{\text{Iw}}^1(V_{f \otimes \eta})$ est un $\mathbf{Q}_p(f, \eta) \otimes \Lambda$ -module libre de rang 1.

(ii) $H_{\text{Iw}}^2(V_{f \otimes \eta})$ et $H_{\text{Iw}}^1(V_{f \otimes \eta})/\mathbf{z}_{\text{Kato}}(f, \eta)$ sont des $\mathbf{Q}_p(f, \eta) \otimes_{\mathbf{z}_p} \Lambda$ -modules de torsion, et $|H_{\text{Iw}}^2(V_{f \otimes \eta})|_{\Lambda}$ divise le produit de $|H_{\text{Iw}}^1(V_{f \otimes \eta})/\mathbf{z}_{\text{Kato}}(f, \eta)|_{\Lambda}$ et $|H^2(\mathcal{G}_{\mathbf{Q}_p}, \Lambda \otimes V_{f \otimes \eta})|_{\Lambda}$.

REMARQUE 3.3. — (i) On conjecture que l'on a en fait

$$|H_{\text{Iw}}^2(V_{f \otimes \eta})|_{\Lambda} = |H_{\text{Iw}}^1(V_{f \otimes \eta})/\mathbf{z}_{\text{Kato}}(f, \eta)|_{\Lambda} \cdot |H^2(\mathcal{G}_{\mathbf{Q}_p}, \Lambda \otimes V_{f \otimes \eta})|_{\Lambda},$$

ce qui est une des formulations de la « conjecture principale ». Dans cette formulation en termes « d'éléments zêta », spécialisation aux formes modulaires d'une conjecture générale de Kato [98], les fonctions L p -adiques n'apparaissent pas. Le lien avec la formulation de Perrin-Riou [141] s'établit en montrant que la fonction L p -adique est l'image de $\mathbf{z}_{\text{Kato}}(f)$ par l'exponentielle de Perrin-Riou (c'est le sujet du chapitre 4; voir en particulier les th. 4.4, 4.11 et 4.15). Remarquons quand même que la formulation de la conjecture principale en termes d'éléments zêta n'impose aucune propriété de « bonne réduction » à la forme modulaire f , alors qu'il reste du travail pour définir une fonction L p -adique en toute généralité.

(ii) Si W est une E-représentation de $\mathcal{G}_{\mathbf{Q}_p}$, le module $H^2(\mathcal{G}_{\mathbf{Q}_p}, \Lambda \otimes W)$ est isomorphe à⁷⁵ $H^0(\mathcal{G}_{\mathbf{Q}_p(\zeta_{p^\infty})}, W^*(1))^*$; il est donc nul en général.

(iii) Si le conducteur de $f \otimes \eta$ n'est pas divisible par N , la divisibilité ci-dessus n'est pas optimale : il peut s'être introduit des facteurs d'Euler en les places divisant N .

(iv) Sous des hypothèses supplémentaires concernant la représentation (satisfaites pour presque tout nombre premier si f n'est pas de type CM), on peut prouver des résultats du même genre sans avoir à inverser p .

(v) La démonstration des théorèmes 3.1 et 3.2 est un peu tortueuse. On projette les systèmes d'Euler $\mathbf{z}_{\text{Kato}}(k, j)$, pour $1 \leq j \leq k - 1$, sur la composante correspondant à f , ce qui permet de construire toute une famille (cf. § 3.3) de systèmes d'Euler pour V_f et ses tordues. On calcule l'image par l'application \exp^* des éléments ainsi construits en utilisant la loi de réciprocité explicite de Kato et en projetant sur la composante correspondant à f , ce qui nous conduit à calculer le produit scalaire de Petersson de f avec un produit de série d'Eisenstein; on utilise la méthode de Rankin (cf. § 3.2) pour ce faire, et le résultat fait intervenir les valeurs de fonctions L du théorème 3.1. Ces systèmes d'Euler ne sont pas optimaux⁷⁶, mais ils permettent, en utilisant la technique des dérivées de Kolyvagin [106, 99, 142, 158, 115] de démontrer que $H_{\text{Iw}}^1(V_{f \otimes \eta})$ est un $\mathbf{Q}_p(f, \eta) \otimes \Lambda$ -module libre de rang 1 pour tout caractère η . On construit⁷⁷ alors $\mathbf{z}_{\text{Kato}}(f, \eta)$ comme combinaison linéaire, à coefficients dans l'anneau des fractions de $\mathbf{Q}(f, \eta) \otimes_{\mathbf{z}_p} \Lambda$ des systèmes d'Euler précédents⁷⁸; le problème étant de montrer que l'élément ainsi construit est encore dans

⁷⁵ et pas à $H^0(\mathcal{G}_{\mathbf{Q}_p(\zeta_{p^\infty})}, W(-1))$ comme il est affirmé dans [54]. . .

⁷⁶ Il y a des tas de facteurs parasites qui interviennent, par exemple à cause de l'introduction de c et d pour définir $\mathbf{z}_{\text{Kato}, c, d}(k, j)$, ou de la disparition des facteurs d'Euler en les places divisant N .

⁷⁷ Pour cette partie de l'argument, le lecteur est invité à consulter le §13 de [101].

⁷⁸ tordus à la Soulé pour passer de $V_f(j)$ à V_f .

$H_{\text{Iw}}^1(V_{f \otimes \eta})$, ce qui se fait en utilisant les calculs précédents pour l'application \exp^* ainsi que le fait que l'on a une infinité de choix à notre disposition, ce qui permet de montrer que l'on n'a pas introduit de pôle parasite. Une fois que l'on a rendu le système d'Euler optimal, on peut réutiliser la technique de Kolyvagin pour démontrer le (ii) du théorème 3.2.

3.2. La méthode de Rankin-Selberg

3.2.1. *Le cas général.* Soit $\Gamma = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbf{Z}) \mid b \equiv c \equiv 0 \pmod{N} \right\}$. Soit ε_1 et ε_2 des caractères de Dirichlet modulo N (pas nécessairement primitifs). Soient $k \geq 2$, $1 \leq j \leq k - 1$, et

$$f = \sum_{n \in \frac{1}{N}\mathbf{Z}, n > 0} a_n q^n \in S_k(\Gamma, \varepsilon_1) \quad \text{et} \quad g = \sum_{n \in \frac{1}{N}\mathbf{Z}, n \geq 0} b_n q^n \in M_{k-j}(\Gamma, \varepsilon_2),$$

des forme propres pour tous les opérateurs $T(\ell)$, ℓ premier ne divisant pas N . On a donc

$$\sum_{n > 0} \frac{a_n}{n^s} = \left(\sum_{n \in \mathbf{Z}[\frac{1}{N}]^*} \frac{a_n}{n^s} \right) \cdot \prod_{\ell \nmid N} \frac{1}{((1 - \alpha_{\ell,1} \ell^{-s})(1 - \alpha_{\ell,2} \ell^{-s}))}, \quad \text{avec } \alpha_{\ell,1} \alpha_{\ell,2} = \varepsilon_1(\ell) \ell^{k-1},$$

$$\sum_{n > 0} \frac{b_n}{n^s} = \left(\sum_{n \in \mathbf{Z}[\frac{1}{N}]^*} \frac{b_n}{n^s} \right) \cdot \prod_{\ell \nmid N} \frac{1}{((1 - \beta_{\ell,1} \ell^{-s})(1 - \beta_{\ell,2} \ell^{-s}))}, \quad \text{avec } \beta_{\ell,1} \beta_{\ell,2} = \varepsilon_2(\ell) \ell^{k-j-1}.$$

Soient $D(f, g, s)$ la série de Dirichlet définie par

$$D(f, g, s) = L(\bar{\varepsilon}_1 \varepsilon_2, j + 2(s - k + 1)) \cdot \sum_{n > 0} \frac{\bar{a}_n b_n}{n^s}$$

et $E_s(\tau)$ la série d'Eisenstein (de poids j) définie par

$$E_s(\tau) = \sum_{c \equiv d - 1 \equiv 0 \pmod{N}} \frac{1}{(c\tau + d)^j} \left(\frac{\text{Im } \tau}{|c\tau + d|^2} \right)^{s+1-k}.$$

PROPOSITION 3.4. — *Sous les hypothèses ci-dessus, on a*

$$D(f, g, s) = \left(\sum_{n \in \mathbf{Z}[\frac{1}{N}]^*} \frac{\bar{a}_n b_n}{n^s} \right) \cdot \prod_{\ell \nmid N} \frac{1}{\left(1 - \frac{\bar{\alpha}_{\ell,1} \beta_{\ell,1}}{\ell^s}\right) \left(1 - \frac{\bar{\alpha}_{\ell,1} \beta_{\ell,2}}{\ell^s}\right) \left(1 - \frac{\bar{\alpha}_{\ell,2} \beta_{\ell,1}}{\ell^s}\right) \left(1 - \frac{\bar{\alpha}_{\ell,2} \beta_{\ell,2}}{\ell^s}\right)}.$$

$$\frac{\Gamma(s)}{(4\pi)^s} D(f, g, s) = \frac{[\mathbf{SL}_2(\mathbf{Z}) : \Gamma(N)]}{N} \langle f, g E_s \rangle.$$

Démonstration. — La première formule se démontre facteur d'Euler par facteur d'Euler et est un petit exercice de sommation de séries géométriques. Pour démontrer le (ii), on utilise le fait que f et g sont des formes modulaires pour Γ pour écrire $\langle f, g E_s \rangle$ sous la

forme $\frac{1}{[\Gamma:\Gamma(\mathbf{N})]} \langle f, gE'_s \rangle$, avec (avec $\Gamma_\infty = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, c = 0 \right\}$) :

$$\begin{aligned} E'_s(\tau) &= \left(\sum_{(e, \mathbf{N})=1} \frac{\bar{\varepsilon}_1 \varepsilon_2(e)}{e^{j+2(s-k+1)}} \right) \cdot \left(\sum_{\substack{\mathbf{N}|c \\ (c,d)=1}} \frac{\bar{\varepsilon}_1 \varepsilon_2(d)}{(c\tau + d)^j} \left(\frac{\text{Im } \tau}{|c\tau + d|^2} \right)^{s+1-k} \right) \\ &= L(\bar{\varepsilon}_1 \varepsilon_2, j + 2(s - k + 1)) \cdot \sum_{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_\infty \setminus \Gamma} \frac{\bar{\varepsilon}_1 \varepsilon_2(d)}{(c\tau + d)^j} \left(\frac{\text{Im } \tau}{|c\tau + d|^2} \right)^{s+1-k}. \end{aligned}$$

Par ailleurs, on a

$$\begin{aligned} \frac{\Gamma(s)}{(4\pi)^s} D(f, g, s) &= L(\bar{\varepsilon}_1 \varepsilon_2, j + 2(s - k + 1)) \cdot \frac{1}{\mathbf{N}} \cdot \int_0^{+\infty} \left(\int_0^{\mathbf{N}} \bar{f}g \, dx \right) y^s \frac{dy}{y} \\ &= L(\bar{\varepsilon}_1 \varepsilon_2, j + 2(s - k + 1)) \cdot \frac{1}{\mathbf{N}} \cdot \int_{\Gamma_\infty \setminus \mathcal{H}} \bar{f}g y^{s+1} \frac{dx \, dy}{y^2} \\ &= L(\bar{\varepsilon}_1 \varepsilon_2, j + 2(s - k + 1)) \cdot \frac{1}{\mathbf{N}} \cdot \int_{\Gamma \setminus \mathcal{H}} \sum_{\gamma \in \Gamma_\infty \setminus \Gamma} \bar{f}(\gamma\tau) g(\gamma\tau) (\text{Im } \gamma\tau)^{s+1} \frac{dx \, dy}{y^2} \\ &= \frac{1}{\mathbf{N}} \cdot \int_{\Gamma \setminus \mathcal{H}} \bar{f}g E'_s y^k \frac{dx \, dy}{y^2}, \end{aligned}$$

le passage de la seconde à la troisième ligne utilisant l'invariance de la forme volume hyperbolique $\frac{dx \, dy}{y^2}$, et le passage de la troisième à la quatrième utilisant les formules de transformations

$$\begin{aligned} f\left(\frac{a\tau + b}{c\tau + d}\right) &= \varepsilon_1(d)(c\tau + d)^k f(\tau), \quad g\left(\frac{a\tau + b}{c\tau + d}\right) = \varepsilon_2(d)(c\tau + d)^{k-j} g(\tau) \\ \text{et } \text{Im}\left(\frac{a\tau + b}{c\tau + d}\right) &= \frac{\text{Im } \tau}{|c\tau + d|^2}, \quad \text{si } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \end{aligned}$$

COROLLAIRE 3.5. — *Sous les hypothèses ci-dessus, il existe une constante $C(k, j, \mathbf{N}) \neq 0$ telle que l'on ait*

$$D(f, g, k - 1) = C(k, j, \mathbf{N}) \langle f, gE_{0, \frac{1}{\mathbf{N}}}^{(j)} \rangle.$$

3.2.2. Projection d'un produit de deux séries d'Eisenstein. Si $\chi_1 : (\mathbf{Z}/M_1\mathbf{Z})^*$ est un caractère de Dirichlet modulo M_1 (pas nécessairement primitif), et $\chi_2 : (\mathbf{Z}/M_2\mathbf{Z})^*$ est un caractère de Dirichlet de conducteur M_2 , et si $\chi_1\chi_2(-1) = (-1)^{k-j}$, on pose

$$F_{\chi_1, \chi_2}^{(k-j)} = \frac{1}{2} \sum_{a=1}^{M_1} \sum_{b=1}^{M_2} \chi_1(a)\chi_2(b) F_{\frac{a}{M_1}, \frac{b}{M_2}}^{(k-j)}.$$

Un petit calcul montre que l'on a

$$(F_{\chi_1, \chi_2}^{(k-j)})|_{k-j} \gamma = \chi_1 \chi_2^{-1}(d) F_{\chi_1, \chi_2}^{(k-j)}, \quad \text{si } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}_2(\mathbf{Z}), \text{ et } b \equiv 0 [M_1], c \equiv 0 [M_2].$$

D'autre part, en utilisant la proposition 1.3, on peut calculer le q -développement $\sum_{n \in \mathbf{Q}_+} c_n q^n$ de $F_{\chi_1, \chi_2}^{(k-j)}$, et obtenir la formule

$$\sum_{n \in \mathbf{Q}_+^*} \frac{c_n}{n^s} = M_1^{s-(k-j)+1} L(\chi_1, s - (k-j) + 1) \cdot G(\chi_2) L(\chi_2^{-1}, s),$$

où $G(\chi_2) = \sum_{b=1}^{M_2} \chi_2(b) e^{2i\pi b/M_2}$ est la somme de Gauss associée à χ_2 .

COROLLAIRE 3.6. — *Si f est une forme primitive de $S_k(\Gamma_0(N), \varepsilon)$, si M_1 est divisible par N , et si M est un multiple de N , M_1 et M_2 , alors il existe une constante $C'(k, j, M) \neq 0$ telle que*

$$\langle f, F_{\chi_1, \chi_2}^{(k-j)} E_{0, \frac{1}{M}}^{(j)} \rangle = C'(k, j, M) \cdot L(f^*, \chi_1, j) \cdot \chi_2^{-1}(M_1) G(\chi_2) L(f^*, \chi_2^{-1}, k-1).$$

Démonstration. — Si $f = \sum_{n \in \mathbf{Q}_+} a_n q^n$, on a $a_n = 0$ si $n \notin \mathbf{N}$. En prenant en compte cette remarque, la première formule de la prop. 3.4, et le calcul du q -développement de $F_{\chi_1, \chi_2}^{(k-j)}$ effectué ci-dessus, on obtient

$$D(f, F_{\chi_1, \chi_2}^{(k-j)}, s) = \chi_2^{-1}(M_1) L(f^*, \chi_1, s - (k-j) + 1) \cdot G(\chi_2) L(\chi_2^{-1}, s),$$

et on conclut en utilisant le corollaire 3.5.

3.3. Projection du système d'Euler de Kato

Soit $\widehat{\Gamma} \subset \mathbf{SL}_2(\widehat{\mathbf{Z}})$ un sous-groupe de congruence; soit $\widetilde{\Gamma}$ son image inverse dans $\Pi_{\mathbf{Q}}$ et soit $\overline{\Gamma}$ l'intersection de $\widetilde{\Gamma}$ avec $\Pi_{\overline{\mathbf{Q}}}$. On suppose que $\det : \widehat{\Gamma} \rightarrow \widehat{\mathbf{Z}}^*$ est surjectif et donc que $\overline{\mathbf{Q}}^{\widetilde{\Gamma}} = \mathbf{Q}$. Si $\phi \in H^0(\widetilde{\Gamma}, \mathrm{LC}(\mathbf{M}_2(\widehat{\mathbf{Z}}^{(p)}), \mathbf{Z}))$, on définit un élément $\mathbf{z}_\phi(f, j) \in H^1(\mathcal{G}_{\mathbf{Q}}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_f(k-j)))$ de la manière suivante. Soit $\mathbf{z}_{\phi,0}(k, j) \in H^2(\widetilde{\Gamma}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_{k,j}))$ la mesure définie par

$$\int_{\widehat{\mathbf{Z}}^{(p)}} \psi \mathbf{z}_{\phi,0}(k, j) = \int_{\mathbf{M}_2(\widehat{\mathbf{Z}}^{(p)})} \psi(\det x) \phi(x) \mathbf{z}_{\mathrm{Kato}}(k, j).$$

Soit $\mathbf{z}_{\phi,1}(k, j)$ l'image de $\mathbf{z}_{\phi,0}(k, j)$ dans⁷⁹

$$H^1(\mathcal{G}_{\mathbf{Q}}, H^1(\overline{\Gamma}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_{k,j}))) \cong H^1(\mathcal{G}_{\mathbf{Q}}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, H^1(\overline{\Gamma}, V_{k,j}))),$$

et soit $\mathbf{z}_\phi(f, j)$ la projection de $\mathbf{z}_{\phi,1}(k, j)$ sur une des composantes de $H^1(\overline{\Gamma}, V_{k,j})_{\pi_f} \cong V_f(k-j)^r$, où r est un entier dépendant de $\widehat{\Gamma}$.

On déduit alors de la proposition 1.10 le fait que la famille $c_M = \int_{(1+M\widehat{\mathbf{Z}}) \cap \widehat{\mathbf{Z}}^{(p)}} \mathbf{z}_\phi(f, j)$, $M \geq 1$ forme un système d'Euler pour $V_f(k-j)$. D'autre part, en utilisant la loi de réciprocité explicite de Kato (th. 1.9) et le cor. 3.6, on montre que l'on peut choisir ϕ de manière à ce que ce système d'Euler soit non identiquement nul (et même, si $L(f^*, j) \neq 0$,

⁷⁹L'isomorphisme $H^1(\overline{\Gamma}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_{k,j})) \cong \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, H^1(\overline{\Gamma}, V_{k,j}))$ provient de ce que $\overline{\Gamma}$ agit trivialement sur $\widehat{\mathbf{Z}}^{(p)}$ et de ce que l'on peut échanger cohomologie et limite projective car on est sous les conditions de Mittag-Leffler, la cohomologie de $\overline{\Gamma}$ dans un module fini étant un groupe fini.

de telle sorte que c_1 ne soit pas de torsion), ce qui permet de faire marcher la machine des dérivées de Kolyvagin.

4. FONCTIONS L p -ADIQUES DES FORMES MODULAIRES

Dans ce chapitre, on explique comment utiliser la machine de Perrin-Riou (i.e. son application exponentielle [140, 144] construite par interpolation des exponentielles de Bloch-Kato) pour construire la fonction L p -adique attachée à une forme modulaire à partir du système d'Euler de Kato.

La construction de la fonction L p -adique via les symboles modulaires est rappelée au § 4.2. La normalisation adoptée est un peu différente de celle de [120], ce qui permet d'obtenir une distribution sur $\widehat{\mathbf{Z}}$ au lieu d'une famille de distributions sur $(\mathbf{Z}/M\mathbf{Z}) \times \mathbf{Z}_p$, $M \geq 1$ premier à p . La description de la machine de Perrin-Riou fait l'objet du § 4.3. Le point de vue adopté est différent de l'approche originale de Perrin-Riou : il utilise la théorie des (φ, Γ) -modules de Fontaine [75, 37]. L'intérêt de cette théorie est de fournir une description explicite (et quasiment gratuite : th. 4.8 et 4.9) du module d'Iwasawa $H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\mathbf{Z}_p^*, V))$ associé à une représentation *arbitraire* V de $\mathcal{G}_{\mathbf{Q}_p}$, grâce à une application Exp^* . Si V est cristalline ou semi-stable, on déduit de lois de réciprocités explicites le fait que l'application Exp^* est (à normalisation près) l'inverse de l'application exponentielle de Perrin-Riou. L'inconvénient d'utiliser les (φ, Γ) -modules est que l'on atterrit dans un module $\mathbf{D}^\dagger(V)^{\psi=1}$ qui est un petit peu mystérieux en général, et le gros du travail consiste à analyser la structure de ce module suivant le type de la représentation V . Cette analyse fait l'objet des §§ 4.4-4.6. Dans le cas semi-stable, on en déduit (cf. n° 4.5.1) une formule purement locale pour la dérivée, en un « zéro supplémentaire », d'une fonction L p -adique, ce qui fournit une démonstration⁸⁰ (cf. th. 4.17) de la conjecture de Mazur-Tate-Teitelbaum pour la dérivée première de la fonction L p -adique d'une forme modulaire dans le cas d'un zéro supplémentaire.

4.1. Mesures, distributions et fonctions analytiques

4.1.1. *Anneaux de séries de Laurent.* Si L est un sous-corps de \mathbf{C}_p , on note \mathcal{R}_L^+ (resp. \mathcal{E}_L^+) l'anneau des fonctions analytiques (resp. analytiques bornées) sur le disque $\{x, v_p(x) \geq 0\}$ et \mathcal{R}_L^\dagger (resp. \mathcal{E}_L^\dagger) l'anneau des fonctions analytiques (resp. analytiques bornées) sur la couronne d'épaisseur nulle et de rayon 1. Un élément d'un des anneaux ci-dessus est défini par une série de Laurent $\sum_{k \in \mathbf{Z}} a_k T^k$, où $a_k \in L$ quel que soit $k \in \mathbf{Z}$, et on a les équivalences suivantes :

⁸⁰Cette démonstration est une traduction dans le langage des (φ, Γ) -modules de celle qu'en a donnée, modulo l'existence d'un système d'Euler relié à la fonction L p -adique, Perrin-Riou [145]; c'est une variante de celle de Kato-Kurihara et Tsuji [103]. (Cf. note 94 pour des compléments.)

- (i) $\sum_{k \in \mathbf{Z}} a_k T^k \in \mathcal{E}_L^+$ si et seulement si $a_k = 0$ si $k \leq -1$ et si la suite de terme général $v_p(a_k)$ est minorée ;
- (ii) $\sum_{k \in \mathbf{Z}} a_k T^k \in \mathcal{R}_L^+$ si et seulement si $a_k = 0$ si $k \leq -1$ et si $\liminf_{k \rightarrow +\infty} \frac{1}{k} v_k(a_k) \geq 0$;
- (iii) $\sum_{k \in \mathbf{Z}} a_k T^k \in \mathcal{E}_L^\dagger$ si et seulement si $\liminf_{k \rightarrow -\infty} \frac{1}{|k|} v_p(a_k) > 0$ et si la suite de terme général $v_p(a_k)$ est minorée ;
- (iv) $\sum_{k \in \mathbf{Z}} a_k T^k \in \mathcal{R}_L$ si et seulement si $\liminf_{k \rightarrow -\infty} \frac{1}{|k|} v_p(a_k) > 0$ et $\liminf_{k \rightarrow +\infty} \frac{1}{k} v_k(a_k) \geq 0$.

On a bien évidemment $\mathcal{E}_L^\dagger \cap \mathcal{R}_L^+ = \mathcal{E}_L^+$.

Si r est un nombre réel, un élément $\sum_{k \in \mathbf{Z}} a_k T^k$ de \mathcal{R}_L est d'ordre r si la suite de terme général $v_p(a_k) + r \log k$, $k \geq 1$, est minorée. On remarquera qu'un élément de \mathcal{R}_L est d'ordre 0 si et seulement s'il appartient à \mathcal{E}_L^\dagger , ce qui fournit un moyen de récupérer \mathcal{E}_L^\dagger à l'intérieur de \mathcal{R}_L .

PROPOSITION 4.1. — *Si h est un entier supérieur ou égal à la partie entière de r , alors quel que soit $n_0 \in \mathbf{N}$, un élément x de \mathcal{R}_L d'ordre r est entièrement déterminé par son développement de Taylor à l'ordre h en les $\zeta_{p^n} - 1$, pour $n \geq n_0$.*

4.1.2. *Mesures et distributions sur \mathbf{Z}_p .* Si X est un ouvert compact de \mathbf{Z}_p , soit $\mathcal{C}^0(X, L)$ l'espace des fonctions continues sur X à valeurs dans L ; on munit cet espace de la norme du sup. qui est bien définie puisque X est compact.

Soit aussi $LA(X, L)$ l'espace des fonctions localement analytiques sur X à valeurs dans L ; si $f \in LA(X, L)$ et si $a \in X$, alors il existe un voisinage de a tel que l'on puisse écrire $f(x)$ sous la forme $f(x) = \sum_{k=0}^{+\infty} c_k(x - a)^k$ dans ce voisinage. L'espace $LA(X, L)$ est naturellement une limite inductive d'espaces de Banach.

Si V est un L -espace vectoriel topologique, on note $\mathcal{D}_0(X, V)$ l'espace des mesures sur X à valeurs dans V (i.e. l'espace des applications linéaires continues de $\mathcal{C}^0(X, L)$ dans V) ; si $f \in \mathcal{C}^0(X, L)$ et $\mu \in \mathcal{D}_0(X, V)$, on note en général $\int_X f \mu$ la valeur de μ sur f . On note aussi $\mathcal{D}(X, V)$ l'espace des distributions sur X à valeurs dans V (i.e. l'espace des applications linéaires continues de $LA(X, L)$ dans V) ; comme précédemment, si $f \in LA(X, L)$ et $\mu \in \mathcal{D}(X, V)$, on note en général $\int_X f \mu$ la valeur de μ sur f . Une fonction localement analytique étant continue, on a une application naturelle de $\mathcal{D}_0(X, V)$ dans $\mathcal{D}(X, V)$ qui est injective, ce qui permet de considérer une mesure comme une distribution d'un type particulier.

A une distribution μ , on associe sa *transformée d'Amice* $\mathcal{A}_\mu(T)$ qui est une série formelle définie par

$$\mathcal{A}_\mu(T) = \sum_{n=0}^{+\infty} \left(\int_X \binom{x}{n} \mu \right) \cdot T^n = \int_X (1 + T)^x \mu.$$

La transformée d'Amice fournit une caractérisation particulièrement agréable [2, 3, 56] des distributions sur \mathbf{Z}_p et, en particulier, permet de construire des distributions à partir de séries entières⁸¹.

⁸¹Voir [165] pour les distributions sur l'anneau des entiers d'une extension finie de \mathbf{Q}_p .

THÉORÈME 4.2. — *L'application $\mu \mapsto \mathcal{A}_\mu(\Gamma)$ induit un isomorphisme de $\mathcal{D}(\mathbf{Z}_p, \mathbf{L})$ sur $\mathcal{R}_\mathbf{L}^+$ et de $\mathcal{D}_0(\mathbf{Z}_p, \mathbf{L})$ sur $\mathcal{E}_\mathbf{L}^+$.*

Une distribution sur X ouvert compact de \mathbf{Z}_p est d'ordre r si sa transformée d'Amice est d'ordre r . On note $\mathcal{D}_r(X, V)$ l'espace des distributions d'ordre r sur X à valeurs dans V . En utilisant l'isomorphisme du théorème 4.2, on peut traduire les résultats sur les fonctions analytiques par :

(i) une distribution est d'ordre 0 si et seulement si c'est une mesure (il n'y a donc pas de conflit de notation...);

(ii) Si h est un entier supérieur ou égal à la partie entière de r , alors quel que soit $n_0 \in \mathbf{N}$, une distribution d'ordre r est entièrement déterminée par les intégrales $\int_{\mathbf{Z}_p} x^i \zeta_{p^n}^x$, pour $n \geq n_0$ et $0 \leq i \leq h$.

Finalement, on a le résultat suivant [3, 56, 190, 56] qui permet de construire une distribution d'ordre r en connaissant les valeurs des intégrales $\int_{\mathbf{Z}_p} x^i \zeta_{p^n}^x$, pour $n \geq n_0$ et $0 \leq i \leq h$.

PROPOSITION 4.3. — *Soit $r \in \mathbf{R}$. Si h est un entier supérieur ou égal à la partie entière de r , et si μ est une forme linéaire à valeurs dans \mathbf{L} sur les fonctions localement polynomiales de degré $\leq h$ telle qu'il existe une constante C telle que l'on ait*

$$\left| \int_{a+p^n \mathbf{Z}_p} (x-a)^h \right| \leq C \cdot p^{n(r-h)} \quad \text{quels que soient } a \in \mathbf{Z}_p \text{ et } n \in \mathbf{N},$$

alors μ s'étend de manière unique en une distribution d'ordre r sur \mathbf{Z}_p .

4.1.3. *Distributions sur $\widehat{\mathbf{Z}}$.* Si X est un ouvert compact de $\widehat{\mathbf{Z}}$, une fonction $f : X \rightarrow \mathbf{L}$ est dite localement analytique si tout point $a = (a_p, a^{[p]}) \in X$ possède un voisinage sur lequel on peut écrire $f(x)$ sous la forme $f(x) = \sum_{k=0}^{+\infty} c_k (x_p - a_p)^k$. Une fonction localement analytique sur X est donc localement constante par rapport à $x^{[p]}$ et localement analytique par rapport à x_p . On note $\text{LA}(X, \mathbf{L})$ l'espace des fonctions localement analytiques sur X à valeurs dans \mathbf{L} et $\mathcal{D}(X, V)$ l'espace des distributions sur X à valeurs dans V . Si $X = X_p \times X^{[p]}$, où X_p (resp. $X^{[p]}$) est un ouvert compact de \mathbf{Z}_p (resp. de $\widehat{\mathbf{Z}}^{[p]}$), alors écrivant $X^{[p]}$ comme limite projective d'ensembles finis, on obtient un isomorphisme

$$\mathcal{D}(X, V) \cong \mathcal{D}_{\text{alg}}(X^{[p]}, \mathcal{D}(X_p, V)).$$

Soit $\mathcal{D}_r(X, V) = \mathcal{D}_{\text{alg}}(X^{[p]}, \mathcal{D}_r(X_p, V)) \subset \mathcal{D}(X, V)$ l'ensemble des distributions d'ordre r . De manière équivalente, une distribution μ sur $\widehat{\mathbf{Z}}$ est d'ordre r si et seulement si la distribution $\phi_p \mapsto \int_{\widehat{\mathbf{Z}}} \phi_p(x_p) \phi^{[p]}(x^{[p]}) \mu$ est une distribution d'ordre r sur \mathbf{Z}_p , quelle que soit la fonction localement constante $\phi^{[p]}$ sur $X^{[p]}$.

4.2. Formes modulaires et fonctions \mathbf{L} p -adiques

Soit f une forme primitive de poids $k \geq 2$, de caractère ε et de niveau N . Le facteur d'Euler en p de $L(f, s)$ peut se mettre sous la forme $(1 - \alpha p^{-s})(1 - \beta p^{-s})$, où α et β sont des

entiers algébriques (éventuellement nuls). On suppose dans toute la suite $v_p(\alpha) < k - 1$; en particulier $\alpha \neq 0$.

Soit f_α la forme modulaire définie par $f_\alpha(z) = f(z) - \beta f(pz)$. Si $\beta = 0$, on a bien évidemment $f_\alpha = f$, et si $\beta \neq 0$, alors f_α est de niveau Np , propre pour tous les opérateurs de Hecke, de valeur propre α pour l'opérateur T_p de niveau Np (i.e. l'opérateur U_p). Soit $f_\alpha = \sum_{n=1}^{+\infty} a_n q^n$ le q -développement de f_α . On a $a_{np} = \alpha a_n$, si $n \in \mathbf{Z}$, ce qui permet de prolonger l'application $n \mapsto a_n$ à $\mathbf{Z}[\frac{1}{p}]$ en forçant l'équation fonctionnelle $a_{np} = \alpha a_n$ si $n \in \mathbf{Z}[\frac{1}{p}]$.

Si $\phi \in \text{LC}_c(\mathbf{Q}_p \times \widehat{\mathbf{Z}}^{[p]}, \overline{\mathbf{Q}})$, soit $L(f_\alpha, \phi, s) = \sum_{n \in \mathbf{Z}[\frac{1}{p}]} \phi(n) a_n n^{-s}$. Si $1 \leq j \leq k - 1$, on définit $\widetilde{L}(f_\alpha, \phi, j) \in \overline{\mathbf{Q}}$ comme précédemment.

Si $\phi \in \text{LC}_c(\mathbf{Q}_p, \overline{\mathbf{Q}})$ est constante modulo p^n , on définit sa transformée de Fourier $\hat{\phi} \in \text{LC}_c(\mathbf{Q}_p, \overline{\mathbf{Q}})$ par la formule (le résultat ne dépend pas du choix de n)

$$\hat{\phi}(x) = p^{-n} \sum_{y \bmod p^n} \phi(x) e^{-2i\pi xy},$$

où $e^{-2i\pi xy}$ est la racine de l'unité d'ordre une puissance de p obtenue en utilisant l'isomorphisme $\mathbf{Q}_p/\mathbf{Z}_p \cong \mathbf{Z}[\frac{1}{p}]/\mathbf{Z}$. Si $\chi : \mathbf{Z}_p^* \rightarrow \overline{\mathbf{Q}}$ est un caractère de Dirichlet de conducteur p^n , on a

$$\hat{\chi}(x) = \begin{cases} \frac{1}{G(\chi^{-1})} \chi^{-1}(p^n x) & \text{si } n \geq 1, \\ \mathbf{1}_{\mathbf{Z}_p}(x) - \frac{1}{p} \mathbf{1}_{p^{-1}\mathbf{Z}_p}(x) & \text{si } n = 0. \end{cases}$$

Si $\phi_p \in \text{LC}_c(\mathbf{Q}_p, \overline{\mathbf{Q}})$, si $\phi^{[p]} \in \text{LC}(\widehat{\mathbf{Z}}^{[p]}, \overline{\mathbf{Q}})$ et si $\phi \in \text{LC}_c(\mathbf{Q}_p \times \widehat{\mathbf{Z}}^{[p]}, \overline{\mathbf{Q}})$ est donnée par la formule $\phi(x) = \phi_p(x_p) \cdot \phi^{[p]}(x^{[p]})$, on définit sa p -transformée de Fourier $\mathcal{F}_p \phi$ comme l'élément de $\text{LC}_c(\mathbf{Q}_p \times \widehat{\mathbf{Z}}^{[p]}, \overline{\mathbf{Q}})$ donné par la formule

$$\mathcal{F}_p \phi(x) = \hat{\phi}_p(x_p) \cdot \phi^{[p]}(x^{[p]}).$$

On étend \mathcal{F}_p en une application $\overline{\mathbf{Q}}$ -linéaire de $\text{LC}_c(\mathbf{Q}_p \times \widehat{\mathbf{Z}}^{[p]}, \overline{\mathbf{Q}})$ dans lui-même.

THÉORÈME 4.4. — *Si $v_p(\alpha) < k - 1$, il existe une unique⁸² distribution $\mu_{f,\alpha}$ d'ordre $v_p(\alpha)$ sur $\widehat{\mathbf{Z}}$, telle que, quel que soit $\phi \in \text{LC}(\widehat{\mathbf{Z}}, \overline{\mathbf{Q}})$ et $1 \leq j \leq k - 1$, on ait*

$$\int_{\widehat{\mathbf{Z}}} \phi(x) x_p^{j-1} \mu_{f,\alpha} = \widetilde{L}(f_\alpha, \mathcal{F}_p \phi, j).$$

De plus, quel que soit $\phi \in \text{LA}(\widehat{\mathbf{Z}}, \mathbf{C}_p)$, on a

$$\int_{p\widehat{\mathbf{Z}}} \phi(p^{-1}x_p, px^{[p]}) \mu_{f,\alpha} = \frac{p}{\alpha} \int_{\widehat{\mathbf{Z}}} \phi(x) \mu_{f,\alpha}.$$

Démonstration. — L'ingrédient principal pour démontrer ce théorème [4, 111, 190, 120] est la théorie des symboles modulaires qui permet de montrer qu'il existe un $\mathcal{O}_{\mathbf{Q}(f)}$ -réseau Λ de $\mathbf{Q}(f)\Omega_f^+ + \mathbf{Q}(f)\Omega_f^- \subset \mathbf{C}$ tel que $\int_r^{i\infty} P(z)f(z) dz \in \Lambda$ quels que soient $r \in \mathbf{Q}$ et $P \in \mathcal{O}_{\mathbf{Q}(f)}[\mathbf{X}]$ de degré $\leq k - 2$.

⁸²Cf. rem 4.12 pour le cas $v_p(\alpha) = k - 1$.

On note $\lambda_{f,\alpha}$ la distribution sur $\widehat{\mathbf{Z}}^{(p)}$ obtenue en multipliant la restriction de $\mu_{f,\alpha}$ à $\widehat{\mathbf{Z}}^{(p)}$ par la fonction x_p^{-1} . On définit la fonction L p -adique de f (associée à α) comme la fonction qui à un caractère⁸³ χ de $\widehat{\mathbf{Z}}^{(p)}$, localement analytique, associe la valeur

$$L_{p,\alpha}(f, \chi) = \int_{\widehat{\mathbf{Z}}^{(p)}} \chi(x) \lambda_{f,\alpha}.$$

Si χ est un tel caractère, on définit une fonction $L_{p,\alpha}(f, \chi, s)$ de la variable $s \in \mathbf{Z}_p$, en posant

$$L_{p,\alpha}(f, \chi, s) = L_{p,\alpha}(f, \chi \cdot \langle x_p \rangle^s) = \int_{\widehat{\mathbf{Z}}^{(p)}} \chi(x) \cdot \langle x_p \rangle^s \lambda_{f,\alpha}.$$

On peut voir un caractère de Dirichlet modulo M comme un caractère de $\widehat{\mathbf{Z}}^{(p)}$ constant modulo $pM\widehat{\mathbf{Z}}$, et la proposition suivante montre que la fonction L p -adique est obtenue par interpolation des valeurs aux entiers des fonctions L complexes :

PROPOSITION 4.5. — *Si $v_p(\alpha) < k - 1$, si χ est un caractère de Dirichlet de conducteur M premier à p , si η est un caractère de Dirichlet de conducteur p^m , et si $1 \leq j \leq k - 1$, alors*

$$L_{p,\alpha}(f, \chi \cdot \eta \cdot x_p^j) = \begin{cases} \left(1 - \frac{\chi^{-1}(p)p^j}{\alpha}\right) \cdot \left(1 - \frac{\beta\chi(p)}{p^j}\right) \cdot \widetilde{L}(f, \chi, j) & \text{si } m = 0, \\ \frac{\chi^{-1}(p^m)}{\alpha^m} \cdot p^{mj} \cdot \frac{\widetilde{L}(f, \chi \cdot \eta^{-1}, j)}{G(\eta^{-1})} & \text{si } m \geq 1. \end{cases}$$

4.3. L'application Exp*

4.3.1. (φ, Γ) -modules. On munit $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ et $\mathcal{R}_{\mathbf{Q}_p}$ d'un frobenius φ et d'une action de $\Gamma = \text{Gal}(\mathbf{Q}_p(\zeta_{p^\infty})/\mathbf{Q}_p)$ via les formules

$$\varphi\left(\sum_{k \in \mathbf{Z}} a_k T^k\right) = \sum_{k \in \mathbf{Z}} a_k ((1 + T)^p - 1)^k \quad \text{et} \quad \gamma\left(\sum_{k \in \mathbf{Z}} a_k T^k\right) = \sum_{k \in \mathbf{Z}} a_k ((1 + T)^{\chi_{\text{cycl}}(\gamma)} - 1)^k.$$

Ces actions commutent entre elles. Si L est une extension finie de \mathbf{Q}_p , on étend ces actions à \mathcal{E}_L^\dagger et \mathcal{R}_L par L -linéarité.

Un (φ, Γ) -module étale sur \mathcal{E}_L^\dagger est un \mathcal{E}_L^\dagger -espace vectoriel D de dimension finie d muni d'actions semi-linéaires de φ et Γ commutant entre elles, l'action de φ étant étale (= de pente 0), ce qui signifie qu'il existe une base de D sur \mathcal{E}_L^\dagger telle que les matrices de φ et φ^{-1} dans cette base soient à coefficients dans l'anneau des entiers de \mathcal{E}_L^\dagger (pour la valuation $v_p(\sum a_k T^k) = \inf v_p(a_k)$).

⁸³ $\chi(xy) = \chi(x)\chi(y)$ si $x, y \in \widehat{\mathbf{Z}}^{(p)}$.

THÉORÈME 4.6 ([75, 37]). — On dispose d'un foncteur naturel⁸⁴ $V \mapsto \mathbf{D}^\dagger(V)$ qui induit une équivalence de catégories de la catégorie des L-représentations de dimension finie de $\mathcal{G}_{\mathbf{Q}_p}$ sur celle des (φ, Γ) -modules étales sur \mathcal{E}_L^\dagger .

La théorie des (φ, Γ) -modules ayant déjà fait l'objet d'un demi-exposé à ce séminaire [55], nous ne nous étendrons pas sur la démonstration de ce théorème. On pourra consulter la note 84 pour une description du foncteur \mathbf{D}^\dagger ; celui-ci fait intervenir l'anneau \mathbf{B}^\dagger de Fontaine qui n'est pas franchement facile à appréhender...

REMARQUE 4.7. — On étend l'action de φ et Γ sur \mathcal{R}_L à $\mathcal{R}_L[\log T]$ par les formules

$$\varphi(\log T) = \log \varphi(T) = p \log T + \log \frac{\varphi(T)}{T^p} \quad \text{et} \quad \gamma(\log T) = \log \gamma(T) = \log T + \log \frac{\gamma(T)}{T},$$

ces formules ayant un sens car la série définissant $\log \frac{\varphi(T)}{T^p}$ converge dans \mathcal{E}_L^\dagger et celle définissant $\log \frac{\gamma(T)}{T}$ dans \mathcal{R}_L^+ vers un élément d'ordre 1. On munit aussi $\mathcal{R}_L[\log T]$ de l'unique \mathcal{R}_L -dérivation N telle que $N(\log T) = -\frac{p}{p-1}$.

Grâce aux travaux de Berger [12, 55], on sait décrire $\mathbf{D}_{\text{cris}}(V)$ et $\mathbf{D}_{\text{st}}(V)$ en termes de $\mathbf{D}^\dagger(V)$, tous ces modules vivant à l'intérieur de $\mathcal{R}_L[\log T] \otimes_{\mathcal{E}_L^\dagger} \mathbf{D}^\dagger(V)[\frac{1}{t}]$, et on peut en tirer une « description » (pas très explicite, mais suffisante pour ce que l'on veut en faire dans ce texte (prop. 4.10 et 4.14)) de $\mathbf{D}^\dagger(V)$ en termes de $\mathbf{D}_{\text{cris}}(V)$, si V est cristalline, et de $\mathbf{D}_{\text{st}}(V)$, si V est semi-stable.

4.3.2. (φ, Γ) -modules et théorie d'Iwasawa. Si D est un (φ, Γ) -module étale sur \mathcal{E}_L^\dagger , tout élément x de D peut s'écrire de manière unique sous la forme $\sum_{i=0}^{p-1} (1+T)^i \varphi(x_i)$, ce qui permet de définir un inverse à gauche ψ de φ par la formule $\psi(x) = x_0$; ce morphisme commute à l'action de Γ et joue un rôle très important [87, 36, 53] pour le calcul de la cohomologie galoisienne ou la théorie d'Iwasawa des L-représentations de $\mathcal{G}_{\mathbf{Q}_p}$.

⁸⁴Notons $\tilde{\mathbf{E}}$ le corps des fractions de $\tilde{\mathbf{E}}^+ = \mathbb{R}(\overline{\mathbf{Q}}_p)$ (cf. n° 2.1.1) et $\tilde{\mathbf{A}} = \mathbf{W}(\tilde{\mathbf{E}})$ l'anneau des vecteurs de Witt à coefficients dans $\tilde{\mathbf{E}}$. Il est naturellement muni d'actions de $\mathcal{G}_{\mathbf{Q}_p}$ et φ commutant entre elles et on dispose d'un plongement de l'anneau des entiers de $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ dans $\tilde{\mathbf{A}}$ commutant aux actions de $\mathcal{G}_{\mathbf{Q}_p}$ et φ , à savoir celui envoyant T sur $\pi = \tilde{\zeta} - 1$. L'adhérence $\mathbf{A}_{\mathbf{Q}_p}$ de son image dans $\tilde{\mathbf{A}}$ est l'ensemble des séries de Laurent $\sum_{k \in \mathbf{Z}} a_k \pi^k$ avec $a_k \in \mathbf{Z}_p$ et $\lim_{k \rightarrow -\infty} v_p(a_k) = +\infty$. On note \mathbf{A} l'adhérence pour la topologie p -adique de l'anneau des entiers de l'extension maximale non ramifiée de $\mathbf{B}_{\mathbf{Q}_p} = \mathbf{A}_{\mathbf{Q}_p}[\frac{1}{p}]$ dans $\tilde{\mathbf{B}} = \tilde{\mathbf{A}}[\frac{1}{p}]$, et on pose $\mathbf{B} = \mathbf{A}[\frac{1}{p}]$. On définit le sous-corps $\tilde{\mathbf{B}}^\dagger$ des éléments surconvergents de $\tilde{\mathbf{B}}$ comme l'ensemble des éléments $x = \sum_{k \gg -\infty} p^k [x_k]$, où $(x_k)_{k \gg -\infty}$ est une suite d'éléments de $\tilde{\mathbf{E}}$ vérifiant $\liminf_{k \rightarrow +\infty} \frac{1}{p} v_{\tilde{\mathbf{E}}}(x_k) > -\infty$, et on pose $\mathbf{B}^\dagger = \mathbf{B} \cap \tilde{\mathbf{B}}^\dagger$; c'est un sous-corps de $\tilde{\mathbf{B}}$ stable par φ et $\mathcal{G}_{\mathbf{Q}_p}$ et on a $\mathcal{E}_{\mathbf{Q}_p}^\dagger = (\mathbf{B}^\dagger)^{\mathcal{H}}$, si $\mathcal{H} = \text{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p(\zeta_{p^\infty}))$, et $(\mathbf{B}^\dagger)^{\varphi=1} = \mathbf{Q}_p$. Ceci nous permet de définir un foncteur \mathbf{D}^\dagger qui à une \mathbf{Q}_p -représentation V de $\mathcal{G}_{\mathbf{Q}_p}$ associe le (φ, Γ) -module étale $\mathbf{D}^\dagger(V) = (\mathbf{B}^\dagger \otimes_{\mathbf{Q}_p} V)^{\mathcal{H}}$ et le foncteur \mathbf{V}^\dagger qui à un (φ, Γ) -module D étale sur $\mathcal{E}_{\mathbf{Q}_p}^\dagger$ associe le \mathbf{Q}_p -espace vectoriel $\mathbf{V}^\dagger(D) = (\mathbf{B}^\dagger \otimes_{\mathbf{B}_{\mathbf{Q}_p}^\dagger} D)^{\varphi=1}$. Ces deux foncteurs sont inverses l'un de l'autre et fournissent une description de l'équivalence de catégories du théorème. (On n'a traité dans cette note que le cas des \mathbf{Q}_p -représentations, mais si on est parti d'une L-représentation, tous les objets considérés sont des L-espaces vectoriels.)

THÉORÈME 4.8. — Si V est une L -représentation de $\mathcal{G}_{\mathbf{Q}_p}$, on dispose d'un isomorphisme naturel⁸⁵

$$\mathrm{Exp}^* : H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\mathbf{Z}_p^*, V)) \longrightarrow \mathbf{D}^\dagger(V)^{\psi=1}.$$

L'application Exp^* doit son nom au résultat suivant qui se déduit assez facilement [36] de la formule explicite donnée dans la note 85.

THÉORÈME 4.9. — Si V est une L -représentation de de Rham de $\mathcal{G}_{\mathbf{Q}_p}$, si $n \in \mathbf{N}$ est assez grand, et si $\mu \in H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\mathbf{Z}_p^*, V))$, alors $\varphi^{-n}(\mathrm{Exp}^*(\mu))$ converge⁸⁶ dans $\mathbf{B}_{\mathrm{dR}}^+ \otimes V$, et on a

$$p^{-n}\varphi^{-n}(\mathrm{Exp}^*(\mu)) = \sum_{k \in \mathbf{Z}} \exp^* \left(\int_{1+p^n \mathbf{Z}_p} x^{-k} \mu \right) \in L(\zeta_{p^n})[[t]] \otimes_L \mathbf{D}_{\mathrm{dR}}(V).$$

Pour pouvoir utiliser l'application Exp^* pour construire les fonctions L p -adiques attachées aux formes modulaires, nous allons avoir besoin d'analyser plus en détail la structure de $\mathbf{D}^\dagger(V_f)^{\psi=1}$, si f est une forme primitive de poids $k \geq 2$. On est amené à distinguer les cas où la représentation V_f est cristalline, semi-stable ou seulement potentiellement semi-stable.

4.4. Le cas cristallin

Soit f une forme primitive de poids $k \geq 2$ et niveau N premier à p . La représentation V_f est alors cristalline et, si on factorise le facteur d'Euler en p de la fonction $L(f, s)$ sous la forme $(1 - \alpha p^{-s})(1 - \beta p^{-s})$, alors α et β sont les valeurs propres de φ sur $\mathbf{D}_{\mathrm{cris}}(V_f)$. Pour nous simplifier la vie, nous allons supposer⁸⁷ $\alpha \neq \beta$ et utiliser le fait que ni α ni β ne peuvent être égaux⁸⁸ à 1 ou p^{k-1} . Soit $L = \mathbf{Q}_p(f, \alpha)$, et considérons V_f comme une L -représentation de dimension 2 de $\mathcal{G}_{\mathbf{Q}_p}$. Comme les poids de Hodge-Tate de V_f sont $1 - k$

⁸⁵ Cet isomorphisme est dû à Fontaine (non publié) et repose sur les travaux de Herr [87]. Il est parfaitement explicite et utilise l'application $x \mapsto (1 - \varphi)x$ qui donne naissance à la suite exacte $0 \rightarrow \mathbf{Q}_p \rightarrow \mathbf{B}^\dagger \rightarrow \mathbf{B}^\dagger \rightarrow 0$. L'autre ingrédient de la construction est le fait que, si γ_n est un générateur de $\mathrm{Gal}(\mathbf{Q}_p(\zeta_{p^\infty})/\mathbf{Q}_p(\zeta_{p^n}))$, alors $1 - \gamma_n$ agissant sur $\mathbf{D}^\dagger(V)^{\psi=0}$ est un isomorphisme (c'est le point délicat de la théorie). Muni de ces deux ingrédients, on peut décrire l'inverse de Exp^* comme suit : si $x \in \mathbf{D}^\dagger(V)^{\psi=1}$, alors $(1 - \varphi)x \in \mathbf{D}^\dagger(V)^{\psi=0}$ et il existe $x_n \in \mathbf{D}^\dagger(V)^{\psi=0}$ vérifiant $(1 - \gamma_n)x_n = (1 - \varphi)x$. Si $b_n \in \mathbf{B}^\dagger \otimes_{\mathbf{Q}_p} V$ est une solution de l'équation $(1 - \varphi)b_n = x_n$, alors le cocycle $\sigma \mapsto \log \chi_{\mathrm{cycl}}(\gamma_n) \left(\frac{\sigma-1}{\gamma_n-1} x + (\sigma-1)b_n \right)$ (sur $\mathcal{G}_{\mathbf{Q}_p(\zeta_{p^n})}$) est à valeurs dans V (car tué par $1 - \varphi$) et sa classe dans $H^1(\mathcal{G}_{\mathbf{Q}_p(\zeta_{p^n})}, V)$ est l'image de $(\mathrm{Exp}^*)^{-1}x \in H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\mathbf{Z}_p^*, V))$ par l'application $\mu \mapsto \int_{1+p^n \mathbf{Z}_p} \mu$.

⁸⁶ $\mathbf{B}_{\mathrm{dR}}^+$ et \mathbf{B} sont construits à partir de $W(\tilde{\mathbf{E}}^+)$ en localisant et en complétant, mais pour des topologies trop différentes pour que l'on puisse plonger \mathbf{B}^\dagger dans $\mathbf{B}_{\mathrm{dR}}^+$; par contre, \mathbf{B}^\dagger est exactement le sous-ensemble de \mathbf{B} des éléments $x = \sum_{k \geq k(x)} p^k [x_k]$ tels que la série $\sum_{k \geq k(x)} p^k [x_k^{p^{-n}}] = \varphi^{-n}(x)$ converge dans $\mathbf{B}_{\mathrm{dR}}^+$ si n est assez grand. Par ailleurs, on a $\int_{1+p^n \mathbf{Z}_p} x^{-k} \mu \in H^1(\mathcal{G}_{\mathbf{Q}_p(\zeta_{p^n})}, V(-k))$ et $\exp^*(\int_{1+p^n \mathbf{Z}_p} x^{-k} \mu) \in t^k L(\zeta_{p^n}) \mathbf{D}_{\mathrm{dR}}^{-k}(V)$ est nul si $k \ll 0$.

⁸⁷ On conjecture que c'est toujours le cas : semi-simplicité du frobenius cristallin plus admissibilité de $\mathbf{D}_{\mathrm{cris}}(V_f)$.

⁸⁸ On a $|\alpha| = |\beta| = p^{(k-1)/2}$ d'après Deligne.

et 0, le φ -module filtré $D = \mathbf{D}_{\text{cris}}(V_f)$ est un L -espace vectoriel de dimension 2 muni d'une action L -linéaire de φ et d'une filtration $(D^i)_{i \in \mathbf{Z}}$ vérifiant $D^i = D$ si $i \leq 0$, $D^i = D^{k-1}$ si $0 < i \leq k-1$, et $D^i = 0$ si $i > k-1$ (en particulier, cette filtration est complètement déterminée par la donnée de la L -droite D^{k-1} dont f est naturellement une base); il est isomorphe à l'un des modules $D_{\alpha, \beta}$ ou $D_\alpha \oplus D_\beta$ ci-dessous⁸⁹.

- si $\alpha, \beta \in L$ vérifient $v_p(\alpha) \geq v_p(\beta) \geq 0$ et $v_p(\alpha) + v_p(\beta) = k-1$, on note $D_{\alpha, \beta}$ le φ -module filtré défini par $D_{\alpha, \beta} = L \cdot e_\alpha \oplus L \cdot e_\beta$, avec $\varphi(e_\alpha) = \alpha e_\alpha$, $\varphi(e_\beta) = \beta e_\beta$, et $D_{\alpha, \beta}^{k-1} = L \cdot (e_\alpha + e_\beta)$;

- si $\alpha, \beta \in L$ vérifient $v_p(\alpha) = k-1$ et $v_p(\beta) = 0$, on note $D_\alpha \oplus D_\beta$ le φ -module filtré défini par $D_\alpha \oplus D_\beta = L \cdot e_\alpha \oplus L \cdot e_\alpha$, avec $\varphi(e_\alpha) = \alpha e_\alpha$, $\varphi(e_\beta) = \beta e_\beta$, et $(D_\alpha \oplus D_\beta)^{k-1} = L \cdot e_\alpha$.

Les modules $D_{\alpha, \beta}$ et $D_\alpha \oplus D_\beta$ sont faiblement admissibles et donc [59], les représentations $\mathbf{V}_{\text{cris}}(D_\alpha \oplus D_\beta)$ et $\mathbf{V}_{\text{cris}}(D_{\alpha, \beta})$ sont des L -représentations cristallines de $\mathcal{G}_{\mathbf{Q}_p}$ de dimension 2 dont les poids de Hodge-Tate sont $1-k$ et 0. On pose

$$D_{\alpha, \beta}^\dagger = \mathbf{D}^\dagger(\mathbf{V}_{\text{cris}}(D_{\alpha, \beta})) \quad \text{et} \quad D_\alpha^\dagger \oplus D_\beta^\dagger = \mathbf{D}^\dagger(\mathbf{V}_{\text{cris}}(D_\alpha \oplus D_\beta)).$$

PROPOSITION 4.10. — (i) Si $\alpha, \beta \in L$ vérifient $v_p(\alpha) \geq v_p(\beta) \geq 0$ et $v_p(\alpha) + v_p(\beta) = k-1$, alors $w = w_\alpha \cdot t^{1-k} e_\alpha + w_\beta \cdot t^{1-k} e_\beta \in \mathcal{R}_L \otimes_L D_{\alpha, \beta}[\frac{1}{t}]$ appartient à $(D_{\alpha, \beta}^\dagger)^{\psi=1}$ si et seulement si il existe des distributions μ_α et μ_β sur \mathbf{Z}_p , à valeurs dans L , vérifiant

(a) μ_α est d'ordre $k-1-v_p(\alpha)$ et μ_β est d'ordre $k-1-v_p(\beta)$,

(b)⁹⁰ $\psi(\mu_\alpha) = p^{1-k} \alpha \cdot \mu_\alpha$ et $\psi(\mu_\beta) = p^{1-k} \beta \cdot \mu_\beta$,

(c) $\alpha^{-n} \int_{\mathbf{Z}_p} x^i \zeta_p^n \mu_\alpha = \beta^{-n} \int_{\mathbf{Z}_p} x^i \zeta_p^n \mu_\beta$ quels que soient $n \geq 1$ et $1 \leq i \leq k-1$,

telles que l'on ait $w_\alpha = \int_{\mathbf{Z}_p} (1+T)^x \mu_\alpha$ et $w_\beta = \int_{\mathbf{Z}_p} (1+T)^x \mu_\beta$.

(ii) Si $\alpha, \beta \in L$ vérifient $v_p(\alpha) = k-1$, $\alpha \neq p^{k-1}$ et $v_p(\beta) = 0$, $\beta \neq 1$, alors $x \in (D_\alpha^\dagger \oplus D_\beta^\dagger)^{\psi=1}$ si et seulement si il existe des mesures μ_α et μ_β sur \mathbf{Z}_p , à valeurs dans L , vérifiant $\psi(\mu_\alpha) = p^{1-k} \alpha \cdot \mu_\alpha$ et $\psi(\mu_\beta) = \beta \cdot \mu_\beta$ telles que l'on ait

$$w = \int_{\mathbf{Z}_p} (1+T)^x \mu_\alpha \cdot t^{1-k} e_\alpha + \int_{\mathbf{Z}_p} (1+T)^x \mu_\beta \cdot e_\beta.$$

Démonstration. — Le (ii) se démontre comme le (i), mais est plus simple donc nous n'allons considérer que le (i). La condition (b) est une traduction de l'égalité $\psi(w) = w$; la condition (c) se traduit en terme du développement limité de $\varphi^{-n}(w)$ à l'ordre $k-2$ et permet de montrer que $w \in \mathcal{R}_L \otimes_{\mathcal{O}_L} D^\dagger$, et la condition (a) implique que w est « d'ordre 0 », ce qui permet de montrer que $w \in D^\dagger$. Dans l'autre sens, l'existence de distributions vient de ce qu'une solution dans \mathcal{R}_L d'une équation du type $\psi(x) - \alpha x \in \mathcal{R}_L^+$, où $\alpha \in L$

⁸⁹Le cas $D_\alpha \oplus D_\beta$ correspond au cas où la représentation V_f est scindée; il semble raisonnable de penser que cela ne peut se produire que si f est de type CM; j'ignore si ceci est connu ou pas.

⁹⁰On définit $\psi(\mu)$, si μ est une distribution à valeurs dans L , en passant par les transformées d'Amice, par la formule $\mathcal{A}_{\psi(\mu)} = \psi(\mathcal{A}_\mu)$. On a alors $\int_{\mathbf{Z}_p} f(x) \psi(\mu) = \int_{\mathbf{Z}_p} f(\frac{x}{p}) \mu$ quelle que soit la fonction localement analytique f .

vérifie $v_p(\alpha) \leq 0$, appartient à \mathcal{R}_L^+ (resp. $\mathcal{R}_L^+ \oplus L \cdot \frac{1}{T}$) si $\alpha \neq 1$ (resp. si $\alpha = 1$), et donc est la transformée d'Amice d'une distribution.

Soit π_α (resp. π_β) la projection sur $L \cdot e_\alpha$ (resp. $L \cdot e_\beta$) parallèlement à $L \cdot e_\beta$ (resp. $L \cdot e_\alpha$). Il ressort de ce qui précède que, si $z \in H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\mathbf{Z}_p^*, V_f))$, alors $\pi_\alpha(\text{Exp}^*(z)) = w_\alpha \cdot t^{1-k} e_\alpha$, où w_α est la transformée d'Amice d'un élément de $\mathcal{D}_{k-1-v_p(\alpha)}(\mathbf{Z}_p, L)$. Les mêmes arguments montrent que, si $z \in H^1(\mathcal{G}_{\mathbf{Q}_p}, \mathcal{D}_0(\widehat{\mathbf{Z}}^{(p)}, V_f))$, alors $\pi_\alpha(\text{Exp}^*(z)) = w_\alpha \cdot t^{1-k} e_\alpha$, où w_α est la transformée d'Amice d'un élément de $\mathcal{D}_{k-1-v_p(\alpha)}(\widehat{\mathbf{Z}}, L)$.

THÉORÈME 4.11. — Si $v_p(\alpha) > 0$, alors $\pi_\alpha(f) \neq 0$ et

$$(1 + T) \frac{d}{dT} (\pi_\alpha(\text{Exp}^*(\mathbf{z}_{\text{Kato}}(f)))) = \left(\int_{\mathbf{Z}_p} (1 + T)^x \mu_{f^*, p^{k-1}\alpha^{-1}} \right) \cdot t^{1-k} \pi_\alpha(f).$$

Démonstration. — C'est une conséquence de la caractérisation (th. 3.1) de $\mathbf{z}_{\text{Kato}}(f)$, de la définition (th. 4.4) de la mesure $\mu_{f^*, p^{k-1}\alpha^{-1}}$ et de la loi de réciprocité explicite du th. 4.9.

REMARQUE 4.12. — On peut aussi projeter $\text{Exp}^*(\mathbf{z}_{\text{Kato}}(f))$ sur la droite propre pour la valeur propre β . Si $v_p(\beta) > 0$, le résultat précédent reste valable. Si $v_p(\beta) = 0$ et la représentation V_f de $\mathcal{G}_{\mathbf{Q}_p}$ n'est pas une somme directe de deux sous-représentations, alors $\pi_\beta(f) \neq 0$, et on récupère une distribution d'ordre $k - 1$ qui permet de définir une fonction L p -adique. Les valeurs de cette fonction L p -adique aux caractères critiques (i.e. les $L_{p,\beta}(f, \eta \cdot x_p^j)$, pour η d'ordre fini et $j \in \{1, \dots, k - 1\}$) sont reliées aux valeurs de la fonction L complexe comme pour α , mais ces valeurs ne suffisent pas à déterminer complètement la fonction L p -adique; il manque les valeurs pour un entier j quelconque n'appartenant pas à $\{1, \dots, k - 1\}$. Une voie d'approche pour restaurer une unicité est d'utiliser le fait que la machine de Perrin-Riou fournit des renseignements en tous les caractères algébriques pour une fonction L p -adique fabriquée à partir d'un système d'Euler⁹¹; on obtient en particulier, si $j = 0$, une expression pour $L_{p,\beta}(f, \eta \cdot x_p^j)$ en termes du régulateur [22, 23] des éléments de Beilinson [8] ou de Scholl [161] (cf. [80] et [145]). Une autre approche est fournie par les symboles modulaires « en famille » (cf. [149]). Le cas d'une forme modulaire de type CM est un peu étrange : comme $\pi_\beta(f) = 0$, la fonction L que l'on construit⁹² s'annule en tous les caractères critiques et il n'y a aucune raison pour qu'elle soit identiquement nulle!

4.5. Le cas semi-stable

Supposons maintenant que le niveau de f est divisible par p , mais que $a_p \neq 0$. La représentation V_f est alors semi-stable, et le (φ, N) -module filtré $D = \mathbf{D}_{\text{st}}(V_f)$ est un L-espace vectoriel de dimension 2 muni d'actions L-linéaires de φ et N vérifiant la relation

⁹¹Par construction pour les entiers $\gg 0$, et grâce à la loi de réciprocité explicite [51, 102, 11] pour les autres.

⁹²On a quand même le problème, pour normaliser cette fonction, de choisir une base de l'espace propre pour β , le choix évident $\pi_\beta(f)$ ne marchant pas.

$N\varphi = p\varphi N$ et d'une filtration $(D^i)_{i \in \mathbf{Z}}$ entièrement déterminée par la donnée de la L-droite D^{k-1} ; il est isomorphe à un des modules $D_{\alpha, \mathcal{L}}$ ci-dessous (avec $\alpha = pa_p$ et $\mathcal{L} = \mathcal{L}_f$ invariant de Fontaine).

Si $\alpha \in L$ vérifie $2v_p(\alpha) - 1 = k - 1$ et $\mathcal{L} \in L$, on note $D_{\alpha, \mathcal{L}}$ le (φ, N) -module filtré défini par $D_{\alpha, \mathcal{L}} = L \cdot e_\alpha \oplus L \cdot Ne_\alpha$, avec $\varphi(e_\alpha) = \alpha e_\alpha$ et $D_{\alpha, \mathcal{L}}^{k-1} = L \cdot (e_\alpha + \mathcal{L} \cdot Ne_\alpha)$. (La relation $N\varphi = p\varphi N$ entraîne que l'on a de plus $\varphi(Ne_\alpha) = p^{-1}\alpha \cdot Ne_\alpha$ et $N(Ne_\alpha) = 0$.)

Le (φ, N) -module filtré $D_{\alpha, \mathcal{L}}$ est admissible et $\mathbf{V}_{\text{st}}(D_{\alpha, \mathcal{L}})$ est une L-représentation semi-stable de $\mathcal{G}_{\mathbf{Q}_p}$ de dimension 2 dont les poids de Hodge-Tate sont $1 - k$ et 0 . Posons $D_{\alpha, \mathcal{L}}^\dagger = \mathbf{D}^\dagger(\mathbf{V}_{\text{st}}(D_{\alpha, \mathcal{L}}))$. Pour décrire $(D_{\alpha, \mathcal{L}}^\dagger)^{\psi=1}$ à l'intérieur de $\mathcal{R}_L[\log T] \otimes_L D_{\alpha, \mathcal{L}}[\frac{1}{t}]$, nous aurons besoin du lemme suivant.

LEMME 4.13. — Si $\alpha \in L$, il existe $\ell_\alpha \in \mathcal{R}_L[\log T]$ (non unique) vérifiant les conditions suivantes :

$$(1 - \alpha\varphi) \cdot N(\ell_\alpha) = - \begin{cases} 1 + T & \text{si } \alpha \notin p^{-\mathbf{N}}, \\ 1 + T - p^i \frac{t^i}{i!} & \text{si } \alpha = p^{-i}, \end{cases}$$

$$\text{et } (\psi - p^{-1}\alpha) \cdot \ell_\alpha = \begin{cases} 0 & \text{si } \alpha \notin p^{-\mathbf{N}}, \\ -\frac{(p-1)\log T}{p^2} \cdot \frac{t^i}{i!} & \text{si } \alpha = p^{-i}. \end{cases}$$

Démonstration. — On commence par construire $\ell_\alpha^{(0)} = -N(\ell_\alpha)$. Si $v_p(\alpha) > -1$, on prend $\ell_\alpha^{(0)}$ de la forme $c_\alpha + \sum_{n=0}^{+\infty} \alpha^n \varphi^n(T)$, et on ajuste la constante c_α pour que ça marche. On passe de α à $p^{-1}\alpha$ par la formule $\partial \ell_{\alpha/p}^{(0)} = \ell_\alpha^{(0)}$, avec $\partial = (1 + T) \frac{d}{dT}$, en ajustant la constante d'intégration. On définit alors ℓ_α par la formule

$$\ell_\alpha = \frac{p-1}{p} \left(\ell_\alpha^{(0)} \frac{\log \varphi(T)}{p} - \sum_{n=0}^{+\infty} (p\alpha^{-1})^n \psi^n \left(p^{-1} \ell_\alpha^{(0)} \log \frac{\varphi(T)}{T^p} \right) \right).$$

PROPOSITION 4.14. — $w = t^{1-k} (w_\alpha \cdot e_\alpha + w_{\alpha/p} \cdot Ne_\alpha) \in \mathcal{R}_L[\log T] \otimes_L D_{\alpha, \mathcal{L}}[\frac{1}{t}]$ appartient à $(D_{\alpha, \mathcal{L}}^\dagger)^{\psi=1}$ si et seulement si

- (i) $w_{\alpha/p} (\zeta_{p^n} e^{t/p^n} - 1) \equiv p^{-n} \mathcal{L} w_\alpha (\zeta_{p^n} e^{t/p^n} - 1) \pmod{t^{k-1} L(\zeta_{p^n})[[t]]}$, pour n assez grand;
- (ii) il existe⁹³ des distributions μ_α et $\mu_{\alpha/p}$ sur \mathbf{Z}_p , à valeurs dans L , vérifiant

(a) μ_α est d'ordre $k - 1 - v_p(\alpha)$ et $\mu_{\alpha/p}$ est d'ordre $k - v_p(\alpha)$,

(b) $\psi(\mu_\alpha) = p^{1-k} \alpha \cdot \mu_\alpha$ et $\psi(\mu_{\alpha/p}) = p^{-k} \alpha \cdot \mu_{\alpha/p}$,

telles que l'on ait $w_\alpha = \int_{\mathbf{Z}_p} (1 + T)^x \mu_\alpha$ et

$$w_{\alpha/p} = \int_{\mathbf{Z}_p} (1 + T)^x \mu_{\alpha/p} + \int_{\mathbf{Z}_p^*} \ell_{p^{1-k}\alpha} ((1 + T)^x - 1) \mu_\alpha + \begin{cases} 0 & \text{si } \alpha \notin p^{k-1-\mathbf{N}}, \\ \left(\int_{\mathbf{Z}_p} x^i \mu_\alpha \right) \frac{(p-1)\log T}{p^2} \cdot \frac{t^i}{i!} & \text{si } \alpha = p^{k-1-i}. \end{cases}$$

⁹³La distribution μ_α est déterminée par w , mais $\mu_{\alpha/p}$ dépend du choix de ℓ_α

Soit π_α la projection sur $L \cdot e_\alpha$ parallèlement à $L \cdot e_{\alpha/p}$. Comme dans le cas cristallin, on déduit de la proposition 4.14 le résultat suivant qui montre que la fonction L p -adique est l'image du système d'Euler de Kato.

THÉORÈME 4.15. — On a $\pi_\alpha(f) \neq 0$ et

$$(1 + T) \frac{d}{dT} (\pi_\alpha(\text{Exp}^*(\mathbf{z}_{\text{Kato}}(f)))) = \left(\int_{\mathbf{Z}_p} (1 + T)^x \mu_{f^*, p^{k-1}\alpha^{-1}} \right) \cdot t^{1-k} \pi_\alpha(f).$$

REMARQUE 4.16. — La proposition 4.14 nous fournit une deuxième fonction L p -adique attachée à f (en utilisant $\mu_{\alpha/p}$ qui dépend du choix de $\ell_{p^{1-k}\alpha}$), mais il est un peu difficile, dans l'état actuel de nos connaissances, de comprendre quelle est sa signification (si signification il y a).

4.5.1. *Zéros supplémentaires des fonctions L p -adiques.* Ce qui précède va nous permettre de donner une démonstration d'un théorème de Kato-Kurihara-Tsuji⁹⁴

THÉORÈME 4.17. — Si f est une forme primitive de poids $k = 2i + 2 \geq 2$, de niveau N divisible par p , et si la valeur propre de T_p est $p^i = \beta$, alors $L_{p,\beta}(f, x_p^{i+1}, 0) = 0$ et

$$L'_{p,\beta}(f, x_p^{i+1}, 0) = \mathcal{L}_f \cdot \tilde{L}(f, i + 1).$$

Démonstration. — Utilisant le th. 4.15, on se ramène à un calcul purement local dans $(D_{\alpha, \mathcal{L}}^\dagger)^{\psi=1}$, avec $\alpha = p^{i+1}$, et on reprend les notations de la proposition 4.14. Comme $\psi(\mu_\alpha) = p^{1-k} \alpha \mu_\alpha$, on a

$$\int_{\mathbf{Z}_p^*} x^i \mu_\alpha = \int_{\mathbf{Z}_p} x^i (1 - p^{1-k} \alpha \varphi) \cdot \mu_\alpha = (1 - \alpha p^{i-k+1}) \int_{\mathbf{Z}_p} x^i \mu_\alpha.$$

En particulier, si $\alpha \neq p^{k-1-i}$, l'intégrale $\int_{\mathbf{Z}_p^*} x^i \mu_\alpha$ ne s'annule que si $\int_{\mathbf{Z}_p} x^i \mu_\alpha = 0$. Par contre, si $\alpha = p^{k-1-i}$ ce qui est le cas sous les hypothèses du théorème, la fonction $s \mapsto \int_{\mathbf{Z}_p^*} x^i \langle x \rangle^s \mu_\alpha$ a un zéro trivial en $s = 0$ et sa dérivée est donnée par la proposition suivante qui permet de conclure.

⁹⁴Ce théorème [103] fait intervenir l'invariant \mathcal{L}_f de Fontaine [114]; un autre candidat pour l'invariant \mathcal{L}_f a été proposé par Coleman [45] via sa théorie de l'intégration p -adique. Dans le cas d'une forme modulaire correspondant à une courbe elliptique E définie sur \mathbf{Q} , l'égalité de ces deux invariants et la formule $\mathcal{L}_f = \mathcal{L}_E$ sont des conséquences faciles du théorème d'uniformisation de Tate [184]. La formule du th. 4.17 a été démontrée par Stevens [181] (la démonstration est une généralisation de celle de [84] en poids 2; les deux ingrédients principaux en sont, d'une part, une formule [180] pour l'invariant de Coleman en termes de dérivées de valeur propres de l'opérateur de Hecke T_p et, d'autre part, la construction d'une fonction L p -adique pour une famille analytique de formes modulaires [181, 133]) avec l'invariant de Coleman, ce qui permet, en regroupant les théorèmes de Stevens et Kato-Kurihara-Tsuji, de montrer (de manière très détournée) que les invariants de Fontaine et Coleman coïncident; ce dernier fait a été vérifié directement par Coleman et Iovita [48]. (Cf. [93] pour un résultat du même type, et aussi [58] pour une démonstration basée sur une formule exprimant l'invariant \mathcal{L} de Fontaine en termes de dérivées de valeurs propres de Frobenius, le résultat de Stevens [180] mentionné ci-dessus et un résultat de Kisin [104] permettant de passer d'une formule à l'autre.)

PROPOSITION 4.18. — Si $\alpha = p^{k-1-i}$, alors

$$\int_{\mathbf{Z}_p^*} x^i \log x \mu_\alpha = \mathcal{L} \cdot \int_{\mathbf{Z}_p} x^i \mu_\alpha.$$

Démonstration. — L'idée est d'évaluer $\partial^i((1-p\varphi) \cdot w) = (1-p^{i+1}\varphi) \cdot \partial^i w$ en $T = 0$; on note $ae_\alpha + bNe_\alpha$ le résultat. La fonction w n'est pas définie en $T = 0$, mais on peut utiliser l'équation fonctionnelle $\psi(w) = w$ pour prolonger $(1-p\varphi) \cdot w$ en 0. Cette équation fonctionnelle se traduit, en utilisant la relation $1 + T = e^t$, par les identités

$$\begin{aligned} (1-p\varphi) \cdot w &= w - p\varphi(\psi(w)) = \sum_{j \in (\mathbf{Z}/p\mathbf{Z})^*} \left(w_\alpha(\zeta_p^j e^t - 1) \frac{e_\alpha}{t^{k-1}} + w_{\alpha/p}(\zeta_p^j e^t - 1) \frac{Ne_\alpha}{t^{k-1}} \right) \\ &= \frac{1}{p^{n-1}} \sum_{j \in (\mathbf{Z}/p^n\mathbf{Z})^*} \left(w_\alpha(\zeta_{p^n}^j e^{t/p^{n-1}} - 1) \varphi^{1-n} \left(\frac{e_\alpha}{t^{k-1}} \right) + w_{\alpha/p}(\zeta_{p^n}^j e^{t/p^{n-1}} - 1) \varphi^{1-n} \left(\frac{Ne_\alpha}{t^{k-1}} \right) \right) \end{aligned}$$

Cette dernière expression converge pour n assez grand. Notons $a_{n,i}^{(j)}$ (resp. $b_{n,i}^{(j)}$) le coefficient de t^i dans le développement de $w_\alpha(\zeta_{p^n}^j e^t - 1)$ (resp. de $w_{\alpha/p}(\zeta_{p^n}^j e^t - 1)$). On a alors

$$a = \frac{1}{p^{n-1}} \sum_{j \in (\mathbf{Z}/p^n\mathbf{Z})^*} \left(\frac{p^i \alpha}{p^{k-1}} \right)^{-(n-1)} a_{n,i}^{(j)} \quad \text{et} \quad b = \frac{1}{p^{n-1}} \sum_{j \in (\mathbf{Z}/p^n\mathbf{Z})^*} \left(\frac{p^i \alpha}{p^k} \right)^{-(n-1)} b_{n,i}^{(j)}.$$

D'après le (i) de la prop. 4.14, on a $b_{n,i}^{(1)} = p^{-n} \mathcal{L} a_{n,i}^{(1)}$, ce qui implique en vertu de l'action de $\text{Gal}(\mathbf{Q}_p(\zeta_{p^\infty})/\mathbf{Q}_p)$, la relation $b_{n,i}^{(j)} = p^{-n} \mathcal{L} a_{n,i}^{(j)}$ pour tout $j \in (\mathbf{Z}/p^n\mathbf{Z})^*$, et finalement, la relation

$$b = p^{-1} \mathcal{L} a.$$

Il reste à calculer explicitement a et b pour pouvoir exploiter cette relation. Soit $f_i = \partial^i \ell_{p^{-i}}$ et $g_i = \partial^i(\log T \cdot \frac{t^i}{i!}) - \log T \in \mathcal{R}_L^+$. On peut alors écrire $t^{k-1} \cdot \partial^i w$ sous la forme $w_1 e_\alpha + w_2 Ne_\alpha$, avec $w_1 = \int_{\mathbf{Z}_p} x^i (1+T)^x \mu_\alpha$ et

$$w_2 = \int_{\mathbf{Z}_p} x^i (1+T)^x \mu_{\alpha/p} + \int_{\mathbf{Z}_p^*} x^i f_i ((1+T)^x - 1) \mu_\alpha + \frac{p-1}{p} \cdot \int_{\mathbf{Z}_p} x^i \mu_\alpha \cdot (\log T + g_i).$$

On a $t^{k-1} \cdot (1-p^{i+1}\varphi) \cdot \partial^i w = (1-p\varphi)w_1 e_\alpha + (1-\varphi)w_2 Ne_\alpha$. On obtient donc en évaluant en $T = 0$,

$$a = (1-p) \int_{\mathbf{Z}_p} x^i \mu_\alpha.$$

Pour évaluer $(1-\varphi)w_2$ en $T = 0$, il faut faire un petit peu attention car les fonctions ne sont définies en $T = 0$ que par un procédé de régularisation (si tout était défini en 0, on obtiendrait 0); le premier et le troisième terme donnent une contribution nulle (la contribution du troisième est nulle car $\log p = 0$), et on obtient

$$b = \left((1-\varphi) \left(\int_{\mathbf{Z}_p^*} x^i f_i ((1+T)^x - 1) \mu_\alpha \right) \right)_{T=0} = \left(\int_{\mathbf{Z}_p^*} x^i ((1-\varphi) \cdot f_i) ((1+T)^x - 1) \mu_\alpha \right)_{T=0}.$$

Pour continuer le calcul, nous aurons besoin du lemme suivant.

LEMME 4.19. — (i) f_i se prolonge de manière unique en une fonction localement analytique sur le disque épointé $0 < v_p(z) < +\infty$ vérifiant l'équation fonctionnelle

$$f_i(z) = \frac{p-1}{p} \log z + \sum_{(y+1)^p=z+1} f_i(y).$$

(ii) $(1-\varphi) \cdot f_i + \frac{p-1}{p} \log \varphi(T)$ se prolonge en une fonction continue sur le disque $0 < v_p(z)$.

(iii) Si $x \in \mathbf{Z}_p^*$, alors $h_x = ((1-\varphi) \cdot f_i)((1+T)^x - 1) - (1-\varphi) \cdot f_i(T)$ se prolonge en une fonction continue sur le disque $0 < v_p(z)$, et on a $h_x(0) = -\frac{p-1}{p} \log x$.

Démonstration. — Le (i) suit de ce que f_i est définie sur une couronne non vide $0 < v_p(z) \leq r$, $r > 0$ et de l'équation fonctionnelle $f_i = p\psi(f_i) + \frac{p-1}{p}(\log T + g_i)$. Le (iii) est une conséquence immédiate du (ii) qui suit du (i) et de l'équation fonctionnelle déjà mentionnée que l'on peut réécrire (en appliquant φ aux deux membres) sous la forme

$$\varphi(f_i) - f_i = \frac{p-1}{p}(\log \varphi(T) + \varphi(g_i)) + \sum_{\zeta^p=1, \zeta \neq 1} f_i((1+T)\zeta - 1).$$

Pour terminer le calcul de b , on peut retrancher $\int_{\mathbf{Z}_p^*} x^i ((1-\varphi) \cdot f_i) \mu_\alpha$, qui est nul car $\int_{\mathbf{Z}_p^*} x^i \mu_\alpha = 0$, à l'intégrale ci-dessus, et la réécrire sous la forme

$$b = \left(\int_{\mathbf{Z}_p^*} x^i h_x(T) \mu_\alpha \right)_{T=0} = \int_{\mathbf{Z}_p^*} x^i h_x(0) \mu_\alpha = -\frac{p-1}{p} \cdot \int_{\mathbf{Z}_p^*} x^i \log x \mu_\alpha.$$

4.6. Les cas potentiellement cristallin et potentiellement semi-stable

On peut, en principe, faire la liste, à isomorphisme près, des \mathbf{D}_{pst} des L-représentations potentiellement semi-stables de $\mathcal{G}_{\mathbf{Q}_p}$ de dimension 2, dont les poids de Hodge-Tate sont 0 et $1-k$, mais on ne sait pas décrire les $D^{\psi=1}$ en termes de distributions (sauf si l'extension sur laquelle la représentation devient semi-stable est abélienne), ce qui rend problématique la construction de fonctions L p -adiques dans ce cas. Mais il est clair que l'on veut que la fonction L p -adique soit, comme dans les th. 4.11 et 4.15, l'image de $\mathbf{z}_{\text{Kato}}(f^*)$ dans $\mathbf{D}^\dagger(V_{f^*})^{\psi=1}$; notre problème est donc juste d'interpréter le résultat.

RÉFÉRENCES

- [1] A. AGBOOLA et B. HOWARD, Anticyclotomic Iwasawa theory of CM elliptic curves, preprint 2003.
- [2] Y. AMICE, Interpolation p -adique, Bull. Soc. France **92** (1964) 117-180.
- [3] Y. AMICE, Duals. Proc. of a conf. on p -adic analysis (Nijmegen 1978), 1-15, Nijmegen, Math. Institut Katholische Univ., 1978.
- [4] Y. AMICE et J. VÉLU, Distributions p -adiques associées aux séries de Hecke, Astérisque **24-25** (1975) 119-131.
- [5] K. BARRÉ-SIRIEIX, G. DIAZ, F. GRAMAIN, G. PHILIBERT, Une preuve de la conjecture de Mahler-Manin, Invent. Math. **124** (1996) 1-9.

- [6] D. BARSKY, Fonctions zêta p -adiques d'une classe de rayon des corps totalement réels, Groupe d'études d'analyse ultramétrique, 1977-1978 ; errata 1978-1979.
- [7] A. BEILINSON, Higher regulators and values of L-functions, J. Soviet Math. **30** (1985) 2036–2070.
- [8] A. BEILINSON, Higher regulators of modular curves, Cont. Math. **55** (1986) 1–34.
- [9] J. BELLAÏCHE, Congruences endoscopiques et représentations galoisiennes, Thèse de l'université Paris 11, 2002.
- [10] J. BELLAÏCHE et G. CHENEVIER, Formes non tempérées pour $U(3)$ et conjectures de Bloch-Kato, prépublication **672**, univ. Nice, 2003.
- [11] D. BENOIS, On Iwasawa theory of crystalline representations, Duke Math. J. **104** (2000) 211–267.
- [12] L. BERGER, Représentations p -adiques et équations différentielles, Invent. Math. **148** (2002), 219–284.
- [13] L. BERGER, Représentations de de Rham et normes universelles, preprint 2003.
- [14] M. BERTOLINI et H. DARMON, Heegner points on Mumford-Tate curves, Invent. Math. **126** (1996) 413–456.
- [15] M. BERTOLINI et H. DARMON, A rigid-analytic Gross-Zagier formula and arithmetic applications, Ann. of Math. **146** (1997) 111–147.
- [16] M. BERTOLINI et H. DARMON, Heegner points, p -adic L-functions, and the Cerednik-Drinfeld uniformisation, Invent. Math. **131** (1998) 453–491.
- [17] M. BERTOLINI et H. DARMON, p -adic periods, p -adic L-functions and the p -adic uniformisation of Shimura curves, Duke Math. J. **98** (1999) 305–334.
- [18] M. BERTOLINI et H. DARMON, The p -adic L-functions of modular elliptic curves, *in 2001 and Beyond*, Springer-Verlag 2001.
- [19] M. BERTOLINI, H. DARMON, A. IOVITA et M. SPIESS, Teitelbaum's conjecture in the anticyclotomic setting, Amer. J. Math. **124** (2002) 411–449.
- [20] M. BERTOLINI et H. DARMON, Iwasawa's main conjecture for elliptic curves in the anticyclotomic setting, preprint.
- [21] D. BERTRAND, Relations d'orthogonalité sur les groupes de Mordell-Weil, Séminaire de théorie des nombres, Paris 1984–85, 33–39, Progr. Math. **63**, Birkhäuser 1986.
- [22] A. BESSER, Syntomic regulators and p -adic integration I : rigid syntomic regulators, Isr. J. Math. **120** (2000) 291–334.
- [23] A. BESSER, Syntomic regulators and p -adic integration II : K_2 of curves, Isr. J. Math. **120** (2000) 335–359.
- [24] A. BESSER, The p -adic height pairings of Coleman-Gross et Nekovář, proceedings of CNTA7, Montréal (à paraître).
- [25] B. BIRCH et H. SWINNERTON-DYER, Notes on elliptic curves. I, J. Reine Angew. Math. **212** (1963) 7–25.
- [26] B. BIRCH et H. SWINNERTON-DYER, Notes on elliptic curves. II, J. Reine Angew. Math. **218** (1965) 79–108.
- [27] S. BLOCH et K. KATO, L functions and Tamagawa numbers of motives, dans The Grothendieck Festschrift, vol. 1, Prog. in Math. **86**, 333–400, Birkhäuser 1990.
- [28] J-B. BOST, Algebraic leaves of algebraic foliations over number fields, Publ. Math. I.H.É.S **93** (2001) 161–221.
- [29] C. BREUIL, B. CONRAD, F. DIAMOND et R. TAYLOR, On the modularity of elliptic curves over \mathbf{Q} : wild 3-adic exercises, J. Amer. Math. Soc. **14** (2001), 843–939.

- [30] A. BRUMER, On the units of algebraic number fields, *Mathematika* **14** (1967) 121–124.
- [31] H. CARAYOL, Sur les représentations l -adiques associées aux formes modulaires de Hilbert, *Ann. Sci. cole Norm. Sup. (4)* **19** (1986) 409–468.
- [32] J. CASSELS, Arithmetic on an elliptic curve, 1963 Proc. Internat. Congr. Mathematicians (Stockholm, 1962) 234–246 Inst. Mittag-Leffler, Djursholm.
- [33] J. CASSELS, Diophantine equations with special reference to elliptic curves, *J. London Math. Soc.* **41** (1966) 193–291.
- [34] P. CASSOU-NOGUÈS, Valeurs aux entiers négatifs des fonctions zêta et fonctions zêta p -adiques. *Invent. Math.* **51** (1979) 29–59.
- [35] A. CHAMBERT-LOIR, Théorèmes d’algébricité en géométrie diophantienne (d’après J.-B. Bost, Y. André, D. et G. Chudnovsky), *Sém. Bourbaki 2000/01*, exp. 886, *Astérisque* **282** (2002) 175–209.
- [36] F. CHERBONNIER, P. COLMEZ, Théorie d’Iwasawa des représentations p -adiques d’un corps local, *J. Amer. Math. Soc.* **12**, 241–268, 1999
- [37] F. CHERBONNIER et P. COLMEZ, Représentations p -adiques surconvergentes, *Invent. Math.* **133** (1998), 581–611.
- [38] D.V. CHUDNOVSKY et G.V. CHUDNOVSKY, Padé approximations and Diophantine geometry. *Proc. Natl. Acad. Sci. USA* **82** (1985) 2212–2216.
- [39] J. COATES, The work of Mazur and Wiles on cyclotomic fields, *Sém. Bourbaki 1980/81*, exp. 575, *Lecture Notes in Math.* **901** (1981) 220–242.
- [40] J. COATES, The work of Gross and Zagier on Heegner points and the derivatives of L-series, *Sem. Bourbaki 1984/85*, exp. 635, *Astérisque* **133-134** (1986) 57–72.
- [41] J. COATES et A. WILES, On the conjecture of Birch and Swinnerton-Dyer, *Invent. Math.* **39** (1977) 223–251.
- [42] J. COATES et A. WILES, On p -adic L-functions and elliptic units, *J. Australian Math. Soc.*, **A 26** (1978) 1–25.
- [43] R. COLEMAN, Division values in local fields, *Invent. Math.* **53** (1979) 91–116.
- [44] R. COLEMAN, The dilogarithm and the norm residue symbol, *Bull. S.M.F.* **109** (1981) 373–402.
- [45] R. COLEMAN, A p -adic Shimura isomorphism and p -adic periods of modular forms, *Contemp. Math.* **165** (1994) 21–51.
- [46] R. COLEMAN et B. GROSS, p -adic heights on curves, *Adv. Math.* **17** (1989) 73–81.
- [47] R. COLEMAN et E. DE SHALIT, p -adic regulators on curves and special values of p -adic L-functions, *Invent. Math.* **93** (1988), 239–266.
- [48] R. COLEMAN et A. IOVITA, Hidden structures on semi-stable curves, preprint 2003.
- [49] C. CORNUT, Mazur’s conjecture on higher Heegner points, *Invent. Math.* **148** (2002) 495–523.
- [50] P. COLMEZ, Résidu en $s = 1$ des fonctions zêta p -adiques, *Invent. Math.* **91** (1988) 371–389.
- [51] P. COLMEZ, Théorie d’Iwasawa des représentations de de Rham d’un corps local, *Ann. of Math.* **148** (1998) 485–571.
- [52] P. COLMEZ, Intégration sur les variétés p -adiques, *Astérisque* **248** (1998).
- [53] P. COLMEZ, Représentations p -adiques d’un corps local, *Proceedings of the International Congress of Mathematicians II (Berlin 1998)*, *Doc. Mat. Extra vol. II* (1998), 153–162.
- [54] P. COLMEZ, Fonctions L p -adiques, *Sém. Bourbaki 1998/99*, exp. 851, *Astérisque* **266** (2000) 21–58.
- [55] P. COLMEZ, Les conjectures de monodromie p -adiques, *Sém. Bourbaki 2001/02*, exp. 897, *Astérisque* (à paraître).

- [56] P. COLMEZ, Arithmétique de la fonction zêta, dans *La fonction zêta*, 37-164, journées X-UPS 2002.
- [57] P. COLMEZ, Espaces de Banach de dimension finie, *J. Inst. Math. Jussieu* **1** (2002), 331–439.
- [58] P. COLMEZ, Invariants \mathcal{L} et dérivées de valeurs propres de Frobenius, preprint 2003.
- [59] P. COLMEZ et J.-M. FONTAINE, Construction des représentations p -adiques semi-stables, *Invent. Math.* **140** (2000), 1–43.
- [60] H. DARMON, Integration on $\mathcal{H}_p \times \mathcal{H}$ and arithmetic applications, *Ann. of Math.* **154** (2001) 589–639.
- [61] H. DARMON et A. IOVITA, The anticyclotomic main conjecture for supersingular elliptic curves, preprint 2003.
- [62] D. DELBOURGO, On the p -adic Birch, Swinnerton-Dyer conjecture for non-semistable reduction, *J. Number Theory* **95** (2002) 38–71.
- [63] P. DELIGNE, Formes modulaires et représentations ℓ -adiques, *Sém. Bourbaki 1968/69*, exp. 343, *SLN* **179** (1971) 139-172.
- [64] P. DELIGNE, Valeurs de fonctions L et périodes d'intégrales, dans *Automorphic forms, representations and L-functions*, *Proc. Symp. Pure Math.* **33** (1979) 313–346.
- [65] P. DELIGNE, Preuve des conjectures de Tate et de Shafarevitch (d'après G. Faltings), *Sém. Bourbaki 1983/84*, exp. 616, *Astérisque* **121-122** (1985) 25–41.
- [66] P. DELIGNE et K. RIBET, Values of Abelian L-functions at Negative Integers Over Totally Real Fields, *Invent. Math.* **59** (1980) 227–286.
- [67] C. DENINGER et A. SCHOLL, The Beilinson conjectures, *L-functions and arithmetic (Durham, 1989)*, 173–209, *London Math. Soc. Lecture Note Ser.* **153**, Cambridge Univ. Press 1991.
- [68] M. DEURING, Die Zetafunktion einer algebraischen Kurve vom Geschlechte Eins, I, . . . , IV, *Gött. Nach.*, 1953-1957.
- [69] B. EDIXHOVEN, Rational elliptic curves are modular (after Breuil, Conrad, Diamond and Taylor), *Sém. Bourbaki 1999/2000*, exp. 871 *Astérisque* No. **276** (2002) 161–188.
- [70] M. EICHLER, Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion, *Archiv der Mat.* **5** (1954) 355–366.
- [71] G. FALTINGS, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, *Invent. Math.* **73** (1983), 349–366.
- [72] G. FALTINGS, Almost étale extensions, *Astérisque* **279** (2002), 185–270.
- [73] J.-M. FONTAINE, Sur certains types de représentations p -adiques du groupe de Galois d'un corps local ; construction d'un anneau de Barsotti-Tate, *Ann. Math.* **115** (1982) 529–577.
- [74] J.-M. FONTAINE, Le corps des périodes p -adiques. dans “*Périodes p -adiques*” exposé II, *Astérisque* **223** (1994) 59–102.
- [75] J.-M. FONTAINE, Représentations p -adiques des corps locaux, dans “*The Grothendieck Festschrift*”, vol. 2, *Prog. in Math.* **87**, 249–309, Birkhäuser 1991.
- [76] J.-M. FONTAINE, Valeurs spéciales de fonctions L des motifs, *Sém. Bourbaki 1991/92*, exp. 751, *Astérisque* **206**, 205-249, 1992.
- [77] J.-M. FONTAINE et B. PERRIN-RIOU, Autour des conjectures de Bloch et Kato : cohomologie galoisienne et valeurs de fonctions L, dans *Motives (Seattle)*, *Proceedings of Symposia in Pure Mathematics*, vol. **55**, part 1, 599-706, 1994.
- [78] T. FUKAYA, The theory of Coleman power series for K_2 , *J. Alg. Geom.* **12** (2003) 1–80.
- [79] T. FUKAYA, Coleman power series for K -groups and explicit reciprocity laws, preprint 2003.
- [80] M. GEALY, Special values of p -adic L-functions associated to modular forms, preprint 2003.

- [81] D. GOLDFELD, The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer, *Ann. Scuola Norm. Sup. Pisa* (4) **3** (1976) 624–663.
- [82] R. GREENBERG, Iwasawa theory for p -adic representations, *Adv. Stud. Pure Math.* **17** (1989) 97–137.
- [83] R. GREENBERG, On the Birch and Swinnerton-Dyer conjecture, *Invent. Math.* **72** (1983) 241–265.
- [84] R. GREENBERG et G. STEVENS, p -adic L-functions and p -adic periods of modular forms, *Invent. Math.* **111** (1993) 407–447.
- [85] B. GROSS et D. ZAGIER, Heegner points and derivatives of L-series, *Invent. Math.* **84** (1986) 225–320.
- [86] L. GUO, General Selmer groups and critical values of Hecke L-functions, *Math. Ann.* **297** (1993) 221–233.
- [87] L. HERR, Sur la cohomologie galoisienne des corps p -adiques, *Bull. S.M.F.* **126** (1998), 563–600.
- [88] H. HIDA, Anticyclotomic Main Conjectures, preprint 2003.
- [89] H. HIDA et J. TILOUINE, On the anticyclotomic main conjecture for CM fields, *Invent. Math.* **117** (1994) 89–147.
- [90] H. HIDA et J. TILOUINE, Anti-cyclotomic Katz p -adic L-functions and congruence modules. *Ann. Sci. E.N.S.* **26** (1993) 189–259.
- [91] O. HYODO, On the Hodge-Tate decomposition in the imperfect residue field case, *J. Reine Angew. Math.* **365** (1986) 97–113.
- [92] A. IOVITA et R. POLLACK, Iwasawa theory of Elliptic Curves at Supersingular Primes over \mathbf{Z}_p -extensions of number fields, *MSRI Proceedings of a Conference on “Rankin’s method in arithmetic”* (à paraître).
- [93] A. IOVITA et M. SPIESS, Derivatives of p -adic L-functions, Heegner cycles and monodromy modules attached to modular forms, *Invent. Math.* **154** (2003) 333–384.
- [94] A. IOVITA et A. WERNER, p -adic height pairings on abelian varieties with semistable ordinary reduction, preprint 2002.
- [95] K. IWASAWA, On explicit formulas for the norm residue symbol, *J. Math. Soc. Japan* **20** (1968) 151–164.
- [96] H. JACQUET et J. SHALIKA, A non vanishing theorem for zeta functions of \mathbf{GL}_n , *Invent. Math.* **38** (1976) 1–16.
- [97] K. KATO, The explicit reciprocity law and the cohomology of Fontaine-Messing, *Bull. S.M.F.* **119** (1991) 397–441.
- [98] K. KATO, Lectures on the approach to Iwasawa theory for Hasse-Weil L-functions via \mathbf{B}_{dR} , in *Arithmetic Algebraic Geometry*, *Lect. Notes in Math.* **1553**, Springer 1993.
- [99] K. KATO, Euler systems, Iwasawa theory and Selmer groups, *Kodai Math. J.* **22** (1999) 313–372.
- [100] K. KATO, Generalized explicit reciprocity laws, *Algebraic number theory (Hapcheon/Saga, 1996)*, *Adv. Stud. Contemp. Math. (Pusan)* **1** (1999), 57–126.
- [101] K. KATO, Hodge theory and values of zeta functions of modular forms, *Astérisque* (à paraître).
- [102] K. KATO, M. KURIHARA et T. TSUJI, Local Iwasawa theory of Perrin-Riou and syntomic complexes, preprint 1996.
- [103] K. KATO, M. KURIHARA et T. TSUJI, Cours au centre Émile Borel, premier semestre 1997.
- [104] M. KISIN, Overconvergent modular forms and the Fontaine-Mazur conjecture, *Invent. Math.* **153** (2003) 373–454.
- [105] S. KOBAYASHI, Iwasawa theory for elliptic curves at supersingular primes, *Invent. Math.* **152** (2003) 1–36.
- [106] V. KOLYVAGIN, Euler systems, *The Grothendieck Festschrift*, vol. **2**, *Prog. in Math.* **87**, 436–483, Birkhäuser 1990.

- [107] D. KUBERT et S. LANG, Units in the modular function fields II, *Math. Ann.* **218** (1975) 175–189.
- [108] M. KURIHARA, On the Tate Shafarevich groups over cyclotomic fields of an elliptic curve with supersingular reduction. I, *Invent. Math.* **149** (2002) 195–224.
- [109] S. LANG, Les formes bilinéaires de Néron et Tate, *Sém. Bourbaki 1963/64*, exp. 274, rééd. *Sém Bourbaki 1948-1968*, vol. **8**, Soc. Math. France 1995.
- [110] S. LANG, Sur la conjecture de Birch-Swinnerton-Dyer (d’après J. Coates et A. Wiles), *Sém. Bourbaki 1976/77*, exp. 503, pp. 189–200, *Lecture Notes in Math.* **677** (1978) 189–200.
- [111] Y. MANIN, Periods of cusp forms, and p -adic Hecke series, *Math USSR-Sb.* **92** (1973) 371–393.
- [112] B. MAZUR, Rational points of abelian varieties in towers of number fields, *Invent. Math.* **18** (1972) 183–266.
- [113] B. MAZUR, Modular curves and arithmetic, *Proceedings of the International Congress of Mathematicians (Warsaw, 1983)*, 185–211, PWN, Warsaw, 1984.
- [114] B. MAZUR, On monodromy invariants occurring in global arithmetic, and Fontaine’s theory, *Contemp. Math.* **165** (1994) 1–20.
- [115] B. MAZUR et K. RUBIN, Kolyvagin systems, to appear in *Memoirs of the American Math. Soc.*
- [116] B. MAZUR et K. RUBIN, Elliptic curves and class field theory, *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, 185–195, Higher Ed. Press, Beijing, 2002.
- [117] B. MAZUR et P. SWINNERTON-DYER, Arithmetic of Weil curves, *Invent. Math.* **25** (1974), 1–61.
- [118] B. MAZUR et J. TATE, Canonical height pairings via biextensions, *Prog. in Math.* **35**, 195–238, Birkhäuser, 1983.
- [119] B. MAZUR et J. TATE, The p -adic sigma function, *Duke Math. J.* **62** (1991) 663–688.
- [120] B. MAZUR, J. TATE et J. TEITELBAUM, On p -adic analogues of the conjectures of Birch and Swinnerton-Dyer, *Invent. Math.* **84** (1986) 1–48.
- [121] B. MAZUR et J. TILOUINE, Représentations galoisiennes, différentielles de Kähler et « conjectures principales », *Publ. Math I.H.É.S.* **71** (1990) 65–103.
- [122] B. MAZUR et A. WILES, Class fields of abelian extensions of \mathbf{Q} , *Invent. Math.* **76** (1984) 179–330.
- [123] J-F. MESTRE, Formules explicites et minorations de conducteurs de variétés algébriques, *Comp. Math.* **58** (1986) 209–232.
- [124] J. NEKOVÁŘ, Kolyvagin’s method for Chow groups of Kuga-Sato varieties, *Invent. Math.* **107** (1992) 99–125.
- [125] J. NEKOVÁŘ, On p -adic height pairings, dans *Séminaire de théorie des nombres 1990-1991*, *Prog. in Math.* **108**, 127–202, Birkhäuser 1993.
- [126] J. NEKOVÁŘ, On the p -adic heights of Heegner cycles, *Math. Ann.* **302** (1995) 609–686.
- [127] J. NEKOVÁŘ, On the parity of ranks of Selmer groups II, *C.R.A.S.* **332** (2001) 99–104.
- [128] A. NÉRON, Quasi-fonctions et hauteurs sur les variétés abéliennes, *Ann. of Math.* **82** (1965) 249–331.
- [129] A. NÉRON, Fonctions thêta p -adiques et hauteurs p -adiques, *Séminaire de Théorie des Nombres, Paris 1980-1981*, *Prog. in Math.* **22**, Birkhäuser, 1982.
- [130] J. OESTERLÉ, Nombres de classes des corps quadratiques imaginaires, *Sém. Bourbaki 1983/84*, exp. 631, *Astérisque* **121-122** (1985), 309–323.
- [131] J. OESTERLÉ, Travaux de Wiles (et Taylor, ...). II, *Sém. Bourbaki 1994/95*, exp. 804, *Astérisque* **237** (1996) 333–355.
- [132] A. PANCHISHKIN, A new method of constructing p -adic L-functions associated with modular forms, *Mosc. Math. J.* **2** (2002) 313–328.

- [133] A. PANCHISHKIN, Two variable p -adic L functions attached to eigenfamilies of positive slope, preprint 2001.
- [134] B. PERRIN-RIOU, Hauteurs p -adiques, Séminaire de Théorie des Nombres, Paris 1982-1983, Prog. in Math. **51**, Birkhäuser, 1984.
- [135] B. PERRIN-RIOU, Points de Heegner et dérivées de fonctions L p -adiques, Invent. Math. **89** (1987) 455–510.
- [136] B. PERRIN-RIOU, Théorie d’Iwasawa p -adique locale et globale, Invent. Math. **99** (1990) 247-292.
- [137] B. PERRIN-RIOU, Travaux de Kolyvagin et Rubin, *Sém. Bourbaki 1989/90*, exp. 717, Astérisque **189-190** (1990) 69–106.
- [138] B. PERRIN-RIOU, Théorie d’Iwasawa et hauteurs p -adiques, Invent. Math. **109** (1992) 137-185.
- [139] B. PERRIN-RIOU, Fonctions L p -adiques d’une courbe elliptique et points rationnels, Ann. Inst. Fourier **43** (1993) 945–995.
- [140] B. PERRIN-RIOU, Théorie d’Iwasawa des représentations p -adiques sur un corps local, Invent. Math. **115** (1994) 81-149.
- [141] B. PERRIN-RIOU, *Fonctions L p -adiques des représentations p -adiques*, Astérisque **229** (1995).
- [142] B. PERRIN-RIOU, Systèmes d’Euler et représentations p -adiques, Ann. Inst. Fourier **48** (1998) 1231-1307.
- [143] B. PERRIN-RIOU, Représentations p -adiques et normes universelles I, le cas cristallin, J. Amer. Math. Soc. **13** (2000) 533–551.
- [144] B. PERRIN-RIOU, Théorie d’Iwasawa des représentations p -adiques semi-stables, Mém. Soc. Math. Fr. **84** (2001).
- [145] B. PERRIN-RIOU, Quelques remarques sur la théorie d’Iwasawa des courbes elliptiques, *Number theory for the millennium, III* (Urbana, IL, 2000) 119–147.
- [146] B. PERRIN-RIOU, Arithmétique des courbes elliptiques à réduction supersingulière en p , preprint 2001.
- [147] R. POLLACK, On the p -adic L-function of a modular form at a supersingular prime, Duke Math. J. **118** (2003) 523–558.
- [148] R. POLLACK et K. RUBIN, The main conjecture for CM elliptic curves at supersingular primes, Ann. of Maths, (à paraître).
- [149] R. POLLACK et G. STEVENS, The “missing” p -adic L-function, preprint 2003.
- [150] K. RIBET, Galois representations attached to modular forms, Invent. Math. **28** (1975) 245–275.
- [151] K. RIBET, A modular construction of unramified extensions of $\mathbf{Q}(\zeta_p)$, Invent. Math. **34** (1976) 151–162.
- [152] K. RIBET, Galois representations attached to modular forms II, Glasgow Math. J. **27** (1985) 185–194.
- [153] D. ROHRlich, On L-functions of elliptic curves and cyclotomic towers, Invent. Math. **75** (1984) 409–423.
- [154] K. RUBIN, Elliptic curves and \mathbf{Z}_p -extensions, Comp. Math. **56** (1985) 237-250.
- [155] K. RUBIN, Local units, elliptic units, Heegner points, and elliptic curves, Invent. Math. **88** (1987) 405–422.
- [156] K. RUBIN, Tate-Shafarevich groups and L-functions of elliptic curves with complex multiplication, Invent. Math. **89** (1987) 527–560.
- [157] K. RUBIN, The “main conjectures” of Iwasawa theory for imaginary quadratic fields, Invent. Math. **103** (1991) 25–68.
- [158] K. RUBIN, *Euler systems*, Ann. of Math. Studies **147**, Princeton Univ. Press (2000).
- [159] A. SCHOLL, Motives for modular forms, Invent. Math. **100** (1990) 419–430.

- [160] A. SCHOLL, Remarks on special values of L-functions, dans *L-functions and Arithmetic*, Proc. of the Durham Symp., London Math. Soc. L.N.S. **153**, Cambridge University Press, 1991, 373–392.
- [161] A. SCHOLL, An introduction to Kato’s Euler systems, dans *Galois representations in arithmetic algebraic geometry*, 379-460, Cambridge University Press, 1998
- [162] A. SCHOLL, Higher regulators and special values of L-functions of modular forms, en préparation.
- [163] A. SCHOLL, Zeta elements for higher weight modular forms, en préparation.
- [164] P. SCHNEIDER, p -adic height pairings, II, *Invent. Math.* **79**, 329-374, 1985
- [165] P. SCHNEIDER et J. TEITELBAUM, p -adic Fourier theory, *Doc. Math.* **6** (2001) 447–481.
- [166] J-P. SERRE, *Abelian ℓ -adic representations and elliptic curves*, W. A. Benjamin 1968.
- [167] J-P. SERRE, Formes modulaires et fonctions zêta p -adiques, *Modular functions of one variable III*, Lect. Notes Math. **350** (1972) 191–268.
- [168] J-P. SERRE, Travaux de Wiles (et Taylor, ...). I, *Sém. Bourbaki 1994/95*, exp. 803, Astérisque **237** (1996) 319–332.
- [169] J-P. SERRE, lettre du 13/11/59, in *Correspondance Grothendieck-Serre*, Documents Mathématiques **2**, Soc. Math. de France 2001.
- [170] G. SHIMURA, Correspondances modulaires et les fonctions ζ de courbes algébriques, *J. Math. Soc. Japan* **10** (1958) 1–28.
- [171] G. SHIMURA, Sur les intégrales attachées aux formes automorphes, *J. Math. Soc. Japan* **11** (1959) 291–311.
- [172] G. SHIMURA, *Introduction to the arithmetic theory of automorphic functions*, Kan Memorial Lectures 1, Pub. of the Math. Soc. of Japan **11**, 1971.
- [173] G. SHIMURA, On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields, *Nagoya Math. J.* **43** (1971) 199–208.
- [174] G. SHIMURA, On the factors of the jacobian variety of a modular function field. *J. Math. Soc. Japan* **25** (1973) 523–544.
- [175] G. SHIMURA, The special values of the zeta functions associated with cusp forms, *Comm. Pure Appl. Math.* **29** (1976) 783–804.
- [176] C. SKINNER et E. URBAN, Sur les déformations p -adiques des formes de Saito-Kurokawa, *C.R.A.S.* **335** (2002) 581–586.
- [177] C. SKINNER et E. URBAN, En préparation.
- [178] C. SOULÉ, Régulateurs, *Sém. Bourbaki 1984/85*, exp. 644, Astérisque **133-134** (1986) 237–253.
- [179] C. SOULÉ, Eléments cyclotomiques en K-théorie, Astérisque **147-148** (1987) 225-257.
- [180] G. STEVENS, Coleman’s \mathcal{L} -invariant and families of modular forms, preprint 1996.
- [181] G. STEVENS, Cours au centre Émile Borel, premier semestre 1997.
- [182] J. TATE, On the conjectures of Birch and Swinnerton-Dyer and a geometric analog, *Sém. Bourbaki 1965/66*, exp. 306, rééd. *Sém. Bourbaki 1948-1968*, vol. **9**, Soc. Math. France 1995.
- [183] J. TATE, p -divisible groups, Proc. of a conference on local fields, Nuffic Summer School at Driebergen, Springer, Berlin, 158-183, 1967.
- [184] J. TATE, A review of non-Archimedean elliptic functions, *Elliptic curves, modular forms, & Fermat’s last theorem (Hong Kong, 1993)*, 162–184, Ser. Number Theory, I, Internat. Press, Cambridge, MA, 1995.
- [185] F. THAINE, On the ideal class group of real abelian number fields, *Ann. of Math.* **128** (1988) 1–18.
- [186] J. TILOUINE, Sur la conjecture principale anticyclotomique, *Duke Math. J.* **59** (1989) 629-673.

- [187] T. TSUJI, p -adic étale cohomology and crystalline cohomology in the semi-stable reduction case. *Invent. Math.* **137** (1999), 233–411.
- [188] T. TSUJI, Semi-stable conjecture of Fontaine-Jannsen : a survey, *Astérisque* **279** (2002), 323–370.
- [189] V. VATSAL, Uniform distribution of Heegner points, *Invent. Math.* **148** (2002) 1–46.
- [190] M. VISHIK, Non-archimedean measures connected with Dirichlet series, *Math. USSR Sbornik* **28** (1976) 216–228.
- [191] M. WALDSCHMIDT, Sur la nature arithmétique des valeurs de fonctions modulaires, *Sém. Bourbaki 1996/97*, exp. 824, *Astérisque* **245** (1997) 105–140.
- [192] A. WEIL, *Elliptic functions according to Eisenstein and Kronecker*, *Erg. der Math.* **88** (1976), Springer-Verlag.
- [193] A. WILES, Higher explicit reciprocity laws, *Ann. of Math.* **107** (1978) 235–254.
- [194] A. WILES, Modular elliptic curves and Fermat’s last theorem, *Ann. of Math.* **141** (1995) 443–551.
- [195] Y. ZARHIN, Local heights and abelian varieties, dans “Groupe d’étude sur les problèmes diophantiens 1988-1989”, *Publications math. de l’Univ. Pierre et Marie Curie N° 90*.

Pierre COLMEZ

Institut de mathématiques de Jussieu

4 place Jussieu

F-75005 PARIS

E-mail : colmez@math.jussieu.fr

NOMBRES DE BETTI L^2 ET FACTEURS DE TYPE II_1
[d'après D. Gaboriau et S. Popa]

par **Alain CONNES**

INTRODUCTION

M. F. Atiyah a introduit dans ([3]) les nombres de Betti L^2 pour les groupes discrets. Ces invariants ont ensuite été généralisés dans ([14]) aux feuilletages mesurés. L'intérêt initial de cette généralisation tient à des applications immédiates telles que l'existence de feuilles compactes stables pour tout feuilletage mesuré dont les feuilles sont de dimension 2 et ont une courbure dont l'intégrale (pour la mesure transverse) est positive.

Dans un article récent ([28]), D. Gaboriau a montré que les nombres de Betti L^2 des feuilletages à feuilles contractiles sont en fait des invariants de la relation d'équivalence mesurée associée. Il a de plus défini les nombres de Betti ℓ^2 , $\{\beta_n(\mathcal{R})\}$, $n \geq 0$ pour toute relation d'équivalence mesurée \mathcal{R} à orbites dénombrables.

Sa construction généralise aussi les nombres de Betti ℓ^2 d'Atiyah et Cheeger-Gromov ([3], [7]) pour les groupes discrets dénombrables Γ , $\{\beta_n(\Gamma)\}_{n \geq 0}$, et Gaboriau démontre que $\beta_n(\Gamma) = \beta_n(R_\Gamma)$, pour toute relation d'équivalence R_Γ engendrée par une action ergodique libre et préservant la mesure du groupe Γ sur un espace de probabilité (X, μ) ([28]).

La notion de coût des relations d'équivalence mesurées a été introduite par Levitt ([36]), le coût $C(\mathcal{R})$ de la relation \mathcal{R} est la borne inférieure des mesures des sous-relations engendrant \mathcal{R} . Grâce à la notion d'arborage, due à Gaboriau (cf.[27]), le coût permet en particulier de retrouver le nombre de générateurs d'un groupe libre Γ comme invariant des relations d'équivalence mesurées provenant d'une action libre de Γ . L'idée des nombres de Betti des relations d'équivalence a été suggérée à Gaboriau par une remarque d'A. Valette, à savoir la validité, dans les exemples d'actions de groupes où l'on sait calculer les deux membres, de l'égalité

$$C(R_\Gamma) - 1 = \beta_1(\Gamma) - \beta_0(\Gamma)$$

L'on ne sait pas si cette égalité reste valable pour toute action ergodique libre d'un groupe discret Γ .

Le développement de la théorie des relations d'équivalence mesurées est depuis le début parallèle à celui de la théorie des facteurs, et plus précisément des facteurs de type II_1 pour les relations ergodiques préservant la mesure sur un espace de probabilité (X, μ) .

À une telle relation \mathcal{R} correspond canoniquement un facteur $L(\mathcal{R})$ de type II_1 par la construction de Murray et von Neumann (on peut également tordre par un 2-cocycle v (voir [26])). Cette construction n'est ni injective – le même facteur peut provenir de relations ergodiques non-isomorphes [19] – ni surjective même si l'on autorise un 2-cocycle ([57]). Il serait donc simpliste de penser que la «transposition» d'une théorie à l'autre est facile et il a souvent fallu attendre plusieurs années avant que la transplantation d'un résultat d'un côté à l'autre devienne possible. En fait il faut également rajouter à ce dictionnaire une troisième colonne contenant la théorie des groupes discrets. Pour tout groupe discret dénombrable Γ le commutant $L(\Gamma)$ de la représentation régulière droite de Γ est une algèbre de von Neumann finie, c'est un facteur de type II_1 si toutes les classes de conjugaison (hormis celle de 1) sont infinies (on dit alors que Γ est i.c.c.). Ici le foncteur $\Gamma \rightarrow L(\Gamma)$ est très loin d'être injectif – tous les groupes moyennables (i.c.c.) donnent le même facteur R ([12]) – et n'est pas surjectif car il existe des facteurs de type II_1 non-antiisomorphes à eux-mêmes. Il y a essentiellement trois types de phénomènes bien répertoriés dans les trois colonnes, ce sont

1) Moyennabilité

Les résultats clefs dans ce domaine sont l'unicité du facteur hyperfini (Murray et von Neumann), et celle du facteur injectif (moyennable) ([12]). Ce dernier résultat implique en particulier que le facteur hyperfini R est isomorphe à tous ses sous-facteurs. La contrepartie du théorème de von Neumann pour les relations d'équivalence est le théorème de Dye ([25]). La contrepartie de ([12]) est l'unicité de la relation d'équivalence associée à une action de groupe moyennable ([42], [20]).

2) Liberté

Les résultats clefs dans ce domaine dans le cadre des facteurs sont la propriété d'approximation compacte H de Haagerup ([31]), et la théorie des probabilités libres de Voiculescu ([58]). Dans le cadre des groupes libres l'action sur les arbres ([56]) et dans le cas des relations d'équivalence l'«arborabilité» et le coût, mentionnés plus haut, jouent un rôle important.

3) Rigidité

Les résultats clefs sont ici la propriété T de Kazhdan ([35]) et le théorème de rigidité de Margulis ([38]). Ce théorème a été adapté par R. Zimmer aux relations d'équivalence mesurées ([59]), ce qui a permis des progrès décisifs dans ce domaine. La propriété T d'un groupe discret Γ ne dépend en fait que du facteur $L(\Gamma)$ associé et les facteurs correspondants ont des propriétés de rigidité remarquables ([18]).

S. Popa a réussi récemment à établir un tunnel entre les colonnes « facteurs » et « relations » en combinant les propriétés H de Haagerup et T de Kazhdan pour toute une classe de relations d'équivalence. Les facteurs associés à de telles relations (dites de classe HT) sont isomorphes si et seulement si les relations sont isomorphes. Cela lui a permis de résoudre un problème fondamental de la théorie des facteurs de type II_1 , et de construire des facteurs M de type II_1 dont le groupe fondamental est trivial et qui sont en particulier non isomorphes à $M_k(\mathbb{C}) \otimes M$ pour tout entier $k \neq 1$.

Soit N un facteur de type II_∞ ; tout automorphisme $\theta \in \text{Aut}(N)$ multiplie la trace normale semi-finie τ sur N par un scalaire

$$\text{Mod}(\theta) := (\tau \circ \theta) / \tau \in \mathbb{R}_+^*.$$

Le groupe fondamental $F(N) := \{\text{Mod}(\theta) \mid \theta \in \text{Aut}(N)\} \subset \mathbb{R}_+^*$ a été défini par Murray et von Neumann dans les années 40. Pour un facteur M de type II_1 on définit $F(M) := F(M \otimes I_\infty)$ où I_∞ est le facteur de type I_∞ . Murray et von Neumann ont montré que $F(M) = \mathbb{R}_+^*$ quand M est isomorphe au facteur hyperfini de type II_1 , R , et plus généralement quand M est de la forme $P \otimes R$.

Les premiers exemples de facteurs M de type II_1 tels que $F(M) \neq \mathbb{R}_+^*$, ont été donnés dans ([11]). Soit Γ un groupe discret (dont les classes de conjugaison sont infinies) vérifiant la propriété T de Kazhdan, et $M := L(\Gamma)$ le facteur de type II_1 engendré par la représentation régulière de Γ . On montre alors avec $N = L(\Gamma) \otimes I_\infty$ que le groupe $\text{Out}(N)$ des classes d'automorphismes de N modulo les automorphismes intérieurs est dénombrable de sorte que $F(L(\Gamma))$ est dénombrable. De plus pour tout sous-groupe dénombrable $D \subset \mathbb{R}_+^*$, l'existence de facteurs M de type II_1 avec $F(M)$ dénombrable contenant D est démontrée dans ([30], [48]). Mais la détermination exacte de $F(M)$ restait ouverte dans tous ces exemples.

1. NOMBRES DE BETTI ℓ^2

D. Gaboriau a défini les nombres de Betti ℓ^2 , $\{\beta_n(\mathcal{R})\}$, $n \geq 0$ pour toute relation d'équivalence mesurée \mathcal{R} à orbites dénombrables préservant la mesure. L'essentiel de sa démarche se comprend simplement si l'on tient compte de la théorie des mesures transverses sur les groupoïdes mesurables développée dans ([14]) pour formuler le théorème de l'indice longitudinal pour les feuilletages mesurés. Il suffit alors d'étendre (en passant du cas des groupes à celui des groupoïdes) les idées que Cheeger et Gromov ont développées dans ([7]) pour définir les nombres de Betti dans le cas « de covolume fini ».

La généralisation au cas des groupoïdes des notions d'action libre (resp. propre) d'un groupe discret sur un complexe simplicial et de la dimension de Murray-von Neumann d'espaces de chaînes L^2 ne pose aucun problème ([14]). Soit \mathcal{R} une relation d'équivalence ergodique à orbites dénombrables, préservant la mesure sur un espace de probabilité

(X, μ) . Munissons $\mathcal{R} \subset X^2$ de sa structure naturelle de groupoïde, les deux projections étant notées r et s , $(r(x, y) := x$ et $s(x, y) := y)$ et la loi de composition $(x, y) \circ (y, z) := (x, z)$. Un \mathcal{R} -espace discret est un foncteur mesurable Y de la petite catégorie \mathcal{R} vers celle des ensembles dénombrables. La réunion $\bar{Y} := \cup_{x \in X} Y(x)$ est un espace borélien standard, la projection $\pi : \bar{Y} \rightarrow X$, $\pi(z) = x, \forall z \in Y(x)$ et l'action $Y(x, y) : Y(y) \rightarrow Y(x)$ sont boréliennes. On ne s'intéresse qu'aux \mathcal{R} -espace discrets *réguliers* obtenus par réduction $\tilde{\mathcal{R}}_B$ de la somme directe $\tilde{\mathcal{R}}$ d'une infinité dénombrable de copies du foncteur $\mathcal{R}(x) := r^{-1}(x)$ par un borélien invariant B . L'invariance de la mesure μ montre alors que la mesure $\tilde{\mu}(F)$ ne dépend ni du choix d'un domaine fondamental F tel que $\mathcal{R}F = B$ ni de l'identification de Y avec $\tilde{\mathcal{R}}_B$. C'est la *mesure transverse* $\Lambda(Y)$ ([14]). En composant Y avec le foncteur ℓ^2 qui à un ensemble dénombrable Z associe l'espace de Hilbert $\ell^2(Z)$, on obtient un foncteur mesurable (représentation) de \mathcal{R} vers la catégorie des espaces de Hilbert séparables. Son commutant est une algèbre de von Neumann semi-finie $End_\Lambda(\ell^2(Y))$ qui possède une unique trace normale semi-finie Tr_Λ telle que

$$Tr_\Lambda(1_Z) = \Lambda(Z)$$

pour tout borélien \mathcal{R} -invariant Z . De plus $End_\Lambda(\ell^2(Y))$ agit dans l'espace de Hilbert \mathcal{H} intégrale directe des $\ell^2(Y(x))$ sur (X, μ) . C'est l'algèbre des endomorphismes de \mathcal{H} pour la structure évidente de $L(\mathcal{R})$ -module hilbertien et la dimension de Murray-von Neumann de ce module est égale à $Tr_\Lambda(1)$ (Voir ([14])).

Un \mathcal{R} -complexe simplicial est un foncteur mesurable Σ vers la catégorie des ensembles simpliciaux dénombrables et l'on ne s'intéresse qu'au cas où le \mathcal{R} -espace discret $\Sigma^{(0)}$ des sommets est régulier (il en va alors de même pour les $\Sigma^{(n)}$). Dans le cas des feuilletages des variétés compactes, la compacité ambiante suffisait à régler les problèmes d'uniformité, mais dans le cas général traité par Gaboriau les nombres de Betti ne sont plus nécessairement finis et il faut ajouter une étape intermédiaire (analogue à ([7])) avant de définir ces nombres.

DÉFINITION 1.1 ([28]). — *Un \mathcal{R} -complexe simplicial Σ est uniformément localement borné ssi il existe $N < \infty$ tel que tout $x \in \Sigma^{(0)}$ soit le sommet d'au plus N simplexes et si $\Lambda(\Sigma^{(0)}) < \infty$.*

En composant Σ avec le foncteur qui associe à un complexe simplicial localement borné le complexe de ses chaînes ℓ^2 ,

$$0 \xleftarrow{\partial_0} C_0^{(2)} \xleftarrow{\partial_1} C_1^{(2)} \xleftarrow{\partial_2} \dots \xleftarrow{\partial_n} C_n^{(2)} \xleftarrow{\partial_{n+1}} C_{n+1}^{(2)} \xleftarrow{\partial_{n+2}} \dots$$

on obtient un complexe de $L(\mathcal{R})$ modules.

DÉFINITION 1.2 ([28]). — (i) *L'homologie L^2 d'un \mathcal{R} -complexe simplicial uniformément borné Σ est le $L(\mathcal{R})$ -module*

$$H_n^{(2)}(\Sigma, \mathcal{R}, \mu) := Ker \partial_n / \overline{Im \partial_{n+1}}$$

(ii) *L'homologie L^2 d'un \mathcal{R} -complexe simplicial Σ est la limite inductive des homologies L^2 de ses sous-complexes uniformément bornés.*

Ces deux définitions sont compatibles entre elles. La deuxième ne donne pas nécessairement un $L(\mathcal{R})$ -module hilbertien mais W . Luck ([37]) a montré comment prolonger la dimension de Murray-von Neumann : Dim_M au cas des modules généraux sur un facteur fini M . Cela permet de définir les nombres de Betti d'un \mathcal{R} -complexe simplicial Σ comme la dimension généralisée,

$$\beta_n(\Sigma, \mathcal{R}) := \text{Dim}_{L(\mathcal{R})}(H_n^{(2)}(\Sigma, \mathcal{R}, \mu)).$$

Un \mathcal{R} -complexe simplicial Σ est dit n -connexe ssi $\Sigma(x)$ est n -connexe pour tout $x \in X$. Le résultat essentiel de D. Gaboriau s'énonce alors ainsi,

THÉORÈME 1.3 ([28]). — *Tous les \mathcal{R} -complexes simpliciaux n -connexes ont le même nombre de Betti $\beta_n(\Sigma, \mathcal{R})$.*

On définit alors $\beta_n(\mathcal{R})$ comme la valeur commune des $\beta_n(\Sigma, \mathcal{R})$ pour Σ un \mathcal{R} -complexe simplicial n -connexe. L'existence d'un tel complexe s'obtient en composant le foncteur $\tilde{\mathcal{R}}$ ci-dessus avec celui qui associe à tout ensemble dénombrable Z le complexe simplicial universel EZ tel que $EZ^{(0)} = Z$.

COROLLAIRE 1.4 ([28]). — (i) *Soit Γ un groupe discret agissant librement sur un espace de probabilité (X, μ) alors $\beta_n(R_\Gamma) = \beta_n(\Gamma)$.*

(ii) *Deux groupes discrets orbitalement équivalents ont les mêmes nombres de Betti ℓ^2 .*

(On dit que deux groupes discrets Γ_j sont orbitalement équivalents s'ils possèdent une action ergodique libre sur un espace de probabilité engendrant la même relation d'équivalence.)

2. FACTEURS DE TYPE II_1

S. Popa introduit la classe \mathcal{HT} des facteurs M de type II_1 qui possèdent une sous-algèbre abélienne maximale A telle que l'inclusion $A \subset M$ satisfasse deux propriétés, l'une (\mathcal{T}) de rigidité, l'autre (\mathcal{H}) d'approximation compacte. Il montre alors que si M possède une telle sous-algèbre abélienne maximale, celle-ci est unique à conjugaison près par un automorphisme intérieur et détermine de manière unique une relation d'équivalence ergodique \mathcal{R} à orbites dénombrables, préservant la mesure sur un espace de probabilité (X, μ) , et un 2-cocycle v à valeurs dans le groupe unitaire de $L^\infty(X, \mu)$ tels que $M = L(\mathcal{R}, v)$. Il en résulte que tout invariant de \mathcal{R} est un invariant de M .

Sous-algèbres de Cartan.

Une sous-algèbre (involutive) abélienne maximale d'un facteur M est dite régulière ([23]) (ou «de Cartan» ([26])) quand son normalisateur $\mathcal{N}(A)$ dans M engendre M comme algèbre de von Neumann. Soit M un facteur de type II_1 séparable¹. Il est immédiat (cf.([26])) que $A \subset M$ est une sous-algèbre de Cartan si et seulement si il existe une relation d'équivalence ergodique à orbites dénombrables, préservant la mesure sur un espace de probabilité (X, μ) , et un 2-cocycle v à valeurs dans le groupe unitaire de $L^\infty(X, \mu)$ tels que $(A \subset M) \sim (L^\infty(X, \mu) \subset L(\mathcal{R}, v))$. Il s'agit là d'une simple traduction et l'association $(A \subset M) \rightarrow (\mathcal{R}, v) \rightarrow (A \subset M)$ transforme l'isomorphisme des inclusions $(A \subset M)$ en isomorphisme des relations d'équivalence \mathcal{R} avec 2-cocycles v (voir [26]).

L'outil essentiel qui permet de détecter les sous-algèbres de Cartan est la théorie des «correspondances» où «bimodules de Hilbert» ([16], [17], [46], [55]).

DÉFINITION 2.1. — (i) Soient M et N deux algèbres de von Neumann, une correspondance de M vers N est un espace de Hilbert \mathcal{H} qui est un $N - M$ -bimodule.

(ii) Une correspondance \mathcal{H} de M vers N est finie quand les algèbres de von Neumann M et N et leurs représentations dans \mathcal{H} sont finies.

Pour un facteur de type II_1 le bimodule ${}_M L^2(M, \tau)_M$ où τ est la trace normalisée, joue le rôle de la correspondance identique. Si $A_j \subset M$ sont des sous-algèbres de von Neumann on note ${}_{A_1} L^2(M, \tau)_{A_2}$ la correspondance obtenue par restriction.

Le résultat suivant est dû à ([53]), s'appuyant sur ([26]).

PROPOSITION 2.2. — Soit M un facteur de type II_1 séparable.

(i) Une sous-algèbre involutive abélienne maximale $A \subset M$ est de Cartan si et seulement si la correspondance ${}_A L^2(M, \tau)_A$ est somme directe de correspondances finies.

(ii) Soient $A_1, A_2 \subset M$ deux sous-algèbres de Cartan de M . Alors A_1, A_2 sont conjuguées par un unitaire de M si et seulement si la correspondance ${}_{A_1} L^2(M, \tau)_{A_2}$ est somme directe de correspondances finies.

Passons maintenant aux propriétés H d'approximation compacte et T de rigidité.

Propriété T et inclusions rigides.

Un groupe localement compact G a la propriété T de Kazhdan ssi la représentation triviale est isolée dans le spectre de G .

Pour les facteurs ce sont les correspondances qui jouent le rôle des représentations et un facteur N de type II_1 vérifie la propriété T ssi la correspondance $L^2(N, \tau)$ (qui joue le rôle de la représentation triviale) est isolée ([18]). Autrement dit, un facteur N de type II_1 a la propriété T ssi

¹séparable pour la norme kk_* donnée par la trace.

(T) Il existe $n < \infty$, $x_1, x_2, \dots, x_n \in N$ et $\varepsilon > 0$ tels que si \mathcal{H} est un N -bimodule Hilbertien et $\xi \in \mathcal{H}$ un vecteur de norme 1 tel que $\|x_i \xi - \xi x_i\| \leq \varepsilon, \forall i$, alors \mathcal{H} contient un vecteur non nul ξ_0 tel que $x \xi_0 = \xi_0 x, \forall x \in N$.

Le facteur $L(\Gamma)$ associé à un groupe discret (i.c.c.) Γ a la propriété T ssi le groupe Γ a la propriété T ([18]). Les facteurs de type II_1 ayant la propriété T vérifient des propriétés remarquables comme ([14]).

Si N vérifie T le groupe $\text{Out}(N) := \text{Aut}(N)/\text{Int}(N)$ est dénombrable discret.

Un point essentiel dans la démonstration de la propriété T par Kazhdan pour $SL(n, \mathbb{R})$ (resp. $SL(n, \mathbb{Z})$), $n \geq 3$, consiste à montrer que les représentations de $\mathbb{R}^2 \rtimes SL(2, \mathbb{R})$ proches de la représentation triviale contiennent en fait un vecteur fixe par le sous-groupe \mathbb{R}^2 . Cette propriété de «rigidité relative» a été formulée par Margulis ([39], [32]) de la manière suivante.

Soit $\Gamma \subset \Gamma'$ une inclusion de groupes discrets. Le couple $\Gamma \subset \Gamma'$ a la propriété T ssi

(T_r) Il existe $n < \infty$, $g_1, g_2, \dots, g_n \in \Gamma'$ et $\varepsilon > 0$ tels que si $\pi : \Gamma' \rightarrow \mathcal{U}(\mathcal{H})$ est une représentation unitaire de Γ' dans un espace de Hilbert \mathcal{H} et $\xi \in \mathcal{H}$ vecteur unité satisfait $\|\pi(g_i)\xi - \xi\| < \varepsilon, \forall i$, il existe un vecteur non-nul $\xi_0 \in \mathcal{H}$ tel que $\pi(h)\xi_0 = \xi_0, \forall h \in \Gamma$.

Dans le cas $\Gamma = \Gamma'$ cette condition se réduit à la propriété T de Kazhdan ([35], [22], [59]).

S. Popa a étendu la propriété T des facteurs au cas des inclusions d'algèbres de von Neumann et cette propriété «relative» joue un rôle crucial dans sa démonstration ([51]). La proposition suivante est un exemple typique de la traduction du langage des correspondances (bimodules) à celui des applications complètement positives lesquelles jouent le même rôle que les «coefficients» des représentations dans la théorie des représentations unitaires. Soient N un facteur de type II_1 et τ sa trace normalisée, \mathcal{H} un N -module hilbertien. Nous dirons d'un vecteur $\xi \in \mathcal{H}$ qu'il est δ -tracial ssi $|\langle x\xi, \xi \rangle - \tau(x)| \leq \delta \|x\|, \forall x \in N$.

PROPOSITION 2.3 ([51]). — Soient N un facteur de type II_1 et $B \subset N$ une sous-algèbre de von Neumann. Les conditions suivantes sont équivalentes :

1) Pour tout $\varepsilon > 0$, il existe $F \subset N$ fini et $\delta > 0$ tels que pour tout N -bimodule hilbertien \mathcal{H} et tout vecteur unité δ -bitrivial tel que $\|y\xi - \xi y\| \leq \delta, \forall y \in F$ il existe un vecteur $\xi_0 \in \mathcal{H}$ tel que $\|\xi_0 - \xi\| \leq \varepsilon$ et $b\xi_0 = \xi_0 b, \forall b \in B$.

2) Pour tout $\varepsilon > 0$ il existe $F \subset N$ fini et $\delta > 0$ tels que si $\phi : N \rightarrow N$ est une application complètement positive normale $\tau \circ \phi \leq \tau, \phi(1) \leq 1$ et $\|\phi(x) - x\|_2 \leq \delta, \forall x \in F$, alors $\|\phi(b) - b\|_2 \leq \varepsilon, \forall b \in B, \|b\| \leq 1$.

DÉFINITION 2.4. — Soient N un facteur de type II_1 et $B \subset N$ une sous-algèbre de von Neumann. L'inclusion $B \subset N$ est rigide ssi $B \subset N$ vérifie les conditions équivalentes de la proposition 2.3.

Propriété H d'approximation compacte.

Dans ([31]) Haagerup démontre que les groupes libres $\Gamma = \mathbb{F}_n, 2 \leq n \leq \infty$ vérifient la propriété H suivante : Il existe une suite φ_n de fonctions de type positif telles que

$$(H) \quad \varphi_n \in c_0(\Gamma) \quad \text{et} \quad \forall g \in \Gamma, \quad \varphi_n(g) \rightarrow 1, \quad \text{pour } n \rightarrow \infty.$$

Cette propriété d'approximation de Haagerup ou propriété H est héritée par tout sous-groupe. Elle a été démontrée pour de nombreux groupes discrets comme les sous-groupes discrets de $SO(n, 1)$ ([6]), et de $SU(n, 1)$ ([21]), ou les groupes $SL(2, \mathbb{K})$ pour tout corps de nombres ([8]).

Il est facile de transposer la propriété H pour les facteurs ([13], [9], [18]) : Un facteur N de type II_1 a la propriété H ssi il existe une suite d'applications complètement positives ϕ_k de N dans N telles que $\tau \circ \phi_k \leq \tau$, et en notant $\tilde{\phi}_k$ les opérateurs correspondants dans $L^2(N, \tau)$ tels que l'on ait

$$(\tilde{H}) \quad \tilde{\phi}_k \text{ compact, et } \forall x \in N, \quad \|\phi_k(x) - x\|_2 \rightarrow 0, \quad \text{pour } k \rightarrow \infty.$$

Le facteur $L(\Gamma)$ associé à un groupe (i.c.c.) Γ vérifie \tilde{H} ssi le groupe Γ vérifie H ([9]). La propriété H d'approximation compacte a été étendue au cas d'une inclusion $B \subset N$ dans ([5]). S. Popa considère la variante suivante de cette définition. L'on note $\langle N, B \rangle$ l'algèbre de von Neumann semi-finie commutant de l'action à droite de B dans $L^2(N, \tau)$ et $J(\langle N, B \rangle)$ l'idéal des opérateurs «compacts» de $\langle N, B \rangle$, c'est-à-dire des opérateurs $T \in \langle N, B \rangle$ vérifiant

Pour tout $\varepsilon > 0$, il existe un projecteur fini $p \in \langle N, B \rangle$ tel que $\|T(1 - p)\| < \varepsilon$.

DÉFINITION 2.5. — *Soient N un facteur de type II_1 , τ sa trace normalisée et $B \subset N$ une sous-algèbre de von Neumann.*

N a la propriété H relative à B ssi il existe une suite ϕ_n d'applications complètement positives de N dans N telles que $\tau \circ \phi_n \leq \tau$ vérifiant les trois conditions suivantes

- (i) $\phi_n(bx b') = b \phi_n(x) b' \quad \forall b, b' \in B, x \in N;$
- (ii) $\tilde{\phi}_n \in J(\langle N, B \rangle), \forall n;$
- (iii) $\forall x \in N \quad \|\phi_n(x) - x\|_2 \rightarrow 0, \quad \text{pour } n \rightarrow \infty.$

La propriété (i) montre que l'opérateur $\tilde{\phi}_n$ appartient au commutant de B dans $L^2(N, \tau)$. Dans le cas des sous-algèbres de Cartan $A \subset M$ associées à l'action ergodique d'un groupe discret Γ sur un espace de probabilité (X, μ) , M a la propriété H relative à A ssi le groupe Γ a la propriété H.

DÉFINITION 2.6. — *Soient N un facteur de type II_1 , τ sa trace normalisée et $B \subset N$ une sous-algèbre de von Neumann. On dit que B est une HT-sous-algèbre de N quand*

(i) N a la propriété H relative à B (définition 2.5).

(ii) Il existe une sous-algèbre de von Neumann $B_0 \subset B$ telle que $B'_0 \cap N \subset B$ et $B_0 \subset N$ rigide (définition 2.4).

Le résultat clef de S. Popa est alors le suivant

THÉORÈME 2.7 ([51]). — Soit M un facteur de type II_1 .

1) Toute sous-algèbre abélienne maximale ayant la propriété HT est de Cartan.

2) Deux sous-algèbres abéliennes maximales A_j ayant la propriété HT sont automatiquement conjuguées par un automorphisme intérieur de M .

Pour démontrer 2), on utilise la propriété H de M relativement à A_1 (2.5) pour obtenir une suite d'applications complètement positives compactes par rapport à A_1 et convergeant vers l'identité. Par la propriété de rigidité de $A_2 \subset M$, (2.4), ces applications laissent A_2 uniformément invariant, ce qui produit des $A_1 - A_2$ -bimodules finis, et la proposition 2.2 achève la preuve.

Ce théorème montre que si M est dans la classe \mathcal{HT} , il lui correspond une unique relation d'équivalence ergodique préservant la mesure R_M^{HT} sur l'espace de probabilité (X, μ) , associé à la sous-algèbre de Cartan $A = L^\infty(X, \mu)$ vérifiant HT. En particulier, tout invariant de R_M^{HT} est un invariant de $M \in \mathcal{HT}$. C'est le cas pour les nombres de Betti de la relation d'équivalence et l'on définit $\{\beta_n^{HT}(M)\}_{n \geq 0}$ pour un facteur M de classe \mathcal{HT} comme les nombres de Betti $\{\beta_n(R_M^{HT})\}_n$ ([28]) de la relation d'équivalence R_M^{HT} .

Étant donné un facteur M de type II_1 et un nombre réel $t \in \mathbb{R}_+^*$, on définit le facteur M^t de type II_1 comme l'algèbre réduite du facteur $N := M \otimes I_\infty$ par un projecteur de dimension t . Le facteur M^t est isomorphe à M ssi t appartient au groupe fondamental $F(M)$. L'opération $M \rightarrow M^t$ est bien définie modulo les automorphismes intérieurs et préserve la classe \mathcal{HT} . Il résulte de plus des propriétés générales des nombres de Betti des relations d'équivalence par amplification, ([28]), que l'on a $\beta_n^{HT}(M^t) = \beta_n^{HT}(M)/t, \forall n$. On notera aussi que \mathcal{HT} est stable par produit tensoriel et que l'on a comme corollaire de ([7], [37], [28]) une formule de Künneth pour $\beta_n^{HT}(M_1 \overline{\otimes} M_2)$ en termes des nombres de Betti de $M_1, M_2 \in \mathcal{HT}$.

L'exemple le plus simple d'un facteur dans la classe \mathcal{HT} est $M = L(\Gamma)$, où $\Gamma = \mathbb{Z}^2 \rtimes SL(2, \mathbb{Z})$. Pour montrer que M est dans la classe \mathcal{HT} , on considère la sous-algèbre abélienne A engendrée par $L(\mathbb{Z}^2)$, et l'on utilise la rigidité de Kazhdan pour l'inclusion $\mathbb{Z}^2 \subset \mathbb{Z}^2 \rtimes SL(2, \mathbb{Z})$ et la propriété d'approximation compacte de Haagerup pour le groupe $SL(2, \mathbb{Z})$. Soit A la sous-algèbre abélienne engendrée par $L(\mathbb{Z}^2)$, elle s'identifie à $L^\infty(\mathbb{T}^2, \mu)$ où \mathbb{T}^2 est le groupe dual de \mathbb{Z}^2 et l'inclusion $A \subset M$ s'identifie à celle de $L^\infty(\mathbb{T}^2, \mu)$ dans le produit croisé $L^\infty(\mathbb{T}^2, \mu) \rtimes_\sigma SL(2, \mathbb{Z})$, où σ est l'action correspondante de $SL(2, \mathbb{Z})$. La

relation d'équivalence R_M^{HT} implémentée par l'action de $SL(2, \mathbb{Z})$ sur A a un seul nombre de Betti non-nul, $\beta_1(R_M^{HT}) = \beta_1(SL(2, \mathbb{Z})) = 1/12$ ([4], [7], [28]), et l'on a donc, avec $M = A \rtimes_{\sigma} SL(2, \mathbb{Z})$, l'égalité $\beta_1^{HT}(M) = 1/12$ et $\beta_n^{HT}(M) = 0, \forall n \neq 1$.

COROLLAIRE 2.8 ([51]). — *Soit $\Gamma = \mathbb{Z}^2 \rtimes SL(2, \mathbb{Z})$. Le groupe fondamental du facteur $M = L(\Gamma)$ est trivial, $F(M) = \{1\}$.*

En particulier le facteur M est non-isomorphe à l'algèbre $M_n(M)$ des matrices n fois n sur M , pour $n \geq 2$, ce qui résout le problème 3 de Kadison ([34]) (voir aussi Sakai, Problem 4.4.38 dans [54]). S. Popa a également construit pour tout sous-groupe dénombrable $D \subset \mathbb{R}_+^*$ un facteur tel que $F(M) = D$ ([52]).

3. REMARQUES

1) N. Ozawa a récemment réussi à pousser plus loin les propriétés de rigidité des facteurs possédant la propriété T . Il montre par exemple dans ([43]) qu'il n'existe aucun facteur de type II_1 séparable «universel», i.e. contenant tous les autres. En fait il montre que le groupe unitaire d'un facteur de type II_1 séparable ne peut contenir tous les groupes discrets dénombrables ayant la propriété T .

2) La notion générale d'«hyperbolicité» est bien développée dans la colonne «groupes». Dans ([44]) N. Ozawa montre que dans le facteur $L(\Gamma)$ associé à un groupe hyperbolique Γ le commutant d'une sous-algèbre diffuse est nécessairement une algèbre injective (moyennable). Il obtient ainsi des exemples simples de facteurs «premiers» c'est-à-dire qui ne peuvent se décomposer comme produit tensoriel de deux facteurs de type II_1 . Le premier exemple de tels facteurs est dû à L. Ge ([29]). Enfin Ozawa et Popa ont montré dans ([45]) un résultat d'unique factorisation en facteurs «premiers» associés comme ci-dessus aux groupes hyperboliques. Ce résultat est à rapprocher de celui de Monod et Shalom ([40]) pour les relations d'équivalence.

3) La notion de nombre de Betti est désormais bien établie pour les colonnes «Groupes» et «Relations». Il reste à l'établir pour la colonne «Facteurs». Son utilisation discutée plus haut n'est en elle-même pas suffisante car la classe des facteurs \mathcal{HT} est trop restreinte. De plus la notion de coût aurait suffi pour démontrer le corollaire 2.8 en utilisant ([33], [27]). Les nombres de Betti du groupe $\Gamma = \mathbb{Z}^2 \rtimes SL(2, \mathbb{Z})$ sont tous nuls, car il admet un sous-groupe moyennable normal infini, de sorte que la notion de nombres de Betti pour les facteurs \mathcal{HT} est une notion «relative» et ne correspond pas aux $\beta_n(\Gamma)$ pour $M = L(\Gamma)$. Il serait évidemment très intéressant de développer la notion «absolue» de nombres de Betti pour les facteurs et de les calculer pour les facteurs $L(\Gamma)$ pour les groupes libres.

BIBLIOGRAPHIE

- [1] C. Akemann, M. Walters – *Unbounded negative definite functions*, *Canad. J. Math.* **33** (1981), no. 4, 862–871
- [2] C. Anantharaman-Delaroche – *On Connes' property T for von Neumann algebras*, *Math. Japon.* **32** (1987), 337-355.
- [3] M. Atiyah – *Elliptic operators, discrete groups and von Neumann algebras*, In Colloque «Analyse et Topologie», in honor of Henri Cartan, *Astérisque* **32** (1976), 43-72.
- [4] A. Borel – *The L^2 -cohomology of negatively curved Riemannian symmetric spaces*, *Acad. Sci. Fenn. Ser. A, Math.* **10** (1985), 95-105.
- [5] F. Boca – *On the method for constructing irreducible finite index subfactors of Popa*, *Pac. J. Math.* **161** (1993), 201-231.
- [6] J. de Cannière, U. Haagerup – *Multipliers of the Fourier algebra of some simple Lie groups and their discrete subgroups*, *Amer. J. Math.* **107** (1984), 455-500.
- [7] J. Cheeger, M. Gromov – *L^2 -cohomology and group cohomology*, *Topology* **25** (1986), 189-215.
- [8] P.-A. Cherix, M. Cowling, P. Jolissaint, P. Julg, A. Valette – «Groups with the Haagerup property (Gromov's a-T-menability)», book.
- [9] M. Choda – *Group factors of the Haagerup type*, *Proc. Japan Acad.* **59** (1983), 174-177.
- [10] E. Christensen – *Subalgebras of a finite algebra*, *Math. Ann.* **243** (1979), 17-29.
- [11] A. Connes – *A type II_1 factor with countable fundamental group*, *J. Operator Theory* **4** (1980), 151-153.
- [12] A. Connes – *Classification of injective factors*, *Ann. of Math.* **104** (1976), 73-115.
- [13] A. Connes – *Classification des facteurs*, *Proc. Symp. Pure Math.* **38** (Amer. Math. Soc., 1982), 43-109.
- [14] A. Connes – *Sur la théorie non-commutative de l'intégration*, in «Algèbres d'opérateurs, Séminaire Les Plans-sur-Bex, 1978», pp. 19-143, *Lecture Notes 725*, Springer, Berlin, 1979.
- [15] A. Connes – *Sur la classification des facteurs de type II*, *C. R. Acad. Sci. Paris Sér. I Math.* **281** (1975), A13-A15.
- [16] A. Connes – *Correspondences*, Notes manuscrites, 1980.
- [17] A. Connes – *Noncommutative geometry*, Academic Press, 1994.
- [18] A. Connes, V.F.R. Jones – *Property T for von Neumann algebras*, *Bull. London Math. Soc.* **17** (1985), 57-62.
- [19] A. Connes, V.F.R. Jones – *A II_1 factor with two non-conjugate Cartan subalgebras*, *Bull. Amer. Math. Soc.* **6** (1982), 211-212.
- [20] A. Connes, J. Feldman, B. Weiss – *An amenable equivalence relation is generated by a single transformation*, *Ergodic Th. and Dynam. Systems* **4** (1982), 431-450.

- [21] M. Cowling, U. Haagerup – *Completely bounded multipliers and the Fourier algebra of a simple Lie group of real rank one*, Invent. Math. **96** (1989), 507-549.
- [22] C. Delaroché, A. Kirillov – *Sur les relations entre l'espace dual d'un groupe et la structure de ses sous-groupes fermés*, Sémin. Bourbaki (1967/1968), exp. n° 343, collection hors série de la S.M.F. **10** (1995), 507-528.
- [23] J. Dixmier – *Sous-anneaux abéliens maximaux dans les facteurs de type fini*, Ann. Math. **59** (1954), 279-286.
- [24] J. Dixmier – «Les algèbres d'opérateurs sur l'espace hilbertien (Algèbres de von Neumann)», Gauthier-Villars, Paris, 1957, 1969.
- [25] H. Dye – *On groups of measure preserving transformations*, I, II, Amer. J. Math. **81** (1959), 119-159, and **85** (1963), 551-576.
- [26] J. Feldman, C.C. Moore – *Ergodic equivalence relations, cohomology, and von Neumann algebras I, II*, Trans. Amer. Math. Soc. **234** (1977), 289-324, 325-359.
- [27] D. Gaboriau – *Coût des relations d'équivalence et des groupes*, Invent. Math. **139** (2000), 41-98.
- [28] D. Gaboriau – *Invariants ℓ^2 de relations d'équivalence et de groupes*, Inst. Hautes Études Scient. Publ. (2002).
- [29] L. Ge – *Prime factors*, Proc. Natl. Acad. Sci. USA, **93** (1996), 12762-12763.
- [30] V.Y. Golodets, N.I. Nessonov – *T-property and nonisomorphic factors of type II and III*, J. Funct. Analysis **70** (1987), 80-89.
- [31] U. Haagerup – *An example of non-nuclear C^* -algebra which has the metric approximation property*, Invent. Math. **50** (1979), 279-293.
- [32] P. de la Harpe, A. Valette – «La propriété T de Kazhdan pour les groupes localement compacts», Astérisque **175** (1989).
- [33] G. Hjorth – *A lemma for cost attained*, UCLA preprint 2002.
- [34] R.V. Kadison – *Problems on von Neumann algebras*, Baton Rouge Conference 1867, unpublished.
- [35] D. Kazhdan – *Connection of the dual space of a group with the structure of its closed subgroups*, Funct. Anal. and its Appl. **1** (1967), 63-65.
- [36] G. Levitt – *On the cost of generating an equivalence relation*, Ergod. Th. & Dynam. Sys. **15** (1995), 1173-1181.
- [37] W. Luck – *Dimension theory of arbitrary modules over finite von Neumann algebras and L^2 -Betti numbers II. Applications to Grothendieck groups, L^2 -Euler characteristics and Burnside groups*, J. Reine Angew. Math. **496** (1998), 213-236.
- [38] G. Margulis – *Discrete groups of motion of manifolds of non-positive curvature*, Am. Math. Soc. Translations. **109** (1977), 33-45.
- [39] G. Margulis – *Finitely-additive invariant measures on Euclidian spaces*, Ergodic. Th. and Dynam. Sys. **2** (1982), 383-396.
- [40] N. Monod, Y. Shalom – *Orbit equivalence rigidity and bounded cohomology*, preprint.

- [41] F. Murray, J. von Neumann – *Rings of operators IV*, Ann. Math. **44** (1943), 716-808.
- [42] D. Ornstein, B. Weiss – *Ergodic theory of amenable group actions*, Bull. Amer. Math. Soc. **2** (1980), 161-164.
- [43] N. Ozawa – *There is no separable universal II_1 -factor*, math/0210411.
- [44] N. Ozawa – *Solid von Neumann algebras*, math/0302082.
- [45] N. Ozawa, S. Popa – *Some prime factorisation results for II_1 factors*, math/0302240.
- [46] S. Popa – *Correspondences*, INCREST preprint 1986, unpublished.
- [47] S. Popa – *Classification of subfactors of type II*, Acta Math. **172** (1994), 163-255.
- [48] S. Popa – *Free independent sequences in type II_1 factors and related problems*, Astérisque **232** (1995), 187-202.
- [49] S. Popa – *On a problem of R.V. Kadison on maximal abelian subalgebras*, Invent. Math. **65** (1981), 269-281.
- [50] S. Popa – *The relative Dixmier property for inclusions of von Neumann algebras of finite index*, Ann. Sci. Éc. Norm. Sup. **32** (1999), 743-767.
- [51] S. Popa – *On a class of type II_1 factors with Betti numbers invariants*.
- [52] S. Popa – *Strong rigidity of II_1 factors coming from malleable actions of weakly rigid groups*. math/0305306
- [53] S. Popa, D. Shlyakhtenko – *Cartan subalgebras and bimodule decomposition of II_1 factors*, Math. Scand. **92** (2003), 93-102.
- [54] S. Sakai – «*C*-algebras and W*-algebras*», Springer-Verlag, Berlin-Heidelberg-New York, 1971.
- [55] J.-L. Sauvageot – *Sur le produit tensoriel relatif d'espaces de Hilbert*, J. Operator Theory **9** (1983), 237-252.
- [56] J.-P. Serre – *Arbres, amalgames, $SL(2)$* , Soc. Math. France, Astérisque **46** (1977).
- [57] D. Voiculescu – *The absence of Cartan subalgebras*, Geom. Funct. Anal. **6** (1996), 172-199.
- [58] D. Voiculescu – *The analogues of entropy and of Fisher's information theory in free probability II*, Invent. Math. **118** (1994), 411-440 .
- [59] R. Zimmer – «*Ergodic theory and semisimple groups*», Birkhäuser-Verlag, Boston 1984.

Alain CONNES

I.H.É.S.

35 route de Chartres

F-91440 BURES-sur-YVETTE

E-mail : connes@ihes.fr

**AMIBES DE VARIÉTÉS ALGÈBRIQUES
ET DÉNOMBREMENT DE COURBES**
[d'après G. Mikhalkin]

par **Ilia ITENBERG**

INTRODUCTION

Le nom *amibes* pour les objets que nous allons considérer a été introduit par I. M. Gelfand, M.M. Kapranov et A.V. Zelevinsky [6] en 1994. Des objets similaires surgissaient de temps en temps (voir, par exemple, [3]) avant l'apparition formelle des amibes sur la scène mathématique. (Il faut signaler que le terme *amibe* est utilisé aussi en logique, mais a une signification complètement différente.) Une amibe typique dans \mathcal{R}^2 est une région non bornée qui ressemble un peu au dessin d'une amibe biologique (voir la Figure 1).

Les amibes habitent dans l'intersection de plusieurs domaines des mathématiques tels que la géométrie algébrique, la géométrie symplectique, la topologie, l'analyse complexe et la combinatoire. Elles permettent de visualiser des variétés algébriques et de les étudier par des méthodes non algébriques. Des résultats impressionnants ont été obtenus au cours de ces neuf années passées depuis l'introduction des amibes. Les premières observations fondamentales ont été faites par Gelfand, Kapranov et Zelevinsky [6]. Les résultats concernant les amibes et provenant de leur traitement analytique sont essentiellement dus à M. Forsberg, M. Passare, H. Rullgård et A. Tsikh (voir [7, 27, 30, 31]). Les interprétations topologiques et la plupart des applications des amibes (en particulier, les relations avec la géométrie algébrique réelle) sont dues à G. Mikhalkin [19, 20, 21, 22, 23].

Le texte actuel est principalement consacré aux applications des amibes. Nous verrons d'abord des exemples d'utilisation des amibes en topologie des variétés algébriques réelles (voir le paragraphe 1.3). Les autres applications présentées concernent des questions de géométrie énumérative. Un rôle important dans ces applications énumératives est joué par les amibes *non archimédiennes* (*i.e.*, les amibes de variétés sur un corps valué non archimédien) qui ont été introduites par Kapranov [12]. Elles sont étroitement liées à un domaine relativement nouveau et très passionnant : la *géométrie algébrique tropicale* (voir, par exemple, [36]).

On peut déformer les amibes d'hypersurfaces complexes en amibes non archimédiennes. Ces amibes non archimédiennes sont des complexes polyédraux dans \mathcal{R}^n qui peuvent être considérés comme des variétés sur un semi-anneau tropical $(\max, +)$, et par conséquent,

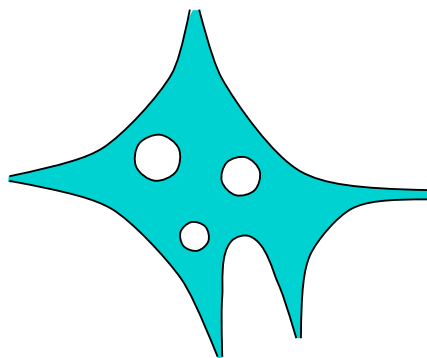


FIGURE 1. Une amibe

portent le nom de *variétés tropicales*. La déformation que l'on vient de mentionner est en rapport direct avec la déquantification de Maslov des nombres réels strictement positifs ([17, 18]), la construction du patchwork due à O. Viro (voir [37, 38] et [28]) et le passage à "*large complex limit*" (voir [16]). M. Kontsevich [16] a proposé d'utiliser les courbes tropicales dans des questions concernant le dénombrement de courbes passant par des points donnés sur une surface complexe. Ce programme a été mis en œuvre par Mikhalkin [22, 23].

Mikhalkin a découvert une correspondance qui fournit un lien direct entre la géométrie algébrique complexe et la géométrie tropicale. Cette correspondance permet de dénombrer certaines courbes nodales de genre donné qui passent par des points génériques donnés sur une surface torique. Formulé de manière informelle, le théorème de Mikhalkin affirme que le nombre de courbes en question est égal au nombre de leurs analogues tropicaux passant par des points génériques donnés dans \mathcal{R}^2 et comptés avec des multiplicités. En plus, Mikhalkin a trouvé un algorithme combinatoire pour dénombrer les courbes tropicales correspondantes.

La correspondance de Mikhalkin permet aussi de dénombrer les courbes réelles passant par des collections spécifiques de points réels sur une surface torique (bien entendu, avec les résultats dépendant du choix de la collection ; voir [22]). Heureusement, une autre découverte importante est tombée à pic : celle faite par J.-Y. Welschinger (voir [41, 42]). Il a trouvé une façon d'attribuer des poids aux courbes réelles rationnelles, conçue pour rendre le nombre de courbes comptées avec poids indépendant de la configuration des points et produire des bornes inférieures pour le nombre de courbes en question. Ainsi est apparu un autre type d'applications de la correspondance de Mikhalkin. Cette approche a déjà donné des résultats, en particulier, l'existence des courbes réelles rationnelles passant par des points réels donnés dans le plan projectif réel (voir [11] et le paragraphe 3.5).

Les textes détaillés de Mikhalkin ne sont pas encore disponibles. Les lecteurs impatients peuvent consulter le texte [33], où E. Shustin, inspiré par les résultats de Mikhalkin et la

technique tropicale, donne une description et une preuve alternatives de la correspondance de Mikhalkin.

Le texte est organisé de la façon suivante. La section 1 contient la définition et une brève description des propriétés des amibes, ainsi que des applications de celles-ci à la topologie des variétés algébriques réelles.

La section 2 est consacrée à la géométrie algébrique tropicale. Le paragraphe 2.1 porte sur la déquantification des nombres réels strictement positifs. Le paragraphe 2.3 contient des notions de bases concernant des variétés tropicales. La définition et les propriétés des amibes non archimédiennes sont présentées dans le paragraphe 2.4.

La section 3 porte sur des résultats énumératifs. Les paragraphes 3.1 et 3.4 sont consacrés aux formulations de questions énumératives (respectivement, complexes et réelles). Les analogues tropicaux des notions algébro-géométriques nécessaires pour le théorème de correspondance sont présentés dans le paragraphe 3.2. Le théorème de correspondance lui-même fait l'objet du paragraphe 3.3. Le paragraphe 3.5 contient des formules combinatoires pour les nombres de courbes étudiés.

Finalement, la section 4 présente les idées des démonstrations des théorèmes principaux.

Remerciements. L'auteur tient à remercier M. Coste, V. Kharlamov, G. Mikhalkin, E. Shustin et O. Viro pour des discussions très utiles. L'auteur remercie G. Mikhalkin pour avoir mis à sa disposition une version préliminaire de [23].

1. AMIBES DE VARIÉTÉS ALGÈBRIQUES

1.1. Définition et propriétés de base

Soient n un entier strictement positif, et $V \subset (\mathcal{C}^*)^n$ une variété algébrique (pour un corps K quelconque, on pose $K^* = K \setminus \{0\}$). Considérons l'application

$$\begin{aligned} \text{Log} : (\mathcal{C}^*)^n &\longrightarrow \mathcal{R}^n, \\ (z_1, \dots, z_n) &\longmapsto (\log |z_1|, \dots, \log |z_n|). \end{aligned}$$

DÉFINITION 1.1. — *L'amibe $\mathcal{A}(V)$ de V est l'image de V par l'application Log .*

Gelfand, Kapranov et Zelevinsky ont mis en évidence les propriétés suivantes des amibes.

- L'amibe $\mathcal{A}(V)$ est un sous-ensemble fermé de \mathcal{R}^n , et le complémentaire de $\mathcal{A}(V)$ est non vide.
- Si $V \subset (\mathcal{C}^*)^n$ est une hypersurface, alors chaque composante connexe du complémentaire de $\mathcal{A}(V) \subset \mathcal{R}^n$ est convexe.

La propriété de convexité des composantes connexes du complémentaire de l'amibe d'une hypersurface ne se transporte pas littéralement au cas des variétés de codimension

plus élevée. Néanmoins, en codimension > 1 , il existe certaines généralisations de la propriété de convexité (voir, par exemple, [9]).

Gelfand, Kapranov et Zelevinsky ont aussi décrit le comportement des amibes à l'infini en termes d'amibes de dimensions plus petites. Soient Δ un polytope convexe à sommets entiers dans \mathcal{R}^n , et Tor_Δ la variété torique associée (voir, par exemple, [8]). La variété Tor_Δ est une compactification de $(\mathcal{C}^*)^n$ telle que l'action multiplicative de $(\mathcal{C}^*)^n$ sur lui-même se prolonge sur Tor_Δ . Le complémentaire de $(\mathcal{C}^*)^n \subset \text{Tor}_\Delta$ se décompose en réunion des variétés toriques associées aux faces de Δ . Notons $\tilde{V} \subset \text{Tor}_\Delta$ la compactification de V .

Le polytope Δ détermine aussi une forme symplectique ω sur Tor_Δ . Cette forme est invariante par l'action du tore $(S^1)^n \subset (\mathcal{C}^*)^n$ formé par les (z_1, \dots, z_n) tels que $|z_1| = \dots = |z_n| = 1$. Considérons l'application des moments correspondante $\mu : \text{Tor}_\Delta \rightarrow \Delta$ (voir, par exemple, [2]).

DÉFINITION 1.2. — *L'amibe compactifiée $\tilde{\mathcal{A}}(\tilde{V})$ de \tilde{V} est $\mu(\tilde{V}) \subset \Delta$.*

Remarquons que l'application $\text{Log} : (\mathcal{C}^*)^n \rightarrow \mathcal{R}^n$ peut aussi être vue comme une application des moments (pour l'action de $(S^1)^n$ sur $(\mathcal{C}^*)^n$ muni de la forme symplectique $\frac{i}{2} \sum_{j=1}^n \frac{dz_j}{z_j} \wedge \frac{d\bar{z}_j}{\bar{z}_j}$).

Les applications $\mu|_{(\mathcal{C}^*)^n}$ et Log sont des submersions et ont les mêmes tores réels pour fibres. Par conséquent, $\mathcal{A}(V)$ peut être envoyé sur $\tilde{\mathcal{A}}(\tilde{V}) \cap \text{Int}(\Delta)$ (où $\text{Int}(\Delta)$ est l'intérieur de Δ) par un difféomorphisme de \mathcal{R}^n sur $\text{Int}(\Delta)$.

Soient Δ' une face de Δ , et $\text{Tor}_{\Delta'} \subset \text{Tor}_\Delta$ la variété torique associée à Δ' . Posons $\tilde{V}' = \tilde{V} \cap \text{Tor}_{\Delta'}$.

PROPOSITION 1.3 (Gelfand, Kapranov et Zelevinsky, [6]). — *On a*

$$\tilde{\mathcal{A}}(\tilde{V}') = \tilde{\mathcal{A}}(\tilde{V}) \cap \Delta'.$$

L'amibe d'une courbe $V \subset (\mathcal{C}^*)^2$ atteint l'infini par des "tentacules" (voir la Figure 1). Chaque tentacule contient exactement un rayon, et tend vers ce rayon en s'approchant de l'infini. Le rayon contenu dans un tentacule est orthogonal à un côté de Δ (on dit que le tentacule est associé à ce côté de Δ). Le nombre de tentacules associés à un côté c est strictement positif et inférieur ou égal à la longueur entière de c (*i.e.*, au nombre de segments en lesquels c est divisé par ses points entiers).

1.2. Amibes d'hypersurfaces

Soit $f(\mathbf{z}) = \sum_{\mathbf{i}} a_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$ un polynôme complexe à n variables (ici $\mathbf{i} = (i_1, \dots, i_n)$ est un multi-indice, $\mathbf{z} = (z_1, \dots, z_n)$ et $\mathbf{z}^{\mathbf{i}} = z_1^{i_1} \dots z_n^{i_n}$), et $V \subset (\mathcal{C}^*)^n$ l'hypersurface définie par f dans $(\mathcal{C}^*)^n$. L'enveloppe convexe Δ des points $\mathbf{i} \in \mathcal{Z}^n \subset \mathcal{R}^n$ tels que $a_{\mathbf{i}} \neq 0$ s'appelle le *polytope de Newton* de f (on dit aussi que Δ est le polytope de Newton de V).

À chaque composante connexe du complémentaire de $\mathcal{A}(V)$ dans \mathcal{R}^n , on peut associer, d'une façon naturelle, un point entier de Δ .

THÉORÈME 1.4 (M. Forsberg, M. Passare et A. Tsikh, [7]). — *Il existe une fonction localement constante $\text{ind} : \mathcal{R}^n \setminus \mathcal{A}(V) \rightarrow \Delta \cap \mathcal{Z}^n$ qui envoie des composantes connexes différentes du complémentaire de $\mathcal{A}(V)$ sur des points entiers différents de Δ . En particulier, le nombre de composantes connexes de $\mathcal{R}^n \setminus \mathcal{A}(V)$ est inférieur ou égal au nombre de points entiers de Δ .*

REMARQUE 1.5. — Mikhalkin a proposé une interprétation topologique de la démonstration du Théorème 1.4. Pour décrire une fonction ind ayant les propriétés formulées ci-dessus, choisissons un point x dans $\mathcal{R}^n \setminus \mathcal{A}(V)$. L'image réciproque de x par Log est un tore dont l'intersection avec V est vide. Pour chaque entier $1 \leq j \leq n$, considérons un j -ème méridien m_j de ce tore, *i.e.*, un cercle le long duquel toutes les coordonnées sauf la j -ème sont constantes. Soit D_j un disque de bord m_j dans \mathcal{C}^n . Notons κ_j l'indice d'intersection de D_j et de l'adhérence de V dans \mathcal{C}^n . L'entier κ_j ne dépend pas du choix de m_j et D_j . L'application $\text{ind} : \mathcal{R}^n \setminus \mathcal{A}(V) \rightarrow \mathcal{Z}^n$ associant à x le point $(\kappa_1, \dots, \kappa_n)$ est localement constante et envoie des composantes connexes différentes de $\mathcal{R}^n \setminus \mathcal{A}(V)$ sur des points entiers différents de Δ . On renvoie à [19] pour les détails.

La borne supérieure pour le nombre de composantes connexes de $\mathcal{R}^n \setminus \mathcal{A}(V)$ formulée dans le Théorème 1.4 est atteinte : pour tout polytope convexe Δ à sommets entiers dans \mathcal{R}^n , il existe une hypersurface $V \subset (\mathcal{C}^*)^n$ de polytope de Newton Δ telle que le nombre de composantes connexes de $\mathcal{R}^n \setminus \mathcal{A}(V)$ soit égal au nombre de points entiers de Δ (voir, par exemple, [20]).

Un outil analytique très important dans l'étude des amibes est la *fonction de Ronkin*. Pour un polynôme $f(\mathbf{z}) = \sum_{\mathbf{i}} a_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$, la fonction de Ronkin $N_f : \mathcal{R}^n \rightarrow \mathcal{R}$ est définie par

$$N_f(\mathbf{x}) = \frac{1}{(2\pi i)^n} \int_{\text{Log}^{-1}(\mathbf{x})} \log |f(\mathbf{z})| \frac{dz_1}{z_1} \wedge \dots \wedge \frac{dz_n}{z_n}.$$

(On considère ici la fonction $\log |f| : (\mathcal{C}^*)^n \rightarrow \mathcal{R} \cup \{-\infty\}$ en posant $\log(0) = -\infty$.) L'énoncé suivant est un corollaire du fait que $\log |f|$ est une fonction plurisousharmonique qui est pluriharmonique sur $(\mathcal{C}^*)^n \setminus V$ et strictement plurisousharmonique sur V .

PROPOSITION 1.6 (L. Ronkin, M. Passare et H. Rullgård, [29], [27]). — *La fonction de Ronkin N_f est convexe. Elle est affine sur chaque composante connexe de $\mathcal{R}^n \setminus \mathcal{A}(V)$ et strictement convexe sur $\mathcal{A}(V)$.*

Remarquons que l'existence d'une fonction ayant les propriétés mentionnées dans la Proposition 1.6 implique que chaque composante connexe de $\mathcal{R}^n \setminus \mathcal{A}(V)$ est convexe.

Soit $\nabla N_f : \mathcal{R}^n \rightarrow \mathcal{R}^n$ le gradient de N_f . D'après la Proposition 1.6, le gradient ∇N_f est constant sur chaque composante connexe de $\mathcal{R}^n \setminus \mathcal{A}(V)$. La proposition suivante peut être démontrée à l'aide de la formule de Jensen.

PROPOSITION 1.7 (M. Passare et H. Rullgård, [27]). — *On a*

$$\text{Int}\Delta \subset \nabla N_f(\mathcal{R}^n) \subset \Delta.$$

Pour tout $x \in \mathcal{R}^n \setminus \mathcal{A}(V)$, l'image $\nabla N_f(x)$ de x est un point entier de Δ . En plus, la restriction de ∇N_f à $\mathcal{R}^n \setminus \mathcal{A}(V)$ coïncide avec la fonction ind du Théorème 1.4 et de la Remarque 1.5.

Puisque N_f est affine sur chaque composante connexe de $\mathcal{R}^n \setminus \mathcal{A}(V)$, on peut définir une fonction N_f^∞ convexe et affine par morceaux en posant $N_f^\infty = \max_{\mathcal{F}} N_{\mathcal{F}}$, où \mathcal{F} parcourt toutes les composantes connexes de $\mathcal{R}^n \setminus \mathcal{A}(V)$, et $N_{\mathcal{F}} : \mathcal{R}^n \rightarrow \mathcal{R}$ est la fonction affine dont la restriction à \mathcal{F} coïncide avec N_f .

Passare et Rullgård [27] ont proposé la définition suivante.

DÉFINITION 1.8. — *Le squelette Sq de l'amibe $\mathcal{A}(V)$ est le lieu des coins de N_f^∞ , i.e., le sous-ensemble de \mathcal{R}^n où N_f^∞ n'est pas lisse.*

Clairement, Sq est un complexe polyédral qui est contenu dans $\mathcal{A}(V)$. En plus, le squelette Sq est un rétract par déformation de $\mathcal{A}(V)$ (voir [27] et [31]).

Nous interrompons ici la présentation des amibes. D'autres propriétés des ces objets remarquables peuvent être trouvées dans les articles de synthèse [20] et [25]. Nous allons nous intéresser aux applications des amibes.

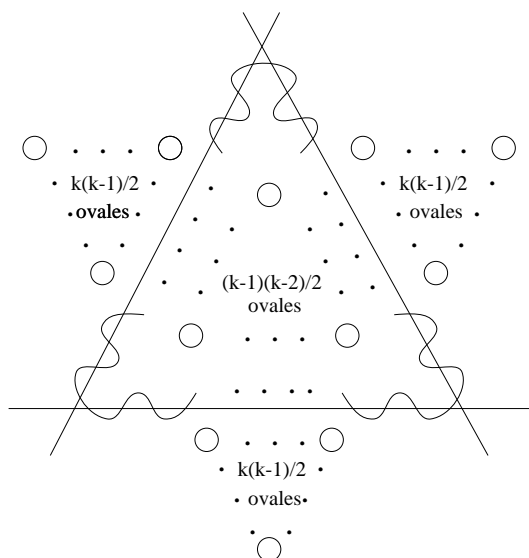
1.3. Premières applications

Les premières applications significatives des amibes ont été proposées par Mikhalkin dans le cadre de la première partie du 16-ème problème de Hilbert (voir [10]). Elles sont liées, en particulier, à la question sur les configurations possibles des composantes connexes (de l'ensemble des points réels) d'une courbe non singulière de degré donné dans le plan projectif réel $\mathcal{R}P^2$. Soient m un entier strictement positif, et $\mathcal{R}X$ une courbe non singulière de degré m dans $\mathcal{R}P^2$. Quels sont les types topologiques possibles de la paire $(\mathcal{R}P^2, \mathcal{R}X)$? La réponse complète à cette question n'est connue que pour $m \leq 7$.

Le nombre de types topologiques réalisables par $(\mathcal{R}P^2, \mathcal{R}X)$, où $\mathcal{R}X$ est une courbe non singulière de degré m , croît de façon exponentielle quand m tend vers l'infini (voir [13, 14] pour des résultats sur l'asymptotique de ce nombre). Beaucoup de restrictions importantes sur les types topologiques en question peuvent être trouvées, par exemple, dans les articles de synthèse [43], [39] et [5]. Le cas le plus restrictif et le plus intéressant est celui des courbes maximales. (Une courbe algébrique non singulière dans $\mathcal{R}P^2$ est dite *maximale* si le nombre de ses composantes connexes est maximal pour le degré donné. Ce nombre maximal est égal à $\frac{(m-1)(m-2)}{2} + 1$ pour le degré m .) Le nombre de types topologiques des paires $(\mathcal{R}P^2, \mathcal{R}X)$ réalisables par les courbes maximales de degré donné croît aussi de façon exponentielle quand le degré tend vers l'infini.

La situation change si on considère $\mathcal{R}P^2$ comme une compactification de $(\mathcal{R}^*)^2$, i.e., si on tient compte de la position d'une courbe étudiée par rapport aux trois droites qui forment $\mathcal{R}P^2 \setminus (\mathcal{R}^*)^2$.

Soient l_1, \dots, l_s des droites en position générique dans $\mathcal{R}P^2$.

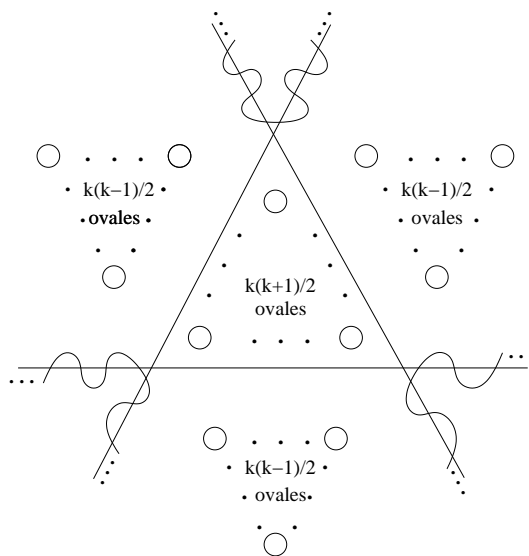
FIGURE 2. Le $(2k, 3)$ -type topologique maximal

DÉFINITION 1.9. — On dit qu'une courbe $\mathcal{R}X$ de degré m dans $\mathcal{R}P^2$ est en position maximale par rapport à la collection l_1, \dots, l_s , si $\mathcal{R}X$ est maximale et s'il existe s arcs disjoints r_1, \dots, r_s sur une composante connexe de $\mathcal{R}X$ tels que, pour chaque $j = 1, \dots, s$, l'arc r_j a m points d'intersection avec l_j .

Le type topologique d'une triade $(\mathcal{R}P^2; \mathcal{R}X, l_1 \cup \dots \cup l_s)$, où $\mathcal{R}X$ est une courbe de degré m en position maximale par rapport à l_1, \dots, l_s , s'appelle un (m, s) -type topologique maximal.

THÉORÈME 1.10 (Mikhalkin, [19]). — Pour tout entier strictement positif m , il existe exactement un $(m, 3)$ -type topologique maximal ; il est représenté sur les Figures 2 et 3 (cas m pair et m impair, respectivement). Il n'existe pas de (m, s) -type topologique maximal avec $s > 3$ et $m \geq 3$.

La deuxième affirmation du Théorème 1.10 est un corollaire de la première. Pour démontrer la première affirmation, Mikhalkin considère une courbe $\mathcal{R}X$ de degré m en position maximale par rapport à trois droites l_1, l_2 et l_3 dans $\mathcal{R}P^2$, identifie le complémentaire de ces trois droites avec $(\mathcal{R}^*)^2$ (en prenant l_1, l_2 et l_3 pour des axes de coordonnées), et étudie l'amibe de la courbe correspondante V dans $(\mathcal{C}^*)^2$. (Puisqu'on considère maintenant $\mathcal{C}P^2$ comme une compactification de $(\mathcal{C}^*)^2$, on va utiliser la notation $\mathcal{R}\tilde{V}$ pour $\mathcal{R}X$.) La démonstration est basée sur le résultat suivant : la condition de maximalité implique que le lieu critique de $\text{Log}|_V$ coïncide avec l'ensemble $\mathcal{R}V$ des points réels de V , et que la frontière de l'amibe $\mathcal{A}(V)$ coïncide avec $\text{Log}(\mathcal{R}V)$. En plus, pour chaque composante connexe \mathcal{F} de $\mathcal{R}^2 \setminus \mathcal{A}(V)$, la description topologique de la fonction ind (voir la Remarque 1.5) permet de retrouver le quadrant de $(\mathcal{R}^*)^2$ contenant les points de $\mathcal{R}V$ qui

FIGURE 3. Le $(2k + 1, 3)$ -type topologique maximal

sont envoyés par Log sur la frontière de \mathcal{F} (en fait, le quadrant est déterminé par les parités des coordonnées de $\text{ind}(x)$, où $x \in \mathcal{F}$). On renvoie à [19] et [20] pour les détails.

Le Théorème 1.10 a une généralisation qui s'applique à d'autres surfaces toriques (voir [19]). Il a aussi une version concernant des hypersurfaces de dimension plus élevée (voir [19]), mais, en l'état actuel, les énoncés du théorème deviennent plus faibles quand la dimension augmente.

Les courbes maximales de degré pair (respectivement, impair) dont la position par rapport aux axes de coordonnées est décrite sur la Figure 2 (respectivement, sur la Figure 3) s'appellent les *courbes simples de Harnack*. Elles ont beaucoup de propriétés géométriques remarquables. Par exemple, si $\mathcal{R}\tilde{V} \subset \mathcal{R}P^2$ est une courbe simple de Harnack, alors $\mathcal{R}V \subset (\mathcal{R}^*)^2$ n'a pas de point d'inflexion logarithmique réel, *i.e.*, n'a pas de point $\mathbf{x} \in \mathcal{R}V$ tel que $\text{Log}(\mathbf{x})$ soit un point d'inflexion de $\text{Log}(\mathcal{R}V)$ (voir [19]).

Les amibes des courbes simples de Harnack ont la plus grande aire parmi les amibes des courbes de degré donné dans $\mathcal{R}P^2$ (voir [24]). L'aire de l'amibe d'une courbe $V \subset (\mathcal{C}^*)^2$ est toujours finie et peut être majorée en termes du polygone de Newton de V . Passare et Rullgård [27] ont proposé la définition suivante : la *mesure de Monge-Ampère* sur $\mathcal{A}(V)$ est la tirée en arrière de la mesure de Lebesgue sur $\Delta \subset \mathcal{R}^2$ par ∇N_f (où f est un polynôme définissant la courbe V). Le théorème suivant provient d'une comparaison des mesures de Lebesgue et de Monge-Ampère sur $\mathcal{A}(V)$.

THÉORÈME 1.11 (Passare et Rullgård, [27]). — *Soit $V \subset (\mathcal{C}^*)^2$ une courbe algébrique ayant le polygone de Newton Δ . Alors, $\text{Aire } \mathcal{A}(V) \leq \pi^2 \text{Aire } \Delta$.*

Ce théorème est spécifique au cas des amibes dans \mathcal{R}^2 , car la Proposition 1.3 implique que les amibes d'hypersurfaces dans $(\mathcal{C}^*)^n$, $n > 2$, ont en général un volume infini.

2. GÉOMÉTRIE ALGÈBRIQUE TROPICALE

2.1. Déquantification des nombres réels strictement positifs

Considérons une famille de semi-anneaux $\{S_t\}$, $t \in (1, +\infty]$. Comme ensemble, chaque semi-anneau S_t coïncide avec \mathcal{R} . Les opérations d'addition et multiplication dans S_t sont définies de la manière suivante :

$$a \oplus_t b = \begin{cases} \log_t(t^a + t^b), & \text{si } t \neq +\infty, \\ \max\{a, b\}, & \text{si } t = +\infty ; \end{cases}$$

$$a \odot_t b = a + b.$$

Ces opérations dépendent de t de façon continue. Chaque semi-anneau S_t avec une valeur finie de t est isomorphe au semi-anneau \mathcal{R}_+^* des nombres réels strictement positifs (munis des opérations habituelles d'addition et multiplication) : l'application $x \mapsto \log_t x$ effectue un isomorphisme entre \mathcal{R}_+^* et S_t . Par contre, S_∞ n'est pas isomorphe à \mathcal{R}_+^* . C'est un semi-anneau idempotent qui s'appelle un *semi-anneau* $(\max, +)$ et noté $\mathcal{R}_{\text{trop}}$. Le semi-anneau $\mathcal{R}_{\text{trop}}$ est un des semi-anneaux *tropicaux* (voir, par exemple, [26]). Les opérations \oplus_∞ et \odot_∞ dans $\mathcal{R}_{\text{trop}}$ s'appellent l'*addition tropicale* et la *multiplication tropicale*, respectivement.

Le passage de valeurs finies de t à l'infini dans la famille $\{S_t\}$ s'appelle la *déquantification de Maslov des nombres réels strictement positifs* (voir [17] et [18]). Des déformations similaires sont connues dans plusieurs domaines des mathématiques. Comme il a été remarqué par O. Viro [40], la déquantification de Maslov est directement liée au patchwork, la méthode de construction de variétés algébriques réelles proposée par Viro il y a une vingtaine d'années (voir [37, 38] et [28]). La déquantification de Maslov est aussi directement liée au passage à "*large complex limit*" (voir [16]) qui fait dégénérer une structure complexe sur une variété (cf. la section 4). Toutes ces déformations fournissent un lien très important entre la géométrie algébrique et la géométrie des complexes polyédraux. C'est ce lien qui a été exploité par Mikhalkin pour obtenir les résultats énumératifs présentés dans la section 3.

Nous allons maintenant faire une brève description de la géométrie algébrique sur le semi-anneau $\mathcal{R}_{\text{trop}}$. Cette description étant orientée vers les problèmes énumératifs présentés par la suite est, bien sûr, incomplète. On renvoie à [36] pour une information plus complète sur les variétés tropicales.

2.2. Fonctions convexes affines par morceaux et transformée de Legendre

Soient n un entier strictement positif, I une collection finie de points entiers dans \mathcal{R}^n , et $\nu : I \rightarrow \mathcal{R}$ une fonction. Notons $\Delta(I)$ l'enveloppe convexe de I . La fonction ν définit une subdivision de $\Delta(I)$ de la façon suivante. Considérons l'enveloppe convexe Γ_ν de l'ensemble $\{(\mathbf{i}, i_{n+1}) \in \mathcal{R}^{n+1} : \mathbf{i} \in I, i_{n+1} \geq \nu(\mathbf{i})\}$. Le polyèdre Γ_ν se projette

naturellement sur $\Delta(I)$. Les faces de Γ_ν qui se projettent de manière injective définissent une subdivision de $\Delta(I)$. Notons $S(I, \nu)$ cette subdivision.

La fonction $\hat{\nu} : \mathcal{R}^n \rightarrow \mathcal{R}$ définie par $\hat{\nu}(\mathbf{x}) = \max_{\mathbf{i} \in I} \{ \langle \mathbf{i}, \mathbf{x} \rangle - \nu(\mathbf{i}) \}$, où $\langle \mathbf{i}, \mathbf{x} \rangle$ est le produit scalaire standard de \mathbf{i} et \mathbf{x} , s'appelle la *transformée de Legendre* de ν . C'est une fonction convexe affine par morceaux. Considérons le lieu des coins de $\hat{\nu}$. La subdivision $\hat{S}(\hat{\nu})$ de \mathcal{R}^n définie par ce lieu des coins est duale de la subdivision $S(I, \nu)$. (Soient Υ un polyèdre de $S(I, \nu)$, et Υ_ν la face de Γ_ν qui se projette sur Υ . Le polyèdre dual de Υ est formé par les \mathbf{x} tels que $(\mathbf{x}, -1)$ appartienne au cône dual du cône de $\Upsilon_\nu \subset \Gamma_\nu$.) Un polyèdre dans $S(I, \nu)$ et le polyèdre correspondant dans $\hat{S}(\hat{\nu})$ n'ont pas forcément de point commun, mais leurs enveloppes affines sont orthogonales (pour plus d'information sur la transformation de Legendre, voir, par exemple, [1]).

2.3. Variétés tropicales

Comme dans le paragraphe précédent, soient n un entier strictement positif, et I une collection finie de points entiers dans \mathcal{R}^n . Considérons un polynôme tropical $P(\mathbf{x}) = \sum_{\mathbf{i} \in I} a_{\mathbf{i}} \mathbf{x}^{\mathbf{i}}$ en n variables x_1, \dots, x_n et à support dans I (ici $\mathbf{i} = (i_1, \dots, i_n)$ est un multi-indice, $\mathbf{x}^{\mathbf{i}} = x_1^{i_1} \dots x_n^{i_n}$, les coefficients $a_{\mathbf{i}}$ sont des nombres réels, et les opérations d'addition et de multiplication sont tropicales). Par définition des opérations tropicales, $P(\mathbf{x})$ est le maximum des fonctions affines $\langle \mathbf{i}, \mathbf{x} \rangle + a_{\mathbf{i}}$ ($\mathbf{i} \in I$), où le signe “+” désigne l'addition habituelle. Donc, P est une fonction convexe affine par morceaux qui est la transformée de Legendre de la fonction $\nu : \mathbf{i} \mapsto -a_{\mathbf{i}}$.

L'*hypersurface tropicale* $T(P)$ définie par P est le lieu des coins de P (cf., par exemple, [21] et [36]). Autrement dit, $T(P)$ est l'ensemble des points de \mathcal{R}^n pour lesquels le maximum des fonctions affines $\langle \mathbf{i}, \mathbf{x} \rangle + a_{\mathbf{i}}$ est réalisé par au moins deux fonctions. Bien sûr, $T(P)$ est complètement déterminée par la paire (I, ν) . Parfois, on dit que $T(P)$ est l'hypersurface tropicale associée à (I, ν) . L'enveloppe convexe $\Delta(I)$ de I s'appelle un *polytope de Newton* de $T(P)$. La définition des hypersurfaces tropicales peut être généralisée au cas de variétés tropicales arbitraires (voir, par exemple, [36]). Ici, on va se limiter aux courbes tropicales dans \mathcal{R}^2 .

Une *droite tropicale* est la courbe tropicale définie par un polynôme tropical $ax + by + c$ (l'addition et la multiplication sont, bien sûr, tropicales). Une droite tropicale est une réunion de trois rayons ayant la même extrémité. Cette réunion peut être obtenue par une translation de la réunion des rayons $\{y = 0, x \leq 0\}$, $\{x = 0, y \leq 0\}$ et $\{x = y, x \geq 0\}$ (voir la Figure 4). En particulier, une droite tropicale est entièrement déterminée par la position de l'extrémité commune des rayons. Remarquons que deux droites tropicales génériques ont exactement un point en commun, et que, par deux points génériques du plan, on peut faire passer exactement une droite tropicale. Plus généralement, une courbe tropicale est dite *plane et projective de degré m* (où m est un entier strictement positif) si elle a le triangle à sommets $(0, 0)$, $(m, 0)$ et $(0, m)$ dans \mathcal{R}^2 pour polygone de Newton.

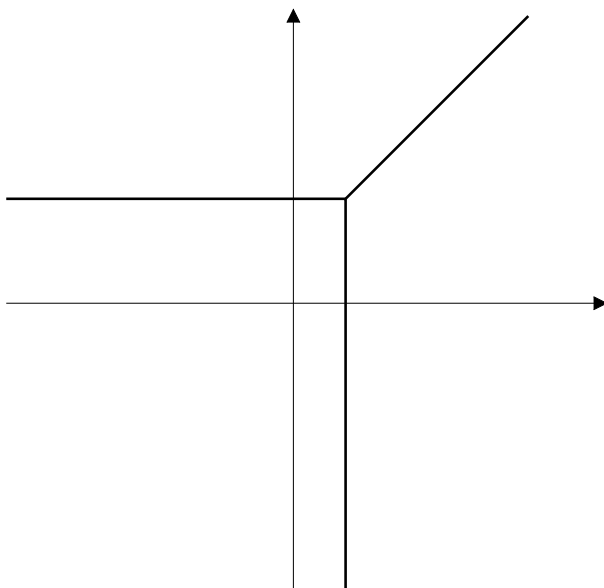


FIGURE 4. Une droite tropicale

Soit P un polynôme tropical en deux variables. La courbe tropicale $T(P)$ définie par P est une réunion de segments et de demi-droites dans \mathcal{R}^2 . Associons à chaque segment et à chaque demi-droite de $T(P)$ un nombre entier strictement positif de la façon suivante. Soit e un segment ou une demi-droite de $T(P)$. Alors, e est la projection d'une arête de la surface affine par morceaux formée par le graphe de P . Cette arête provient de l'intersection de deux plans. Soient $\langle \mathbf{i}^{(1)}, \mathbf{x} \rangle + a_{\mathbf{i}}^{(1)}$ et $\langle \mathbf{i}^{(2)}, \mathbf{x} \rangle + a_{\mathbf{i}}^{(2)}$ les deux fonctions affines correspondantes. Le poids associé à e est égal à la longueur entière du segment reliant $\mathbf{i}^{(1)}$ et $\mathbf{i}^{(2)}$. À partir de maintenant, on va toujours supposer que les segments et les demi-droites d'une courbe tropicale dans \mathcal{R}^2 sont munis des poids définis ci-dessus.

Les courbes tropicales dans \mathcal{R}^2 peuvent être décrites de la manière suivante. Un sous-ensemble de \mathcal{R}^2 est appelé un *graphe rectiligne* s'il peut être représenté comme une réunion finie de cellules fermées de dimension 0 (*sommets*) et de dimension 1 (*arêtes*) telles que

- chaque arête est un segment ou une demi-droite,
- la pente de chaque arête est rationnelle,
- le bord de chaque arête est une réunion de sommets,
- les intérieurs relatifs de cellules différentes sont disjoints.

Un graphe rectiligne est dit *pondéré* si chaque arête de ce graphe est munie d'un nombre entier strictement positif (ce nombre est appelé le *poids* de l'arête).

Chaque courbe tropicale dans \mathcal{R}^2 est un graphe rectiligne pondéré. En plus, on a la caractérisation suivante (voir, par exemple, [21]).

PROPOSITION 2.1. — *Un graphe rectiligne pondéré Γ représente une courbe tropicale si et seulement si, pour tout sommet v de Γ , on a $\alpha_1 e_1 + \dots + \alpha_k e_k = 0$, où e_1, \dots, e_k sont*

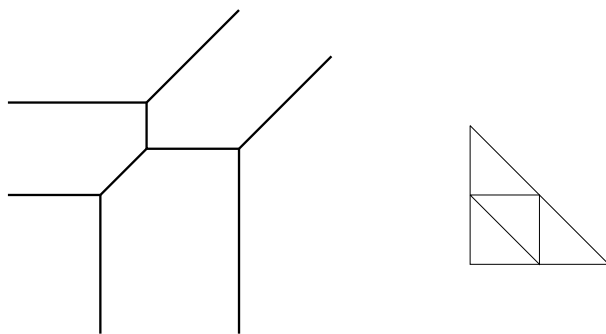


FIGURE 5. Une courbe tropicale et la subdivision correspondante

les vecteurs entiers primitifs (*i.e.*, non divisibles) sortant de v le long des arêtes de Γ , et α_i est le poids de l'arête correspondante à e_i ($i = 1, \dots, k$).

Une généralisation de cette caractérisation des courbes tropicales au cas d'hypersurfaces de dimension plus élevée peut être trouvée dans [21].

La même courbe tropicale peut être définie par différents polynômes tropicaux. Néanmoins, à partir d'une courbe tropicale (avec des poids associés aux segments et demi-droites), on peut reconstruire son polygone de Newton à translation près, ainsi que la subdivision $S(I, \nu)$ de $\Delta(I)$, une fois que le polygone de Newton $\Delta(I)$ est fixé. Pour une courbe tropicale T et son polygone de Newton Δ , notons $D_\Delta(T)$ la subdivision correspondante de Δ .

Le théorème tropical de Bernstein est une généralisation du fait que deux droites tropicales génériques ont exactement un point en commun. Avant de formuler le théorème, définissons la multiplicité d'un point d'intersection de deux courbes tropicales dans \mathcal{R}^2 . Considérons un point d'intersection isolé d'un segment (ou d'un rayon) s_1 d'une courbe tropicale T_1 et d'un segment (ou d'un rayon) s_2 d'une courbe tropicale T_2 . Soient (a_i, b_i) un vecteur primitif le long de s_i , et α_i le poids de s_i ($i = 1, 2$). Alors, la *multiplicité* du point d'intersection de s_1 et s_2 est égale à $\alpha_1 \alpha_2 |a_1 b_2 - a_2 b_1|$.

Le théorème suivant est un corollaire du fait que la réunion de deux courbes tropicales génériques dans \mathcal{R}^2 forme une subdivision duale d'une subdivision mixte de la somme de Minkowski des polygones de Newton de ces courbes (voir [35] pour la définition d'une subdivision mixte).

THÉORÈME 2.2 (Théorème tropical de Bernstein, (voir, par exemple, [36])). — *Le nombre de points d'intersection, comptés avec les multiplicités, de deux courbes tropicales génériques dans \mathcal{R}^2 est égal à l'aire mixte des polygones de Newton de ces courbes.*

COROLLAIRE 2.3 (Théorème tropical de Bézout). — *Le nombre de points d'intersection, comptés avec les multiplicités, de deux courbes tropicales génériques planes et projectives de degrés m_1 et m_2 , respectivement, est égal à $m_1 m_2$.*

Le fait que, par deux points génériques du plan, on peut faire passer exactement une droite tropicale a aussi une généralisation : c'est le théorème de Mikhalkin d'énumération de courbes que l'on va présenter dans la section 3.

Pour terminer ce paragraphe, introduisons les courbes tropicales paramétrées. Soient $\bar{\Lambda}$ un graphe (abstrait) fini pondéré, et $\Lambda \subset \bar{\Lambda}$ le complémentaire des sommets de puissance 1 dans $\bar{\Lambda}$. Une application propre $\theta : \Lambda \rightarrow \mathcal{R}^2$ s'appelle une *courbe tropicale paramétrée* si

- pour toute arête e de Λ , la restriction $\theta|_e$ est un plongement, et l'image $\theta(e)$ de e est contenue dans une droite ayant une pente rationnelle,
- pour tout sommet v de Λ , on a $\alpha_1 e_1 + \dots + \alpha_k e_k = 0$, où e_1, \dots, e_k sont les vecteurs entiers primitifs sortant de $\theta(v)$ le long des images des arêtes de Λ qui sont adjacentes à v , et α_i est le poids de l'arête correspondante à e_i ($i = 1, \dots, k$).

L'image $\theta(\Lambda)$ peut être vue comme un graphe rectiligne pondéré représentant une courbe tropicale. On dit que $\theta : \Lambda \rightarrow \theta(\Lambda)$ est une *paramétrisation* de la courbe tropicale $\theta(\Lambda)$.

2.4. Amibes non archimédiennes

Dans le paragraphe précédent, nous avons vu que les courbes tropicales ont plusieurs propriétés en commun avec les courbes algébriques. Une des explications de ce phénomène est le fait que chaque courbe tropicale peut être vue comme l'amibe d'une courbe algébrique sur un corps valué non archimédien.

Soient K un corps avec une norme, n un entier strictement positif, et $V \subset (K^*)^n$ une variété algébrique. La définition de l'amibe se généralise facilement à cette situation : l'amibe $\mathcal{A}(V)$ est l'image de V par l'application

$$\begin{aligned} \text{Log} : (K^*)^n &\longrightarrow \mathcal{R}^n, \\ (z_1, \dots, z_n) &\longmapsto (\log |z_1|, \dots, \log |z_n|), \end{aligned}$$

où $|z_i|$ est la norme de z_i .

Une *valuation* (à valeurs réelles) d'un corps K est une fonction $v : K^* \rightarrow \mathcal{R}$ telle que, pour tous z et w dans K , on a $v(zw) = v(z) + v(w)$ et $v(z + w) \geq \min\{v(z), v(w)\}$. Un corps muni d'une valuation est dit *valué non archimédien*. Si K est un corps valué non archimédien et v est sa valuation, on définit la norme $|z|$ de $z \in K^*$ par $|z| = e^{-v(z)}$, et on pose $|0| = 0$. L'amibe d'une variété algébrique $V \subset (K^*)^n$ est l'image de V par l'application $\text{Log} : (z_1, \dots, z_n) \mapsto (-v(z_1), \dots, -v(z_n))$.

Un exemple important d'un corps valué non archimédien est le corps $\cup_{l \geq 1} k((t^{1/l}))$ des séries de Puiseux sur un corps k . Les éléments de ce corps sont les séries formelles $b(t) = \sum_{r \in S} b_r t^r$ à coefficients dans k et à une variable t telles que $S \subset \mathcal{Q}$ soit minoré et contenu dans une suite arithmétique. La valuation est donnée par le plus petit r tel que $b_r \neq 0$. Si k est algébriquement clos et de caractéristique 0, alors le corps des séries de Puiseux sur k est aussi algébriquement clos (voir, par exemple, [44]).

Un autre exemple est le corps \mathcal{K} des séries transfinies $b(t) = \sum_{r \in S} b_r t^r$ à coefficients complexes et exposants réels (ici $S \subset \mathcal{R}$ est un ensemble bien ordonné, *i.e.*, chaque sous-ensemble de S contient un élément minimal). Ce corps est algébriquement clos. La valuation non archimédienne $v : \mathcal{K}^* \rightarrow \mathcal{R}$ est encore une fois donnée par le plus petit r tel que $b_r \neq 0$.

Soit K un corps valué non archimédien algébriquement clos tel que $v(K^*) \supset \mathcal{Q}$. Considérons un polynôme $f(\mathbf{z}) = \sum_{\mathbf{i} \in I} a_{\mathbf{i}} \mathbf{z}^{\mathbf{i}}$, où $\mathbf{i} = (i_1, \dots, i_n)$ est un multi-indice, $\mathbf{z} = (z_1, \dots, z_n) \in (K^*)^n$, et $a_{\mathbf{i}} \in K^*$. Notons $V(f)$ l'hypersurface définie par f dans $(K^*)^n$. L'amibe $\mathcal{A}(V(f))$ de $V(f)$ peut être décrite de la façon suivante. Soit $\nu : I \rightarrow \mathcal{R}$ la fonction définie par $\nu(\mathbf{i}) = v(a_{\mathbf{i}})$, et soit $T(f)$ l'hypersurface tropicale associée à (I, ν) .

THÉORÈME 2.4 (M. Kapranov, [12]). — *L'adhérence de l'amibe $\mathcal{A}(V(f)) \subset \mathcal{R}^2$ coïncide avec l'hypersurface tropicale $T(f)$. Si la valuation $v : K^* \rightarrow \mathcal{R}$ est surjective, alors $\mathcal{A}(V(f))$ coïncide avec $T(f)$.*

REMARQUE 2.5. — Le théorème de Kapranov montre en particulier que, contrairement au cas complexe, (l'adhérence de) l'amibe d'une hypersurface dans $(K^*)^n$ ne dépend que des valuations des coefficients d'un polynôme définissant l'hypersurface.

Un autre corollaire du théorème de Kapranov est le fait que les squelettes des amibes d'hypersurfaces complexes peuvent être vues comme des amibes d'hypersurfaces sur un corps valué non archimédien.

Les amibes non archimédiennes sont des limites dans la métrique de Hausdorff de certaines familles d'amibes de courbes complexes. Soient I une collection finie de points entiers dans \mathcal{R}^2 , et $\nu : I \rightarrow \mathcal{R}$ une fonction. Une famille de polynômes complexes $f_t(z, w) = \sum_{(i,j) \in I} a_{i,j} t^{\nu(i,j)} z^i w^j$ à deux variables s'appelle une *famille de patchwork* (cf. [37, 38]). Notons V_t la courbe définie par f_t dans $(\mathcal{C}^*)^2$. Une famille de patchwork peut être vue comme un polynôme $f(z, w) = \sum_{(i,j) \in I} a_{i,j} t^{\nu(i,j)} z^i w^j$ sur le corps \mathcal{K} des séries transfinies à coefficients complexes et exposants réels. Notons $V_{\mathcal{K}}$ la courbe définie par f dans $(\mathcal{K}^*)^2$. Pour $t > 1$, considérons l'application $h_t : \mathcal{R}^2 \rightarrow \mathcal{R}^2$ définie par $(x, y) \mapsto (\frac{x}{\log t}, \frac{y}{\log t})$.

PROPOSITION 2.6 (Mikhalkin, [21], Rullgård, [31]). — *L'amibe $\mathcal{A}(V_{\mathcal{K}}) \subset \mathcal{R}^2$ est la limite dans la métrique de Hausdorff des images $h_t(\mathcal{A}(V_t))$ des amibes $\mathcal{A}(V_t)$ quand t tend vers l'infini.*

3. RÉSULTATS ÉNUMÉRATIFS

Les résultats présentés dans cette section sont consacrés au dénombrement de courbes qui passent par des points fixés sur une surface algébrique.

3.1. Courbes complexes nodales

Soit Δ un polygone convexe à sommets entiers dans \mathcal{R}^2 . Notons $\iota(\Delta)$ le nombre de points entiers à l'intérieur de Δ , et $\phi(\Delta)$ le nombre de points sur la frontière de Δ diminué de 1. Considérons le système linéaire $\mathcal{L}(\Delta)$ des courbes dans $(\mathcal{C}^*)^2$ définies par les polynômes de Laurent complexes dont les polygones de Newton sont contenus dans Δ . Remarquons que $\mathcal{L}(\Delta)$ est un espace projectif complexe de dimension $\iota(\Delta) + \phi(\Delta)$, et qu'une courbe générique de $\mathcal{L}(\Delta)$ est non singulière de genre $\iota(\Delta)$.

Notons $\mathcal{L}_\delta(\Delta)$ le sous-ensemble de $\mathcal{L}(\Delta)$ formé par les courbes qui ont exactement δ points doubles non dégénérés et n'ont pas d'autres singularités dans $(\mathcal{C}^*)^2$. La clôture algébrique $\tilde{\mathcal{L}}_\delta(\Delta)$ de $\mathcal{L}_\delta(\Delta)$ dans $\mathcal{L}(\Delta)$ s'appelle une *variété de Severi*. Les courbes de $\mathcal{L}_\delta(\Delta)$ ont le genre $g = \iota(\Delta) - \delta$. (Pour une courbe $C = \cup_{i=1}^s C_i$, où C_i sont les composantes irréductibles de C , on pose le genre de C égal à la somme des genres des composantes C_i diminuée de $s - 1$. En particulier, le genre d'une courbe peut être strictement négatif.)

Posons $N_\Delta(g)$ égal au degré de $\tilde{\mathcal{L}}_\delta(\Delta)$ dans $\mathcal{L}(\Delta)$. Autrement dit, $N_\Delta(g)$ est le nombre de courbes appartenant à $\mathcal{L}(\Delta)$ qui passent par $g + \phi(\Delta)$ points génériques donnés dans $(\mathcal{C}^*)^2$ et qui ont le genre g (la dernière condition peut être remplacée par celle d'avoir $\delta = \iota(\Delta) - g$ points doubles). Puisque $\tilde{\mathcal{L}}_0(\Delta) = \mathcal{L}(\Delta)$, on a $N_\Delta(\iota(\Delta)) = 1$. Notons $N_\Delta^{\text{irr}}(g)$ le nombre de courbes **irréductibles** appartenant à $\mathcal{L}(\Delta)$ qui passent par $g + \phi(\Delta)$ points génériques donnés dans $(\mathcal{C}^*)^2$ et qui ont le genre g .

Pour un entier strictement positif m , considérons le triangle Δ_m à sommets $(0, 0)$, $(m, 0)$ et $(0, m)$ dans \mathcal{R}^2 . Le nombre $N_{\Delta_m}(g)$ (resp., $N_{\Delta_m}^{\text{irr}}(g)$) coïncide avec le nombre de courbes (resp., des courbes irréductibles) de degré m dans $\mathcal{C}P^2$ qui passent par $g + 3m - 1$ points génériques donnés et qui ont le genre g .

Les nombres $N_{\Delta_m}^{\text{irr}}(g)$ déterminent les nombres $N_{\Delta_m}(g)$ et inversement (voir, par exemple, [4]). Les nombres $N_{\Delta_m}^{\text{irr}}(g)$ sont les invariants de Gromov-Witten de $\mathcal{C}P^2$ (voir [15]). Une formule récursive pour calculer les nombres $N_{\Delta_m}^{\text{irr}}(0)$ a été démontrée par M. Kontsevich (voir [15]). L. Caporaso et J. Harris [4] ont donné un algorithme pour calculer les nombres $N_{\Delta_m}(g)$ pour un genre g arbitraire.

G. Mikhalkin a proposé une nouvelle formule pour $N_{\Delta_m}(g)$ qui se généralise immédiatement à une formule pour $N_\Delta(g)$, où Δ est un polygone convexe arbitraire à sommets entiers. Le théorème de Mikhalkin est basé sur une reformulation du problème énumératif présenté ci-dessus en un problème énumératif correspondant en géométrie tropicale.

3.2. Analogues tropicaux

Ce paragraphe contient un "dictionnaire" qui permet de retrouver des analogues tropicaux de certaines notions de la géométrie algébrique complexe, et en particulier des notions introduites dans 3.1 (un dictionnaire plus complet pourra être trouvé dans [23]).

L'analogue tropical du polygone de Newton d'une courbe algébrique a été déjà introduit dans 2.3 : c'est le polygone de Newton d'une courbe tropicale. Nous allons nous intéresser

aux analogues tropicaux des courbes irréductibles, des courbes nodales, des courbes non singulières, du genre d'une courbe, et des points génériques.

Une courbe tropicale T est dite *irréductible* si elle ne peut pas être représentée comme une réunion de courbes tropicales toutes différentes de T . Une courbe tropicale T est dite *élémentaire* si elle peut être paramétrée par une immersion $\theta : \Lambda \rightarrow T$ telle que tous les sommets de Λ aient la puissance 3, chaque point de T ait au plus deux images réciproques, et l'image de chaque sommet de Λ ait une seule image réciproque. Une courbe tropicale T ayant le polygone de Newton Δ est élémentaire si et seulement si la subdivision $\mathcal{D}_\Delta(T)$ de Δ (voir 2.3) vérifie les conditions suivantes :

- tout polygone de $\mathcal{D}_\Delta(T)$ est soit un triangle, soit un parallélogramme,
- tout point entier sur la frontière de Δ est un sommet de $\mathcal{D}_\Delta(T)$.

Une telle subdivision $\mathcal{D}_\Delta(T)$ est aussi dite *élémentaire*. Les courbes tropicales élémentaires sont des analogues tropicaux des courbes complexes nodales.

Une courbe tropicale T de polygone de Newton Δ est dite *non singulière* si la subdivision $\mathcal{D}_\Delta(T)$ de Δ est une triangulation unimodulaire, *i.e.*, si $\mathcal{D}_\Delta(T)$ est formée par des triangles de l'aire 1 (on normalise l'aire de telle façon que l'aire d'un triangle dont les seuls points entiers sont ses sommets soit égale à 1).

Soit T une courbe tropicale élémentaire ayant le polygone de Newton Δ . Le *rang* de T est la différence, diminuée de 1, entre le nombre de sommets de $\mathcal{D}_\Delta(T)$ et le nombre de parallélogrammes dans $\mathcal{D}_\Delta(T)$. Le rang d'une courbe tropicale élémentaire est un analogue de la dimension de la variété de Severi à laquelle appartient une courbe complexe nodale donnée.

Le *genre* d'une courbe tropicale paramétrée $\theta : \Lambda \rightarrow T$ est $1 - \dim H_0(\Lambda) + \dim H_1(\Lambda)$. Le *genre* $g(T)$ d'une courbe tropicale $T \subset \mathcal{R}^2$ est le genre minimal parmi toutes les paramétrisations de T . Si T est élémentaire, alors le genre de T est égal à la différence entre le nombre de sommets de $\mathcal{D}_\Delta(T)$ qui se trouvent à l'intérieur de Δ et le nombre de parallélogrammes dans $\mathcal{D}_\Delta(T)$. Cette observation est un analogue tropical de la formule de Riemann-Roch. On renvoie à [23] pour la formule de Riemann-Roch tropicale dans le cas de courbes tropicales dans \mathcal{R}^n .

Introduisons encore une caractéristique d'une courbe tropicale élémentaire T . La multiplicité $\zeta(T)$ (et la multiplicité $\zeta(\mathcal{D}_\Delta(T))$ de $\mathcal{D}_\Delta(T)$) est le produit des aires de tous les triangles dans $\mathcal{D}_\Delta(T)$.

On dit que des points x_1, \dots, x_k sont *tropicalement génériques* si toute courbe tropicale $T \subset \mathcal{R}^2$ passant par x_1, \dots, x_k et telle que $g(T) + \phi(\Delta(T)) \geq k$ (où $\Delta(T)$ est un polygone de Newton de T) vérifie les conditions suivantes : T est élémentaire, et aucun des points x_1, \dots, x_k ne coïncide avec un sommet de T .

Soient k un entier strictement positif, et Π un ensemble de k points tropicalement génériques dans \mathcal{R}^2 . Considérons la collection $\mathcal{C}_\Delta(\Pi)$ des courbes tropicales élémentaires qui ont le polygone de Newton Δ , passent par tous les points de Π , et ont le rang k

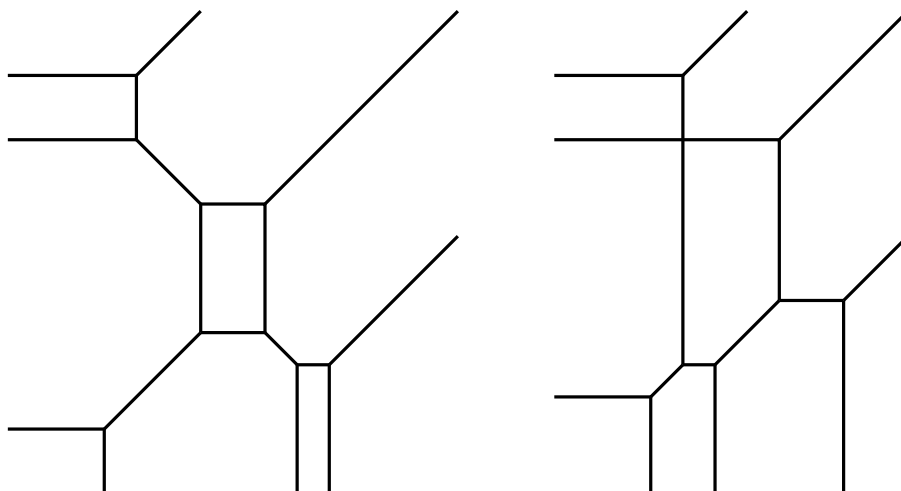


FIGURE 6. Une cubique tropicale non singulière et une cubique tropicale rationnelle

(la dernière condition peut être remplacée par celle d'avoir le genre $k - \phi(\Delta)$). Notons $\mathcal{T}_\Delta(\Pi)$ le nombre de courbes dans $\mathcal{C}_\Delta(\Pi)$ comptées avec les multiplicités. *A priori*, le nombre $\mathcal{T}_\Delta(\Pi)$ dépend de la collection Π choisie. Le Théorème 3.1 du paragraphe suivant montre que $\mathcal{T}_\Delta(\Pi_1) = \mathcal{T}_\Delta(\Pi_2)$ pour des collections quelconques Π_1 et Π_2 de k points tropicalement génériques, et que le nombre $\mathcal{T}_\Delta(g) = \mathcal{T}_\Delta(\Pi)$ est l'analogie tropicale de $N_\Delta(g)$, où $g = k - \phi(\Delta)$.

3.3. Théorème de correspondance

Soient, comme dans le paragraphe précédent, $\Delta \subset \mathcal{R}^2$ un polygone convexe à sommets entiers, k un entier strictement positif, et Π un ensemble de k points tropicalement génériques dans \mathcal{R}^2 . Un des résultats centraux présentés dans ce texte est le théorème suivant de Mikhalkin.

THÉORÈME 3.1 (G. Mikhalkin, [23]). — *Les nombres $\mathcal{T}_\Delta(\Pi)$ et $N_\Delta(g)$, où $g = k - \phi(\Delta)$, sont égaux.*

REMARQUE 3.2. — La démonstration de Mikhalkin du Théorème 3.1 fournit plus que l'égalité des deux nombres : elle établit une bijection entre $\mathcal{C}_\Delta(\Pi)$ (vu comme multi-ensemble) et l'ensemble des courbes complexes de genre g qui appartiennent à $\mathcal{L}(\Delta)$ et passent par une certaine collection de k points génériques dans $(\mathcal{C}^*)^2$. Une autre approche établissant une bijection entre ces deux ensembles a été proposée par E. Shustin [33].

3.4. Courbes réelles et invariant de Welschinger

Soient g un entier positif, et Ω une collection de $g + \phi(\Delta)$ points génériques dans $(\mathcal{R}^*)^2$. Posons $N_\Delta^{\mathcal{R}}(g, \Omega)$ égal au nombre de courbes **réelles** (*i.e.*, invariantes par rapport à la conjugaison complexe) de genre g dans $(\mathcal{C}^*)^2$ qui ont Δ pour polygone de Newton et

passent par tous les points de Ω . De façon similaire, posons $N_{\Delta}^{\mathcal{R},\text{irr}}(g, \Omega)$ égal au nombre de courbes **réelles irréductibles** de genre g dans $(\mathcal{C}^*)^2$ qui ont Δ pour polygone de Newton et passent par tous les points de Ω .

Contrairement au cas complexe, $N_{\Delta}^{\mathcal{R}}(g, \Omega)$ et $N_{\Delta}^{\mathcal{R},\text{irr}}(g, \Omega)$ dépendent de la collection Ω choisie. Une question très intéressante et largement ouverte est celle de bonnes bornes inférieures et supérieures pour $N_{\Delta}^{\mathcal{R}}(g, \Omega)$ et $N_{\Delta}^{\mathcal{R},\text{irr}}(g, \Omega)$. On renvoie à [34] pour plus d'information sur des questions de géométrie énumérative réelle.

Nous allons nous intéresser maintenant au cas $\Delta = \Delta_m$. Introduisons deux autres nombres associés à une collection Ω de $g + 3m - 1$ points génériques dans $\mathcal{R}P^2$. Soit $N_{\Delta_m}^{+,\text{irr}}(g, \Omega)$ (respectivement, $N_{\Delta_m}^{-,\text{irr}}(g, \Omega)$) le nombre de courbes réelles irréductibles de degré m et de genre g dans $\mathcal{C}P^2$ qui passent par tous les points de Ω et ont un nombre pair (respectivement, impair) de points doubles isolés (un point double est dit *isolé* s'il est localement défini par $x^2 + y^2 = 0$). Le *nombre de Welschinger* $W_m(g, \Omega)$ est la différence $N_{\Delta_m}^{+,\text{irr}}(g, \Omega) - N_{\Delta_m}^{-,\text{irr}}(g, \Omega)$.

THÉORÈME 3.3 (J.-Y. Welschinger, [41, 42]). — *Si $g = 0$, le nombre $W_m(g, \Omega)$ ne dépend pas du choix de la collection générique Ω .*

REMARQUE 3.4. — En fait, le Théorème 3.3 est un cas très particulier du théorème de Welschinger qui est formulé dans un cadre symplectique. On renvoie à [41, 42] pour l'énoncé général et la démonstration.

Le nombre $W_m(0, \Omega)$, où Ω est une collection de $3m - 1$ points génériques dans $\mathcal{R}P^2$, s'appelle un *invariant de Welschinger* et est noté W_m . La valeur absolue de l'invariant de Welschinger W_m fournit une borne inférieure pour le nombre de courbes rationnelles réelles de degré m passant par $3m - 1$ points génériques dans $\mathcal{R}P^2$ (les valeurs de W_m pour des petits m peuvent être trouvées dans le paragraphe 3.5).

Le théorème de Welschinger ne se généralise pas directement au cas $g > 0$: pour tout entier $m \geq 4$, il existe des collections génériques Ω_1 et Ω_2 de $3m$ points dans $\mathcal{R}P^2$ telles que $W_m(1, \Omega_1) \neq W_m(1, \Omega_2)$ (voir [11]).

La bijection mentionnée dans la Remarque 3.2 permet aussi d'exprimer l'invariant de Welschinger W_m en termes de courbes tropicales.

Soit T une courbe tropicale élémentaire ayant le polygone de Newton Δ . La courbe T et la subdivision correspondante $\mathcal{D}_{\Delta}(T)$ sont dites *impaires* si chaque triangle dans $\mathcal{D}_{\Delta}(T)$ a une aire impaire. La courbe T et la subdivision $\mathcal{D}_{\Delta}(T)$ sont dites *positives* (respectivement, *négatives*) si la somme des nombres de points entiers intérieurs dans tous les triangles de $\mathcal{D}_{\Delta}(T)$ est paire (respectivement, impaire). Notons $\mathcal{T}_{\Delta}^{+,\text{irr}}(\Pi)$ (respectivement, $\mathcal{T}_{\Delta}^{-,\text{irr}}(\Pi)$) le nombre de courbes impaires positives (respectivement, négatives) irréductibles dans $\mathcal{C}_{\Delta}(\Pi)$ (comptées **sans** multiplicités).

THÉORÈME 3.5 (cf. [22, 33]). — *L'invariant de Welschinger W_m est égal à la différence des nombres $\mathcal{T}_{\Delta_m}^{+, \text{irr}}(\Pi)$ et $\mathcal{T}_{\Delta_m}^{-, \text{irr}}(\Pi)$.*

Les Théorèmes 3.1 et 3.5 expriment les nombres $N_{\Delta}(g)$ et W_m en termes des nombres de certaines courbes tropicales, mais ils ne donnent pas de recettes pour calculer ces derniers nombres. Mikhalkin [22] a proposé un algorithme combinatoire permettant de calculer les nombres en question. Cet algorithme est présenté dans le paragraphe suivant.

3.5. Algorithme combinatoire

Soit Δ un polygone convexe à sommets entiers dans \mathcal{R}^2 . Choisissons une fonction linéaire $\lambda : \mathcal{R}^2 \rightarrow \mathcal{R}$ telle que sa restriction à \mathcal{Z}^2 soit injective. Notons p (respectivement, q) le point où λ atteint son minimum (respectivement, son maximum) sur Δ . Les sommets p et q divisent la frontière $\partial\Delta$ de Δ en deux parties $\partial\Delta_+$ et $\partial\Delta_-$.

Soit l un entier strictement positif. Un chemin $\gamma : [0, l] \rightarrow \Delta$ est dit λ -admissible si

- $\gamma(0) = p$ et $\gamma(l) = q$,
- la composition $\lambda \circ \gamma$ est injective,
- pour tout entier $0 \leq i \leq l-1$, le point $\gamma(i)$ est entier, et $\gamma([i, i+1])$ est un segment.

Le nombre l s'appelle la *longueur* de γ . Un chemin λ -admissible γ divise Δ en deux polygones : le polygone $\Delta_+(\gamma)$ borné par γ et $\partial\Delta_+$, et le polygone $\Delta_-(\gamma)$ borné par γ et $\partial\Delta_-$. (L'intérieur de $\Delta_{\pm}(\gamma)$ n'est pas forcément connexe et peut être vide.)

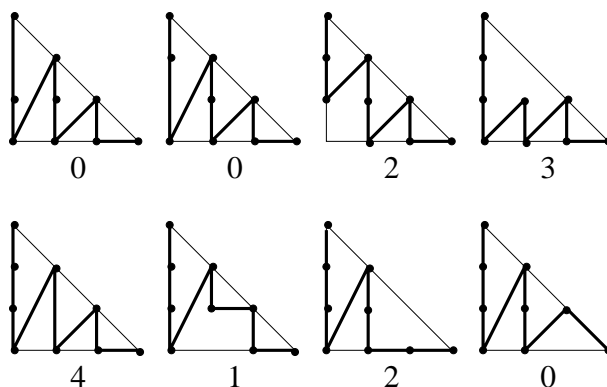
Définissons l'opération de *compression* de $\Delta_+(\gamma)$ de la façon suivante. Notons j le plus petit entier entre 1 et $l-1$ tel que $\gamma(j)$ soit un sommet de $\Delta_+(\gamma)$ avec un angle strictement inférieur à π . (Si un tel entier j n'existe pas, une compression de $\Delta_+(\gamma)$ n'est pas définie.) Une compression de $\Delta_+(\gamma)$ est le polygone $\Delta_+(\gamma')$, où γ' est ou bien le chemin défini par

$$\gamma'(i) = \gamma(i) \text{ pour } i < j, \text{ et } \gamma'(i) = \gamma(i+1) \text{ pour } i \geq j,$$

ou bien le chemin défini par

$$\begin{aligned} \gamma'(i) &= \gamma(i) \text{ pour } i \neq j, \text{ et } \gamma'(j) = \gamma(j-1) + \gamma(j+1) - \gamma(j), \\ &\text{(à condition que } \gamma(j-1) + \gamma(j+1) - \gamma(j) \in \Delta). \end{aligned}$$

Remarquons que γ' est aussi un chemin λ -admissible. Une suite de compressions qui commence par $\Delta_+(\gamma)$ et se termine par un polygone ayant l'intérieur vide (*i.e.*, un polygone $\Delta_+(\bar{\gamma})$ tel que l'image de $\bar{\gamma}$ soit entièrement contenue dans $\partial\Delta_+$) définit une subdivision de $\Delta_+(\gamma)$ qui s'appelle une *subdivision de compression*. Une compression et une subdivision de compression de $\Delta_-(\gamma)$ sont définies de manière complètement similaire. Une paire $(\mathcal{S}_+(\gamma), \mathcal{S}_-(\gamma))$, où $\mathcal{S}_{\pm}(\gamma)$ est une subdivision de compression de $\Delta_{\pm}(\gamma)$, produit une subdivision de Δ . Notons $\mathcal{N}_{\lambda}(\gamma)$ la collection des subdivisions élémentaires de Δ qui peuvent être obtenues de cette façon à partir de γ .

FIGURE 7. Calcul du nombre $N_{\Delta_3}(0)$

THÉORÈME 3.6 (G. Mikhalkin, [22, 23]). — Soit $1 \leq k \leq \iota(\Delta) + \phi(\Delta)$ un entier. Alors, il existe une collection Π de k points tropicalement génériques dans \mathcal{R}^2 telle que l'application \mathcal{D}_Δ , qui associe à une courbe tropicale élémentaire T de polygone de Newton Δ la subdivision correspondante $\mathcal{D}_\Delta(T)$, établisse une bijection entre $\mathcal{C}_\Delta(\Pi)$ et la réunion disjointe $\coprod_\gamma \mathcal{N}_\lambda(\gamma)$, où γ parcourt tous les chemins λ -admissibles de longueur k dans Δ . En particulier, $\mathcal{T}_\Delta(\Pi) = \sum_\gamma \sum_{\mathcal{S} \in \mathcal{N}_\lambda(\gamma)} \zeta(\mathcal{S})$, où $\zeta(\mathcal{S})$ est la multiplicité de \mathcal{S} .

COROLLAIRE 3.7 (G. Mikhalkin, [22, 23]). — Dans les notations du Théorème 3.6, on a $N_\Delta(g) = \sum_\gamma \sum_{\mathcal{S} \in \mathcal{N}_\lambda(\gamma)} \zeta(\mathcal{S})$, où $g = k - \phi(\Delta)$,

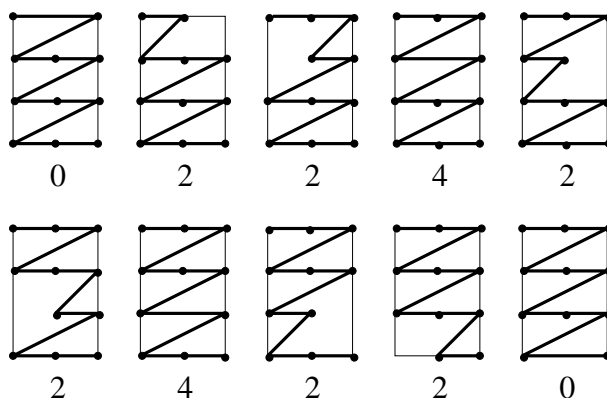
La Figure 7 montre un calcul du nombre $N_{\Delta_3}(0)$. C'est le nombre de cubiques rationnelles passant par 8 points génériques dans $\mathcal{C}P^2$.

La Figure 8 montre un calcul du nombre $N_{\Delta_{2,3}}(1)$, où $\Delta_{2,3}$ est le rectangle à sommets $(0,0)$, $(2,0)$, $(2,3)$ et $(0,3)$ dans \mathcal{R}^2 . C'est le nombre de courbes de genre 1 et de bi-degré $(2,3)$ passant par 10 points génériques dans $\mathcal{C}P^1 \times \mathcal{C}P^1$.

Le Théorème 3.6 a aussi une version qui, pour une collection Π de k points tropicalement génériques choisis de façon appropriée (voir le paragraphe 4.3 pour une description de telles collections) permet de compter le nombre de courbes tropicales dans $\mathcal{C}_\Delta(\Pi)$ (vu comme multi-ensemble) correspondant aux courbes réelles via la bijection mentionnée dans la Remarque 3.2. On renvoie à [22, 23] pour les détails.

Pour une collection Π ayant les propriétés décrites dans le Théorème 3.6, notons $\mathcal{N}_\lambda^+(\gamma)$ (respectivement, $\mathcal{N}_\lambda^-(\gamma)$) le nombre de subdivisions impaires positives (respectivement, négatives) dans $\mathcal{N}_\lambda(\gamma)$ qui sont les images de courbes tropicales irréductibles par la bijection $\mathcal{D}_\Delta|_{\mathcal{C}_\Delta(\Pi)} : \mathcal{C}_\Delta(\Pi) \rightarrow \coprod_\gamma \mathcal{N}_\lambda(\gamma)$. L'énoncé suivant est un corollaire immédiat des Théorèmes 3.5 et 3.6.

PROPOSITION 3.8. — L'invariant de Welschinger W_m est égal à $\sum_\gamma (\mathcal{N}_\lambda^+(\gamma) - \mathcal{N}_\lambda^-(\gamma))$, où γ parcourt tous les chemins λ -admissibles de longueur $3m - 1$ dans Δ_m .

FIGURE 8. Calcul du nombre $N_{\Delta_{2,3}}(1)$

La Proposition 3.8 permet de calculer W_m pour des petites valeurs de m . Par exemple, en observant la Figure 7, on conclut presque immédiatement que $W_3 = 8$ (les nombres W_m pour $m \leq 3$ sont faciles à calculer par d'autres moyens : on a $W_1 = W_2 = 1$, et on peut obtenir l'égalité $W_3 = 8$ en calculant la caractéristique d'Euler-Poincaré de $\mathcal{R}P^2$ à l'aide d'un pinceau de cubiques passant par 9 points réels ; voir [5]). Par contre, les valeurs $W_4 = 240$ et $W_5 = 18264$ ont été trouvées en utilisant la Proposition 3.8.

La Proposition 3.8 permet aussi d'obtenir une borne inférieure pour W_m , et, en particulier, de démontrer que $\lim_{m \rightarrow +\infty} W_m = +\infty$ et $W_m > 0$ pour tout entier $m \geq 1$ (voir [11]). Par conséquent, pour tout entier $m \geq 1$ et toute collection Ω de $3m - 1$ points dans $\mathcal{R}P^2$, il existe au moins une courbe rationnelle réelle de degré m passant par les points de Ω .

4. IDÉES DE DÉMONSTRATIONS

Dans ce paragraphe, nous allons présenter les grandes lignes des démonstrations des Théorèmes 3.1 et 3.6. Les détails pourront être trouvés dans [23].

4.1. Déformation de la structure complexe et courbes J_∞ -holomorphes

Fixons un polygone convexe $\Delta \subset \mathcal{R}^2$ à sommets entiers et un entier $0 \leq g \leq \iota(\Delta)$. Posons $k = g + \phi(\Delta)$, et choisissons des points génériques z_1, \dots, z_k dans $(\mathcal{C}^*)^2$ tels que les points $x_1 = \text{Log}(z_1), \dots, x_k = \text{Log}(z_k)$ soient tropicalement génériques. Rappelons que $N_\Delta(g)$ est le nombre de courbes de genre g dans $(\mathcal{C}^*)^2$ ayant le polygone de Newton Δ et passant par z_1, \dots, z_k .

Pour tout nombre réel $t > 0$, considérons le difféomorphisme $H_t : (\mathcal{C}^*)^2 \rightarrow (\mathcal{C}^*)^2$ défini par

$$H_t : (z, w) \mapsto (|z|^{(1-t)/t} z, |w|^{(1-t)/t} w).$$

Ce difféomorphisme induit une nouvelle structure complexe J_t sur $(\mathcal{C}^*)^2$.

Une courbe C_t est holomorphe par rapport à J_t si et seulement si $C_t = H_t(C)$, où C est une courbe holomorphe par rapport à la structure complexe standard. Si C_t est une courbe holomorphe par rapport à J_t , le polygone de Newton de la courbe $H_t^{-1}(C_t)$ est aussi appelé le polygone de Newton de C_t . Si t est suffisamment générique, le nombre de courbes J_t -holomorphes de genre g ayant le polygone de Newton Δ et passant par z_1, \dots, z_k est égal à $N_\Delta(g)$. Remarquons que l'amibe d'une courbe J_t -holomorphe C_t est l'image de l'amibe de $H_t^{-1}(C_t)$ par l'homothétie $(x, y) \mapsto (x/t, y/t)$.

Il n'y a pas de limite (dans le sens habituel) des structures complexes J_t quand t tend vers l'infini. Néanmoins, on peut définir certaines courbes qui sont des limites de courbes J_t -holomorphes, $t \rightarrow \infty$.

Définissons l'application $v^c : \mathcal{K}^* \rightarrow \mathcal{C}^*$ par $v^c(b(t)) = e^{v(b(t))+i \arg(b_m)}$, où \mathcal{K} est le corps des séries transfinies à coefficients complexes et exposants réels, et b_m est le coefficient du monôme ayant l'exposant $v(b(t))$ dans $b(t)$. Soient $W : (\mathcal{K}^*)^2 \rightarrow (\mathcal{C}^*)^2$ l'application définie par $W(b_1(t), b_2(t)) = (v^c(b_1(t)), v^c(b_2(t)))$, et $C_{\mathcal{K}} \subset (\mathcal{K}^*)^2$ une courbe algébrique.

DÉFINITION 4.1. — *L'image $W(C_{\mathcal{K}}) \subset (\mathcal{C}^*)^2$ de $C_{\mathcal{K}}$ par W s'appelle une courbe J_∞ -holomorphe.*

Le polygone de Newton de $W(C_{\mathcal{K}})$ est, par définition, le polygone de Newton de la courbe $C_{\mathcal{K}}$.

Soit $C_{\mathcal{K}} \subset (\mathcal{K}^*)^2$ une courbe définie par un polynôme

$$F(b_1(t), b_2(t)) = \sum_{(i,j) \in I} a_{i,j} (b_1(t))^i (b_2(t))^j, \quad a_{i,j} \in \mathcal{K}^*$$

de polygone de Newton Δ . D'après le théorème de Kapranov (voir le Théorème 2.4), $\text{Log}(W(C_{\mathcal{K}}))$ est la courbe tropicale associée à la paire (I, ν) , où $\nu : I \rightarrow \mathcal{R}$ est la fonction définie par $\nu(i, j) = v(a_{i,j})$.

De façon géométrique, les courbes J_∞ -holomorphes peuvent être décrites comme les complexes polyédraux de dimension 2 dans $(\mathcal{C}^*)^2$ qui se projettent par Log sur des courbes tropicales et qui ont les propriétés locales formulées si-dessous.

Soient $T \subset \mathcal{R}^2$ une courbe tropicale, $x \in T$ un point, et U un voisinage convexe de x dans \mathcal{R}^2 tel que $T \cap U$ soit un cône sur x . Considérons la cellule de T (vue comme un graphe rectiligne) dont l'intérieur relatif contient x , et notons Δ' l'élément dual de cette cellule dans la subdivision $\mathcal{D}_\Delta(T)$ de Δ . Notons $\text{Ver}(\Delta')$ l'ensemble des sommets de Δ' .

On dit qu'un complexe polyédral $C_\infty \subset (\mathcal{C}^*)^2$ de dimension 2 est $(T \cap U)$ -compatible, si $\text{Log}^{-1}(U) \cap C_\infty = \text{Log}^{-1}(U) \cap \mathcal{W}$, où \mathcal{W} est une translation de $W(C_{\mathcal{K}})$, et $C_{\mathcal{K}} \subset (\mathcal{K}^*)^2$ est la courbe définie par un polynôme $\sum_{(i,j) \in \text{Ver}(\Delta')} a_{i,j} (b_1(t))^i (b_2(t))^j$, dont les valuations de tous les coefficients $a_{i,j}$ sont égales à 0.

THÉORÈME 4.2 (Mikhalkin, [23]). — *Soit $C_\infty \subset (\mathcal{C}^*)^2$ un complexe polyédral de dimension 2. Les conditions suivantes sont équivalentes :*

- C_∞ est une courbe J_∞ -holomorphe,
- $T = \text{Log}(C_\infty)$ est une courbe tropicale et, pour tout point $x \in T$, il existe un voisinage ouvert convexe U de x tel que C_∞ soit $(T \cap U)$ -compatible,
- il existe une suite $\{C_{t_s}\}_{s \in \mathcal{N}}$ de courbes J_{t_s} -holomorphes telle que $\lim_{s \rightarrow \infty} t_s = \infty$ et C_∞ soit la limite dans la métrique de Hausdorff des courbes C_{t_s} .

4.2. Correspondance entre courbes J_t -holomorphes et courbes J_∞ -holomorphes

Soit C_∞ une courbe J_∞ -holomorphe. Comme dans le cas des courbes tropicales, on peut localement “normaliser” C_∞ en considérant une *paramétrisation* de C_∞ par une surface \tilde{C}_∞ (i.e., une application $\psi : \tilde{C}_\infty \rightarrow C_\infty$ ayant certaines propriétés locales ; on renvoie à [23] pour la définition), et définir le genre de C_∞ comme le genre minimal de \tilde{C}_∞ parmi toutes les paramétrisations $\tilde{C}_\infty \rightarrow C_\infty$. (Le genre de \tilde{C}_∞ est égal à $1 + \sum_C (g(C) - 1)$, où C parcourt toutes les composantes connexes de \tilde{C}_∞ , et $g(C)$ est le genre de C .) Pour un nombre réel $\varepsilon > 0$, notons $TN_\varepsilon(C_\infty)$ le ε -voisinage de C_∞ dans $(\mathcal{C}^*)^2$.

Notons \mathcal{H}_t (respectivement, \mathcal{H}_∞) la collection des courbes J_t -holomorphes (respectivement, des courbes J_∞ -holomorphes) de genre g qui ont le polygone de Newton Δ et passent par les points z_1, \dots, z_k .

En utilisant les caractérisations des courbes J_∞ -holomorphes données dans le Théorème 4.2, on peut démontrer le théorème suivant qui établit, pour t suffisamment grand, une bijection entre les collections \mathcal{H}_t et \mathcal{H}_∞ .

THÉORÈME 4.3 (Mikhalkin, [23]). — *Soit $C_\infty \in \mathcal{H}_\infty$. Alors, pour tout $\varepsilon > 0$ suffisamment petit, il existe un nombre réel $t_0 > 0$ tel que, pour tout $t > t_0$, la collection \mathcal{H}_t ait une et une seule courbe entièrement contenue dans $TN_\varepsilon(C_\infty)$. L’application $\mathcal{H}_\infty \rightarrow \mathcal{H}_t$ définie de cette façon (pour tout t suffisamment grand) est une bijection.*

Le Théorème 3.1 est un corollaire du Théorème 4.3 et des résultats ci-dessous.

LEMME 4.4 (Mikhalkin, [23]). —

- (1) Si $C_\infty \subset (\mathcal{C}^*)^2$ est une courbe J_∞ -holomorphe et $T = \text{Log}(C_\infty)$, alors le genre de T est inférieur ou égal au genre de C_∞ .
- (2) Toute courbe tropicale de genre $\leq g$ ayant le polygone de Newton Δ et passant par x_1, \dots, x_k est élémentaire et a le genre g .

THÉORÈME 4.5 (Mikhalkin, [23]). — *Le nombre de courbes J_∞ -holomorphes dans \mathcal{H}_∞ qui se projettent par Log sur la même courbe tropicale élémentaire T est égal à la multiplicité $\zeta(T)$ de T .*

4.3. Idée de la démonstration du Théorème 3.6

Pour démontrer le Théorème 3.6, Mikhalkin choisit une collection très particulière de k points x_1, \dots, x_k dans \mathcal{R}^2 . Soit L une droite orthogonale aux lignes de niveau de la

fonction $\lambda : \mathcal{R}^2 \rightarrow \mathcal{R}$. Remarquons que L a une pente irrationnelle. Les points x_1, \dots, x_k sont choisis sur L un par un : x_1 est un point arbitraire sur L , et si les points x_1, \dots, x_j ($1 \leq j \leq k-1$) sont déjà choisis, on choisit un point x_{j+1} sur L de telle façon que $\lambda(x_{j+1}) \gg \lambda(x_i)$ pour tout entier $1 \leq i \leq j$. Remarquons que les points x_1, \dots, x_k sont tropicalement génériques ! Le dénombrement des courbes tropicales élémentaires de genre g ayant Δ comme polygone de Newton et passant par les points x_1, \dots, x_k choisis mène facilement à l'algorithme présenté dans le paragraphe 3.5 (voir [23] pour les détails).

BIBLIOGRAPHIE

- [1] V.I. ARNOLD – *Mathematical methods of classical mechanics*. “Nauka”, Moscou, 1974 (en russe); traduction anglaise : Graduate Texts in Mathematics, 60. Springer-Verlag, New York, 1989.
- [2] M.F. ATIYAH – *Angular momentum, convex polyhedra and algebraic geometry*, Proc. Edinburgh Math. Soc. **26** (1983) 121–133.
- [3] G.M. BERGMAN – *The logarithmic limit set of an algebraic variety*, Trans. AMS **157** (1971), 459–469.
- [4] L. CAPORASO et J. HARRIS – *Counting plane curves of any genus*, Invent. Math. **131** (1998), 345–392.
- [5] A. DEGTYAREV et V. KHARLAMOV – *Topological properties of real algebraic varieties: Rokhlin’s way*, Russ. Math. Surveys **55** (2000), no. 4, 735–814.
- [6] I.M. GELFAND, M.M. KAPRANOV et A.V. ZELEVINSKY – *Discriminants, resultants and multidimensional determinants*. Birkhäuser, Boston, 1994.
- [7] M. FORSBERG, M. PASSARE et A. TSIKH – *Laurent determinants and arrangements of hyperplane amoebas*, Advances in Math. **151** (2000), 45–70.
- [8] W. FULTON – *Introduction to Toric Varieties*. Annals of Mathematics Studies 131, Princeton University Press, Princeton, 1993.
- [9] A. HENRIQUES – *An analogue of convexity for complements of amoebas of varieties of higher codimensions*. Prépublication, Berkeley, 2001.
- [10] D. HILBERT – *Mathematische Probleme*, Arch. Math. Phys. (3) **1** (1901), 213–237.
- [11] I. ITENBERG, V. KHARLAMOV et E. SHUSTIN – *Welschinger invariant and enumeration of real rational curves*, (à paraître dans Intern. Math. Res. Notices).
- [12] M.M. KAPRANOV – *Amoebas over non-Archimedean fields*. Prépublication, 2000.
- [13] V. KHARLAMOV et S. OREVKOV – *Asymptotic growth of the number of classes of real plane algebraic curves as the degree grows*, Zapiski Nauchn. Semin. POMI **266** (2000), 218–233 (en russe); traduction anglaise : J. of Math. Sciences **113** (2003), 666–674.

- [14] V. KHARLAMOV et S. OREVKOV – *The number of trees half of whose vertices are leaves and asymptotic enumeration of plane real algebraic curves*. Prépublication arXiv: math.AG/0301245.
- [15] M. KONTSEVICH et Yu. MANIN – *Gromov-Witten classes, quantum cohomology and enumerative geometry*, Commun. Math. Phys. **164** (1994), 525–562.
- [16] M. KONTSEVICH et Ya. SOIBELMAN – *Homological mirror symmetry and torus fibrations*. Prépublication arXiv: math.SG/0011041.
- [17] G.L. LITVINOV et V.P. MASLOV – *The correspondence principle for Idempotent Calculus and some computer applications*. Dans : Idempotency, J. Gunawardena (Editor), Cambridge University Press, Cambridge, 1998, pp. 420–443.
- [18] G.L. LITVINOV, V.P. MASLOV et A.N. SOBOLEVSKII – *Idempotent Mathematics and Interval Analysis*. Prépublication arXiv: math.SC/9911126.
- [19] G. MIKHALKIN – *Real algebraic curves, moment map and amoebas*, Ann. of Math. **151** (2000), 309–326.
- [20] G. MIKHALKIN – *Amoebas of algebraic varieties*. Prépublication arXiv: math.AG/0108225.
- [21] G. MIKHALKIN – *Decomposition into pairs-of-pants for complex algebraic hypersurfaces*. Prépublication arXiv: math.GT/0205011.
- [22] G. MIKHALKIN – *Counting curves via lattice paths in polygons*, C. R. Acad. Sci. Paris, Sér. I, **336** (2003), 629–634.
- [23] G. MIKHALKIN – *Counting holomorphic curves in toric surfaces and tropical algebraic geometry*, (en préparation).
- [24] G. MIKHALKIN et H. RULLGÅRD – *Amoebas of maximal area*, Intern. Math. Res. Notices **9** (2001), 441–451.
- [25] G. MIKHALKIN et O. VIRO – *Amoebas of algebraic varieties*, (en préparation).
- [26] J.-E. PIN – *Tropical semirings*. Dans : Idempotency (Bristol 1994), pp. 50–69. Publ. Newton Inst. 11, Cambridge Univ. Press, Cambridge, 1998.
- [27] M. PASSARE et H. RULLGÅRD – *Amoebas, Monge-Ampère measures, and triangulations of the Newton polytope*. Prépublication, Université de Stockholm, 2000.
- [28] J.-J. RISLER – *Construction d'hypersurfaces réelles [d'après Viro]*. Sémin. Bourbaki 1992/93, exp. n° 763, Astérisque **216** (1993), 69–86.
- [29] L. RONKIN – *On zeroes of almost periodic functions generated by holomorphic functions in a multicircular domain*. Dans : Complex Analysis in Modern Mathematics, Fazis, Moscou, 2000, 243–256.
- [30] H. RULLGÅRD – *Stratification des espaces de polynômes de Laurent et la structure de leurs amibes*, C. R. Acad. Sci. Paris, Série I, **331** (2000), 355–358.
- [31] H. RULLGÅRD – *Polynomial amoebas and convexity*. Prépublication, Université de Stockholm, 2001.

- [32] F. SEVERI – *Vorlesungen über algebraische Geometrie*. Teubner, 1921 (resp. Johnson, 1968).
- [33] E. SHUSTIN – *Patchworking singular algebraic curves, non-Archimedean amoebas and enumerative geometry*. Prépublication arXiv: math.AG/0211278.
- [34] F. SOTTILE – *Enumerative Real Algebraic Geometry*. Prépublication arXiv: math.AG/0107179.
- [35] B. STURMFELS – *On the Newton polytope of the resultant*, J. of Alg. Combinatorics **3** (1994), 207–236.
- [36] B. STURMFELS – *Solving systems of polynomial equations*. CBMS Regional Conference Series in Mathematics. AMS, Providence, RI 2002.
- [37] O. VIRO – *Gluing of algebraic hypersurfaces, smoothing of singularities and construction of curves*. Dans : Proc. Leningrad Int. Topological Conf., Leningrad, 1982, Nauka, Leningrad, 1983, pp. 149–197 (en russe).
- [38] O. VIRO – *Gluing of plane real algebraic curves and construction of curves of degrees 6 and 7*. Dans : Lect. Notes Math. **1060**, Springer, Berlin etc., 1984, pp. 187–200.
- [39] O. VIRO – *Progress in the topology of real algebraic varieties over the last six years*, Rus. Math. Surv. **41** (1986), no. 3, 55–82.
- [40] O. VIRO – *Dequantization of real algebraic geometry on a logarithmic paper*, Proceedings of the European Congress of Mathematicians (2000).
- [41] J.-Y. WELSCHINGER – *Invariants of real rational symplectic 4-manifolds and lower bounds in real enumerative geometry*, C. R. Acad. Sci. Paris, Sér. I, **336** (2003), 341–344.
- [42] J.-Y. WELSCHINGER – *Invariants of real symplectic 4-manifolds and lower bounds in real enumerative geometry*, Prépublication arXiv: math.AG/0303145.
- [43] G. WILSON – *Hilbert’s sixteenth problem*, Topology **17** (1978), no. 1, 53–73.
- [44] R. WALKER – *Algebraic curves*. Princeton Univ. Press, Princeton, N.J., 1950.

Ilia ITENBERG

IRMAR

Campus de Beaulieu

F-35042 RENNES cedex

à partir du 1 septembre 2003 :

UFR de Mathématiques et d’Informatique

Université Louis Pasteur

7, rue René Descartes

F-67084 Strasbourg cedex

E-mail : itenberg@math.u-strasbg.fr

SUR LA THÉORIE ÉLÉMENTAIRE DES GROUPES LIBRES [d'après Sela]

par **Frédéric PAULIN**

Étant donné un groupe G , on appelle *théorie élémentaire de G* l'ensemble $\mathbf{T}(G)$ des formules closes du langage des groupes (en logique du premier ordre) qui sont vérifiées dans G , et *théorie universelle de G* l'ensemble $\mathbf{T}_\forall(G)$ de ces formules en forme préfixe dont les quantificateurs sont tous \forall (voir le chapitre 1). Deux groupes G, G' sont dits *élémentairement équivalents* si $\mathbf{T}(G) = \mathbf{T}(G')$ et *universellement équivalents* si $\mathbf{T}_\forall(G) = \mathbf{T}_\forall(G')$.

Dans une série de travaux peu lisibles, Sela [Sel1]-[Sel6] annonce¹ des résultats profonds sur la théorie élémentaire des groupes libres, répondant à des questions soulevées par Tarski vers 1945 (voir [Tar] dans un contexte plus large et [LS, page 51]). En attendant les arbitrages en cours, nous les énonçons encore sous forme de conjecture.

CONJECTURE 1. — *Deux groupes libres de type fini, non cycliques, sont élémentairement équivalents.*

Sela annonce la liste exacte des groupes de type fini, qui sont élémentairement équivalents à un groupe libre de type fini, non cyclique : ce sont les *tours hyperboliques* (voir le chapitre 6.1), construites par «recolléments» successifs de groupes libres et de groupes fondamentaux de surface. On a en particulier l'énoncé frappant suivant.

CONJECTURE 2. — *Le groupe fondamental d'une surface compacte connexe de caractéristique d'Euler au plus -2 est élémentairement équivalent à un groupe libre de type fini, non cyclique.*

Le problème de Tarski de savoir quels sont les groupes élémentairement équivalents à un groupe donné, se pose pour d'autres beaux groupes, comme $SL_n(\mathbf{Z})$ et les réseaux des groupes de Lie connexes. Mais en tant qu'objets universels, les groupes libres sont fondamentaux et prioritaires. De plus, leur «flexibilité» fait que le problème est beaucoup plus difficile à résoudre pour eux que pour des groupes plus «rigides». En utilisant les

¹Le premier résultat ci-dessus, ainsi que la décidabilité de la théorie élémentaire d'un groupe libre de type fini non cyclique, avaient été annoncés en 1998 par Kharlampovich-Myasnikov [KM0]. Avec le recul de cinq ans, il semble clair que les auteurs n'avaient pas de preuve écrite complète en 1998. Voir [KM1]-[KM5] pour l'état de leurs travaux.

mêmes techniques que pour les groupes libres, Sela [Sel7] annonce aussi la caractérisation des groupes de type fini qui sont élémentairement équivalents à un groupe sans torsion hyperbolique (au sens de Gromov, voir [Ghy1]).

Cet article est consacré aux résultats structurels sur les *groupes limites*, i.e. les groupes de type fini universellement équivalents à un groupe libre de type fini. Nous avons suivi de près l'approche topologique de Champetier-Guirardel [CG].

Dans le chapitre 3, nous donnons diverses caractérisations simples, dues à Remeslenikov [Rem], Kharlampovich-Myasnikov [KM1, KM2], Sela [Sel1], Champetier-Guirardel [CG], des groupes limites (ce sont aussi les groupes « multi-résiduellement libres » de type fini, les sous-groupes de type fini d'un ultraproduct de groupe libre, les groupes « limites de groupes libres ») ainsi que leurs premières propriétés.

Dans le chapitre 5, nous donnons les théorèmes principaux de [KM2, Sel1] sur la structure des groupes limites. En particulier (voir [Sel1]), en prenant pour racine un groupe limite G donné, on construit naturellement un arbre fini enraciné de quotients stricts successifs, de sommets terminaux des groupes libres, appelé *diagramme de Makanin-Razborov*, qui permet de « décrire » l'ensemble de tous les morphismes de G dans un groupe libre.

Les outils principaux sont ceux des actions de groupes sur les arbres, au sens de [Ser], avec les contributions fondamentales de Rips-Sela [RS] sur les actions à stabilisateurs d'arêtes cycliques, que nous rappelons dans le chapitre 4. Nous en donnons des versions relatives, utiles pour gérer les contraintes sur les paramètres.

Donnons une justification à l'intérêt de l'étude des groupes limites pour résoudre le problème de Tarski pour les groupes libres (outre le fait qu'un groupe de type fini élémentairement équivalent à un groupe libre de type fini est bien sûr un groupe limite). La structure des formules (quantification sur des variables de combinaisons booléennes d'équations de groupes en ces variables) montre l'intérêt de l'étude des équations dans les groupes. De nombreux travaux portent sur ce sujet, dont ceux fondateurs de Lyndon [Lyn1, Lyn2]. Les plus frappants jusqu'ici étaient ceux de Makanin [Mak1, Mak2], qui montrent qu'il existe un algorithme pour décider si un système fini d'équations dans un groupe libre admet une solution, et que les théories universelles et positives d'un groupe libre de type fini sont décidables, et de ceux de Razborov [Raz], qui décrivent la structure de l'ensemble de toutes les solutions d'un système fini d'équations dans un groupe libre. Le lien avec le chapitre 5 est le suivant. Considérons un système fini d'équations $\underline{m}(\underline{p}_1, \dots, \underline{p}_n, \underline{x}) = 1$ avec \underline{x} un uplet d'inconnues, $\underline{p}_1, \dots, \underline{p}_n$ des uplets de paramètres (nécessaires pour le traitement de la récurrence sur le nombre d'alternances de quantificateurs), et \underline{m} un uplet de mots ès lettres de $\underline{x}, \underline{x}^{-1}, \underline{p}_i, \underline{p}_i^{-1}$. Les solutions de ce système dans un groupe libre F , lorsque les paramètres varient dans F soumis à certaines contraintes, sont en bijection avec les morphismes de groupes, à valeurs dans F , du groupe G de présentation $\langle \underline{p}_1, \dots, \underline{p}_n, \underline{x} \mid \underline{m}(\underline{p}_1, \dots, \underline{p}_n, \underline{x}) = 1 \rangle$, marqué par la suite génératrice (correspondant à) $(\underline{p}_1, \dots, \underline{p}_n, \underline{x})$, avec contraintes sur les images des \underline{p}_i . Enfin, comprendre

les morphismes d'un groupe de type fini dans un groupe libre se ramène à comprendre les morphismes d'un nombre fini de groupes limites dans les groupes libres (voir paragraphe 5.2). Mis à part Lyndon et Remeslennikov sous une forme différente, c'est sans doute Rips qui a eu le premier l'idée que l'on pourrait utiliser les techniques d'actions de groupes sur les arbres (pas forcément simpliciaux), en particulier à stabilisateurs d'arête cycliques, pour résoudre le problème de Tarski.

Nous concluons cette introduction en donnant un aperçu du schéma de la preuve de Sela de l'énoncé 1. Notons que Bestvina-Feighn annoncent d'autres preuves des résultats des articles [Sel3, Sel4].

Le cœur des travaux de Sela [Sel2]-[Sel5], dont nous ne parlerons malheureusement pas ici, et qui est sans doute incontournable pour une solution du problème de Tarski, est un procédé d'élimination des quantificateurs. Soient $\mathbf{L}_n, \mathbf{L}_m$ deux groupes libres de rang $n, m \geq 2$, dont on souhaite montrer qu'ils sont élémentairement équivalents, et a_1, a_2 les deux premiers éléments communs d'une partie génératrice libre de $\mathbf{L}_n, \mathbf{L}_m$. Par un résultat classique (voir [Sac] disant que \mathbf{L}_n et \mathbf{L}_m ont la même théorie universelle-existentielle, il suffit, par récurrence, de montrer que toute partie d'une puissance de \mathbf{L}_n , définissable par une formule $\exists \forall \exists$, est une combinaison booléenne de parties définissables par des formules $\forall \exists$, et ceci de manière indépendante de n . C'est ce que fait Sela dans [Sel5]. Avec un traitement nécessaire d'une partition en deux des variables libres, il s'agit de comprendre les parties des puissances de \mathbf{L}_n définissables par une formule $\forall \exists$. Lorsque celle-ci est close et positive, disons $\forall \underline{y} \exists \underline{z} \underline{m}(\underline{y}, \underline{z}) = 1$, Merzliakov [Mer], dans sa preuve de l'égalité des théories positives de \mathbf{L}_n et \mathbf{L}_m , introduit un procédé de « solutions formelles » : par des méthodes de « petite simplification » (comme dans [Sac] d'ailleurs), Merzliakov montre qu'il existe un uplet de mots $\underline{z}(\underline{y}, a_1, a_2)$ en \underline{y}, a_i^\pm , tels que la formule $\forall \underline{y} \exists \underline{z} \underline{m}(\underline{y}, \underline{z}) = 1$ est satisfaite dans $\mathbf{L}_n, \mathbf{L}_m$ si et seulement si $\forall \underline{y} \underline{m}(\underline{y}, \underline{z}(\underline{y}, a_1, a_2)) = 1$ l'est. Sela étend ces solutions formelles au cas où la formule $\forall \exists$ n'est plus positive [Sel2] (voir aussi [KM3]) et possède des variables libres [Sel5]. C'est la partie la plus technique, qui amène Sela à introduire de nouvelles variables, ce qui l'oblige dans [Sel4] à un travail important pour assurer la finitude de ses procédés d'élimination de quantificateurs.

Remerciements : Cet article doit beaucoup aux travaux en préparation [CG, Gui] de Christophe Champetier et Vincent Guirardel, et je les remercie pour leurs nombreuses conversations, suggestions et corrections. Je remercie aussi G. Baumslag, F. Dahmani, T. Delzant, O. Kharlampovich, G. Levitt, A. Myasnikov, P. Papasoglu, Z. Sela.

1. THÉORIE ÉLÉMENTAIRE DES GROUPES

1.1. Formules

Dans ce texte, nous appellerons *formule* toute expression logique de la forme

$$Q_1 x_1 Q_2 x_2 \dots Q_{n'} x_{n'} \bigvee_{i=1}^k \left(\bigwedge_{j=1}^{k_i} m_{ij}(x_1, \dots, x_n) = 1 \right) \wedge \left(\bigwedge_{j=1}^{k'_i} m'_{ij}(x_1, \dots, x_n) \neq 1 \right)$$

avec

- n, n', k, k_i, k'_i dans \mathbf{N} tels que $n' \leq n$,
- Q_i l'un des quantificateurs \forall, \exists ,
- \vee, \wedge les connecteurs logiques «ou», «et»,
- x_i une variable, avec $x_1, \dots, x_{n'}$ les *variables liées* et $x_{n'+1}, \dots, x_n$ les *variables libres*,
- m_{ij}, m'_{ij} des mots réduits dans les lettres x_1, \dots, x_n et leurs inverses $x_1^{-1}, \dots, x_n^{-1}$.

Toute formule en ce sens est une formule du langage des groupes et toute formule du langage des groupes est équivalente, modulo la théorie des groupes, à une formule en ce sens (voir par exemple [CK]).

Une formule est *sans quantificateur* si $n' = 0$, *close* si $n' = n$, *universelle* si Q_i est \forall pour tout i , *existentielle* si Q_i est \exists pour tout i , *positive* si $k'_i = 0$ pour $i = 1, \dots, k$. Une *formule* $\forall\exists$ est une formule comme ci-dessus telle qu'il existe n'' dans \mathbf{N} avec $0 \leq n'' \leq n'$ tel que Q_i est \forall pour $1 \leq i \leq n''$ et Q_i est \exists pour $n'' < i \leq n'$. On définit de même une *formule* $\exists\forall\exists$.

Si G est un groupe, on appelle *théorie élémentaire* (respectivement *universelle*, *existentielle*, *positive*, $\forall\exists$, $\exists\forall\exists$) de G l'ensemble des formules closes (respectivement closes universelles, closes existentielles, closes positives, closes $\forall\exists$, closes $\exists\forall\exists$) vérifiées dans G , et on la note $\mathbf{T}(G)$ (respectivement $\mathbf{T}_\forall(G)$, $\mathbf{T}_\exists(G)$, $\mathbf{T}_+(G)$, $\mathbf{T}_{\forall\exists}(G)$, $\mathbf{T}_{\exists\forall\exists}(G)$).

Un *plongement élémentaire* $f : G \rightarrow G'$ est un morphisme injectif de groupes tels que, pour toute formule $\varphi(\underline{x})$, de k -uplet de variables libres \underline{x} , et pour tout k -uplet \underline{a} dans G , l'expression logique $\varphi(\underline{a})$ est vérifiée dans G si et seulement si $\varphi(f(\underline{a}))$ est vérifiée dans G' . Dans [Sel6], Sela annonce (voir aussi [KM5]) que le plongement standard $\mathbf{L}_n \rightarrow \mathbf{L}_m$ pour $n \leq m$ est un plongement élémentaire, ce qui, par restriction aux formules closes, implique la validité de la conjecture 1.

REMARQUES. (1) Il est facile de voir que $\mathbf{T}(\mathbf{Z}^n) = \mathbf{T}(\mathbf{Z}^m)$ si et seulement si $n = m$, et plus généralement qu'un groupe de type fini est élémentairement équivalent à \mathbf{Z}^n si et seulement s'il lui est isomorphe.

(2) Deux groupes libres de type fini non cycliques sont universellement équivalents (i.e. $\mathbf{T}_\forall(\mathbf{L}_n) = \mathbf{T}_\forall(\mathbf{L}_m)$ pour tous $n, m \geq 2$), car ils se plongent l'un dans l'autre, et toute formule close universelle vérifiée dans un groupe est aussi vérifiée dans un sous-groupe.

(3) Un résultat classique (voir [Sac]) dit que $\mathbf{T}_{\forall\exists}(\mathbf{L}_n) = \mathbf{T}_{\forall\exists}(\mathbf{L}_m)$ pour tous $n, m \geq 2$. Merzlyakov [Mer] a démontré que $\mathbf{T}_+(\mathbf{L}_n) = \mathbf{T}_+(\mathbf{L}_m)$ pour tous $n, m \geq 2$.

1.2. Des propriétés élémentaires des groupes libres

Pour illustrer la notion de formule, voici quelques propriétés vérifiées par un groupe G élémentairement équivalent à un groupe libre non cyclique F , ainsi qu'une famille de formules closes vérifiées dans F , dont la validité dans G entraîne ces propriétés.

(1) G est non abélien

$$\exists xy \ [x, y] \neq 1 .$$

(2) G est sans torsion

$$\{ \forall x \ (x = 1) \vee (x^n \neq 1) \}_{n \in \mathbb{N} - \{0\}} .$$

(3) G est *commutatif-transitif*

$$\forall xyz \ y = 1 \vee [x, y] \neq 1 \vee [y, z] \neq 1 \vee [x, z] = 1 .$$

(4) Tout sous-groupe abélien maximal A de G est *malnormal* (i.e. si $gAg^{-1} \cap A$ est non trivial, alors g appartient à A)

$$\{ \forall xyz \ y = 1 \vee [x, y] \neq 1 \vee [y, z] \neq 1 \vee [x, z] = 1 ,$$

$$\forall xy \ x = 1 \vee y = 1 \vee [x, yxy^{-1}] \neq 1 \vee [x, y] = 1 \} .$$

(5) Tout sous-groupe abélien libre de G est cyclique

$$\forall xyz \ \exists u \ [x, y] \neq 1 \vee [y, z] \neq 1 \vee xyu^2 = 1 \vee xzu^2 = 1 \vee yzu^2 = 1 .$$

Les formules closes (2) (3) (4) sont des formules universelles, la première est une formule existentielle, la dernière une formule $\forall\exists$. Remarquons que si G est commutatif-transitif, alors le centralisateur de tout élément non trivial est abélien, donc tout sous-groupe abélien non trivial est contenu dans un unique sous-groupe abélien maximal.

Par contre, l'expression logique suivante (tout sous-groupe abélien de G est cyclique), n'est pas du premier ordre, car on y quantifie sur les parties (ou sur les entiers).

$$\forall P \subset G, (\forall xy \in P, xy^{-1} \in P \wedge [x, y] = 1) \Rightarrow (\exists x \in G, \forall y \in P, \exists n \in \mathbb{N}, y = x^n) .$$

Cette propriété est vérifiée dans un groupe de type fini élémentairement équivalent à un groupe libre (voir corollaire 5.4). Mais il n'existe pas d'ensemble de formules du premier ordre la définissant, comme me l'a fait remarquer F. Point.

1.3. Ultrafiltres

Soit I un ensemble non vide. Un *filtre* sur I est une partie non vide ω de $\mathbf{P}(I)$ telle que

(1) $\emptyset \notin \omega$,

(2) si X, Y appartiennent à ω , alors $X \cap Y$ aussi,

(3) si X appartient à ω et si X est contenue dans Y , alors Y appartient à ω .

Un filtre ω est un *ultrafiltre* s'il est maximal (pour l'inclusion). Un ultrafiltre est *non principal* s'il est différent de $\{J \subset I / a \in J\}$ pour tout a dans I . Par le théorème de Zorn, tout ensemble infini admet un ultrafiltre non principal. Tout ultrafiltre non principal sur un ensemble infini contient le filtre de Fréchet des complémentaires des parties finies.

REMARQUE. Les deux propriétés suivantes seront utilisées par la suite (voir [Bou]).

(1) Si ω est un ultrafiltre et J est une partie de I , alors ou bien J appartient à ω , ou bien son complémentaire $I - J$ appartient à ω . L'application $\mathbf{P}(I) \rightarrow \{0, 1\}$, définie par $J \mapsto 1$ si et seulement si J appartient à ω , est donc une mesure finiment additive, que l'on note encore ω .

(2) Si X est un espace topologique, et $\{x_i\}_{i \in I}$ une famille relativement compacte dans X , alors il existe un unique point x_ω dans l'adhérence de cette famille, tel que, pour tout voisinage V de x_ω , le point x_i appartienne à V pour ω -presque tout i . On note $\lim_\omega x_i$ ce point, qui vérifie les propriétés usuelles des limites.

Si X est un ensemble et ω un ultrafiltre sur I , on note ${}^\omega X$ (et parfois $*X$ lorsque ω est sous-entendu), et on appelle *ultraproduit* de X (pour ω), le quotient X^I / \sim avec \sim la relation d'équivalence définie par $(x_i)_{i \in I} \sim (y_i)_{i \in I}$ si et seulement si $x_i = y_i$ pour ω -presque tout i . On note $[x_i]_\omega$ l'image dans ${}^\omega X$ d'une famille $(x_i)_{i \in I}$. Si G est un groupe, la loi de groupe produit sur G^I passe au quotient en une loi de groupe, dont on munit l'ultraproduit ${}^\omega G$. On définit de même l'ultraproduit d'un anneau ou d'un corps.

Terminons ce chapitre en citant deux théorèmes qui justifient l'utilité des ultrafiltres pour le problème de trouver les groupes élémentairement équivalents à un groupe libre (voir [CK] pour une preuve).

THÉORÈME 1.1. — (1) (Shelah) *Deux groupes G et H sont élémentairement équivalents si et seulement s'il existe un ensemble I muni d'un ultrafiltre ω tel que les groupes ${}^\omega G$ et ${}^\omega H$ sont isomorphes.*

(2) (Łoś) *L'application $G \rightarrow {}^\omega G$ définie par $g \mapsto [(g)_{i \in I}]_\omega$ est un plongement élémentaire.*

▷

2. GROUPES MARQUÉS

Le contenu de cette partie, prérequis pour caractériser les groupes universellement équivalents à un groupe libre comme «limites de groupes libres», est repris de l'article de Champetier-Guirardel [CG] (voir aussi [Ghy2], où l'espace **GM** est évoqué).

On note \mathbf{L} le groupe libre sur \mathbf{N} , identifié à l'ensemble des mots (finis) réduits dans $\mathbf{N} \cup \mathbf{N}^{-1}$. Pour éviter toute confusion, on note \bar{s}_j l'élément de \mathbf{L} correspondant à l'élément j de \mathbf{N} . Pour n dans \mathbf{N} , on note \mathbf{L}_n le sous-groupe de \mathbf{L} engendré par $\bar{s}_1, \dots, \bar{s}_n$.

On définit la catégorie des groupes marqués de la manière suivante. Un *marquage* d'un groupe G est un morphisme de groupes surjectif $\mathbf{L} \rightarrow G$ tel que $\bar{s}_j \mapsto 1$ pour tout j assez grand. Le groupe trivial admet un unique marquage. Un *groupe marqué* est un groupe G muni d'un marquage. Un *morphisme* de groupes marqués est un morphisme de groupes commutant avec les marquages. Remarquons que, dans cette catégorie, il existe au plus un morphisme entre deux objets, et que tout morphisme est un épimorphisme.

Un marquage $\mathbf{L} \rightarrow G$ étant uniquement déterminé par la donnée d'une famille ordonnée $S = \{s_n\}_{n \in \mathbf{N}}$, engendrant G , avec $s_n = 1$ pour n assez grand (par $\bar{s}_j \mapsto s_j$), on notera (G, S) le groupe marqué correspondant, et par abus $S = (s_1, \dots, s_k)$, étant entendu que $s_j = 1$ pour $j > k$. Si $m(\bar{s}_1, \dots, \bar{s}_p)$ est un élément de \mathbf{L} (en sous-entendant que seules les lettres $\bar{s}_1, \dots, \bar{s}_p$ et leurs inverses peuvent apparaître dans ce mot), on notera $m(s_1, \dots, s_p)$ l'élément de G correspondant par le marquage.

Si G est un groupe et S un marquage de G , tout morphisme surjectif de groupes $\varphi : G \rightarrow G'$ induit un morphisme de groupes marqués $\varphi : (G, S) \rightarrow (G', \varphi(S))$.

On note \mathbf{GM} l'ensemble des classes d'isomorphisme de groupes marqués, qui est gradué $\mathbf{GM} = \bigcup_{n \in \mathbf{N}} \mathbf{GM}_n$, avec \mathbf{GM}_k l'ensemble des groupes marqués (G, S) tels que $s_j = 1$ pour $j > k$. Cet espace a été introduit par Grigorchuk [Gri] et Champetier [Cham, CG].

L'ensemble \mathbf{GM} est muni d'une topologie uniforme naturelle, dont une base d'entourages est $\{V_n / n \in \mathbf{N}\}$, avec V_n l'ensemble des couples de groupes marqués $((G, S), (G', S'))$ tels que, pour tout mot $m(\bar{s}_1, \dots, \bar{s}_k)$ dans \mathbf{L} de longueur au plus n , on a $m(s_1, \dots, s_k) = 1$ si et seulement si $m(s'_1, \dots, s'_k) = 1$, avec les notations évidentes.

Cette topologie est métrisable (par $d(x, y) = 2^{-\sup\{n \in \mathbf{N} / (x, y) \in V_n\}}$), totalement discontinue (car la distance d est ultramétrique), et localement compacte (car \mathbf{GM}_k est compact par extraction diagonale, et est ouvert-fermé (car égal à son $\frac{1}{2}$ -voisinage ouvert)).

Remarques. Faisons quelques remarques culturelles sur ces définitions.

1) Un groupe marqué est pointé en son élément neutre (noté 1). Un groupe marqué (G, S) est muni d'une distance invariante par translation à gauche, la *distance des mots*, qui est la distance maximale sur G telle que $d(1, s_j) = 1$ si $s_j \neq 1$. Un groupe marqué (G, S) est muni d'une action du groupe libre \mathbf{L} , composée du marquage de G et de l'action par translation à gauche de G sur lui-même.

Tout ensemble d'espaces métriques pointés munis d'une action d'un groupe fixé admet une topologie naturelle, la topologie de Gromov équivariante (voir [Pau1]), deux éléments sont proches si les boules de grand rayon dans chacun d'eux sont presque isométriques, de manière presque équivariante pour une grande partie finie du groupe.

Il est immédiat de voir que la topologie de \mathbf{GM} coïncide avec la topologie de Gromov équivariante sur \mathbf{GM} .

(2) Un autre système fondamental d'entourages de \mathbf{GM} est $\{V'_n / n \in \mathbf{N}\}$ avec V'_n l'ensemble des couples de groupes marqués (G, G') tels qu'il existe une bijection (donc une

isométrie) de la boule $B_G(1, n)$ (de centre l'élément neutre et de rayon n pour la distance des mots dans G) sur la boule $B_{G'}(1, n)$ rendant le diagramme suivant commutatif :

$$\begin{array}{ccc} & B_{\mathbf{L}}(1, n) & \\ \swarrow & & \searrow \\ B_G(1, n) & \longrightarrow & B_{G'}(1, n) \end{array} .$$

(3) Pour un groupe marqué (G, S) , demander que G soit de présentation finie équivaut à demander que G soit de présentation finie sur la famille génératrice finie S (en enlevant les 1), par un théorème de B. H. Neumann [Bau2, page 52]. Tout groupe marqué (G', S') suffisamment proche d'un groupe marqué (G, S) de présentation finie est un quotient de (G, S) . En effet, si $\{m_i(\bar{s}_1, \dots, \bar{s}_k)\}_{i \in I}$ est une famille finie engendrant normalement le noyau du marquage $\mathbf{L} \rightarrow G$, et si (G', S') est suffisamment proche de (G, S) , alors les relations $m_i(s'_1, \dots, s'_k) = 1$ sont vérifiées par $S' = \{s'_i\}_{i \in \mathbb{N}}$, et donc l'unique application ordonnée $S \rightarrow S'$ induit un morphisme $(G, S) \rightarrow (G', S')$.

(4) De la même manière, le sous-espace des groupes abéliens marqués de \mathbf{GM} est fermé, car la propriété pour un groupe d'être commutatif est équivalente au fait que les éléments d'une partie génératrice donnée du groupe commutent.

(5) À tout système fini d'équations

$$(\mathbf{S}) \quad \left\{ w(\bar{a}, \bar{x}, \bar{p}_1, \dots, \bar{p}_n) = 1 \right.$$

avec \bar{x} un uplet d'inconnues (ou variables liées), \bar{a} un uplet de coefficients (ou constantes), $\bar{p}_1, \dots, \bar{p}_n$ des uplets de paramètres (ou variables libres), et w un uplet de mots réduits dans $\bar{a}^\pm, \bar{x}^\pm, \bar{p}_1^\pm, \dots, \bar{p}_n^\pm$, on associe le groupe $G_{(\cong)} = L(\bar{a}, \bar{x}, \bar{p}_1, \dots, \bar{p}_n) / \mathbf{N}(\{w\})$, où $\mathbf{N}(\{w\})$ est le sous-groupe distingué engendré par les éléments de w . On note $S_{(\cong)}$ le marquage de $G_{(\cong)}$ par l'image (a, x, p_1, \dots, p_n) dans $G_{(\cong)}$ de $(\bar{a}, \bar{x}, \bar{p}_1, \dots, \bar{p}_n)$. Soit G un groupe muni d'un uplet d'éléments a' (autant que dans a) et d'une famille de sous-groupes $(G_i)_{i=1, \dots, n}$. On a une bijection entre d'une part l'ensemble des solutions de (\mathbf{S}) dans G avec coefficients \bar{a} évalués en a' et un choix de valeurs des paramètres \bar{p}_i dans G_i , et d'autre part l'ensemble des morphismes de groupes de $G_{(\cong)}$ dans G qui envoient (en préservant l'ordre) a sur a' et les p_i dans G_i . Cette bijection associe, à un uplet de telles solutions s de (\mathbf{S}) , l'unique tel morphisme de $G_{(\cong)}$ dans G envoyant (en préservant l'ordre) x sur s , a sur a' et les p_i sur les valeurs choisies dans G_i .

3. PREMIÈRES PROPRIÉTÉS ET CARACTÉRISATIONS DES GROUPES LIMITES

Le résultat suivant est une concaténation de résultats de Remeslennikov [Rem], Kharlampovich-Myasnikov [KM1, KM2], Sela [Sel1], Champetier-Guirardel [CG]. Nous avons largement suivi la présentation faite dans cette dernière référence.

Rappelons qu'un groupe G est *multi-résiduellement libre* si pour toute partie finie P de G , il existe un morphisme de G dans un groupe libre, dont la restriction à P est injective, ou de manière équivalente si G est de type fini, si pour tout k dans \mathbf{N} , et tous g_1, \dots, g_k dans $G - \{1\}$, il existe un morphisme $\varphi : G \rightarrow \mathbf{L}_2$ non trivial sur chaque g_j .

THÉORÈME 3.1. — *Soit G un groupe de type fini (respectivement de type fini non abélien). Les assertions suivantes sont équivalentes:*

- (1) *le groupe G , muni de n'importe quel marquage, est limite de groupes libres de type fini (respectivement de type fini non abéliens) marqués ;*
- (2) *il existe un marquage S de G tel que (G, S) est limite de groupes libres de type fini (respectivement de type fini non abéliens) marqués ;*
- (3) *pour tout ultrafiltre ω non principal sur \mathbf{N} , le groupe G se plonge dans ${}^\omega\mathbf{L}_2$;*
- (4) *il existe un ultrafiltre ω non principal sur \mathbf{N} et $n \geq 2$ tel que G se plonge dans ${}^\omega\mathbf{L}_n$;*
- (5) *le groupe G est multi-résiduellement libre ;*
- (6) *le groupe G a la même théorie universelle qu'un groupe libre de type fini (respectivement de type fini non abélien) ;*
- (7) *le groupe G a la même théorie existentielle qu'un groupe libre de type fini (respectivement de type fini non abélien).*

Un groupe vérifiant l'une de ces conditions équivalentes sera appelé un *groupe limite*.

PREUVE. Nous supposons que G est non abélien, en laissant l'autre cas au lecteur.

(1) \Rightarrow (2) : C'est clair.

(2) \Rightarrow (3) : On suppose que $(G, S) = \lim_{i \rightarrow \infty} (G_i, S_i)$ avec G_i un groupe libre de type fini, donc que l'on peut considérer comme contenu dans \mathbf{L}_2 , et $S_i = (s_{1,i}, \dots, s_{k,i})$ et $S = (s_1, \dots, s_k)$. Alors il est facile de voir que l'application $G \rightarrow {}^\omega\mathbf{L}_2$, définie par $g \mapsto [m(s_{1,i}, \dots, s_{j,i})]_\omega$ si $g = m(s_1, \dots, s_j)$ pour $m(\bar{s}_1, \dots, \bar{s}_j)$ un élément de \mathbf{L} , est bien définie, et est un morphisme injectif de groupes.

(3) \Rightarrow (4) : C'est clair.

(4) \Rightarrow (1) : Soit $\rho : G \rightarrow {}^\omega\mathbf{L}_n$ un plongement de G , et $S = (s_1, \dots, s_k)$ un marquage de G . Si $\rho(s_j) = [s_{j,i}]_\omega$, on pose $S_i = (s_{1,i}, \dots, s_{k,i})$, et G_i le sous-groupe (libre) de \mathbf{L}_n engendré par S_i . Alors il est facile de voir que $(G, S) = \lim_\omega (G_i, S_i)$.

(4) \Rightarrow (5) [Rem] : Il est bien connu que le groupe $\Gamma_3 = \ker(SL_2(\mathbf{Z}) \rightarrow SL_2(\mathbf{Z}/3\mathbf{Z}))$ est libre de type fini et non cyclique, donc le groupe \mathbf{L}_n se plonge dans Γ_3 . Si G se plonge dans ${}^\omega\mathbf{L}_n$, alors il se plonge dans ${}^\omega\Gamma_3 = \ker(SL_2({}^\omega\mathbf{Z}) \rightarrow SL_2({}^\omega\mathbf{Z}/3{}^\omega\mathbf{Z}))$, et on identifie G avec son image. Soit $E = \{g_1, \dots, g_k\}$ une partie finie de $G - \{1\}$, montrons qu'il existe un morphisme de G dans le groupe libre Γ_3 (qui se plonge dans \mathbf{L}_2), qui est non trivial sur chaque g_j . On peut supposer que E est génératrice. Soient a_1, \dots, a_p dans ${}^\omega\mathbf{Z}$ les coefficients non nuls des matrices $g_j - id$. Il suffit de montrer qu'il existe un morphisme

d'anneaux $\rho : \mathbf{Z}[a_1, \dots, a_p] \rightarrow \mathbf{Z}$ non nul sur chaque a_j , car alors ρ induira un morphisme de groupes de G dans Γ_3 non trivial sur chaque g_j .

Soient P_1, \dots, P_q des polynômes engendrant l'idéal J de l'anneau (noethérien) des polynômes $\mathbf{Z}[X_1, \dots, X_p]$ tel que $\mathbf{Z}[X_1, \dots, X_p]/J = \mathbf{Z}[a_1, \dots, a_p]$. Posons $a_j = [a_{j,i}]_\omega$. Alors pour ω -presque tout i , on a $P_\ell(a_{1,i}, \dots, a_{p,i}) = 0$, donc le morphisme d'anneaux ρ défini par $a_j \mapsto a_{j,i}$ convient, pour ω -presque tout i .

(5) \Rightarrow (4) : Si $\rho_i : G \rightarrow \mathbf{L}_n$ est un morphisme injectif sur la boule de centre l'élément neutre et de rayon i dans G (pour n'importe quelle métrique des mots), alors le morphisme $G \rightarrow {}^\omega\mathbf{L}_n$ défini par $g \mapsto [\rho_i(g)]_\omega$ est injectif.

(1) \Rightarrow (6) : Si l'assertion (1) est vraie, alors $\mathbf{T}_\vee(\mathbf{L}_2)$ est contenue dans $\mathbf{T}_\vee(G)$, car le fait de vérifier une formule close universelle donnée est clairement une propriété fermée dans \mathbf{GM} (voir la remarque (3) précédente, en utilisant le fait que si $m(\bar{s}_1, \dots, \bar{s}_\ell)$ est un mot ès $\bar{s}_j^{\pm 1}$, et si m_1, \dots, m_ℓ sont des mots ès $\bar{s}_j^{\pm 1}$, alors $m(m_1, \dots, m_\ell)$ est un mot ès $\bar{s}_j^{\pm 1}$).

Si l'assertion (1) est vraie, montrons que \mathbf{L}_2 se plonge dans G , ce qui implique que $\mathbf{T}_\vee(G)$ est contenue dans $\mathbf{T}_\vee(\mathbf{L}_2)$. Soient a, b dans G ne commutant pas, montrons que le sous-groupe $\langle a, b \rangle$ de G engendré par a, b est libre sur a, b . Comme (1) implique (3), le groupe G se plonge dans ${}^\omega\mathbf{L}_2$, et il en est de même de $\langle a, b \rangle$. Comme (4) implique (1), il existe une suite de groupes marqués libres (G_i, S_i) avec $S_i = (a_i, b_i)$ convergeant vers $(\langle a, b \rangle, (a, b))$. Comme une limite de groupes abéliens marqués est encore un groupe abélien marqué, pour i suffisamment grand, a_i, b_i ne commutent pas, donc engendrent un groupe libre sur S_i . Donc par passage à la limite, a et b ne vérifient pas de relation (réduite) non triviale, et $\langle a, b \rangle$ est libre sur (a, b) .

(6) \Rightarrow (7) : C'est clair par négation.

(7) \Rightarrow (1) : Soit $S = (s_1, \dots, s_k)$ un marquage de G . Soit N dans \mathbf{N} . Notons $m_1(\bar{s}_1, \dots, \bar{s}_k), \dots, m_p(\bar{s}_1, \dots, \bar{s}_k)$ les mots réduits de longueur au plus N dans $\{\bar{s}_1, \dots, \bar{s}_k, \bar{s}_1^{-1}, \dots, \bar{s}_k^{-1}\}$. Considérons le système fini Σ_N d'équations et d'inéquations $\bigwedge_{i=1}^p \text{eqn}_i$ en les variables x_1, \dots, x_k , avec eqn_i l'équation $m_i(x_1, \dots, x_k) = 1$ si $m_i(s_1, \dots, s_k) = 1$ dans G , et eqn_i l'inéquation $m_i(x_1, \dots, x_k) \neq 1$ sinon. La formule existentielle $\exists x_1 \dots x_k \Sigma_N$ étant vérifiée dans G , supposé existentiellement équivalent à un groupe libre F , il existe une solution $S_N = (s'_1, \dots, s'_k)$ dans F au système Σ_N . Si F_N est le sous-groupe (qui est libre) de F engendré par S_N , alors il est immédiat que la suite des (F_N, S_N) converge vers (G, S) . Comme G n'est pas abélien, le groupe F_N n'est pas abélien pour N assez grand. \triangleright

Remarquons qu'un groupe existentiellement équivalent à un groupe libre non abélien est non abélien, et qu'un groupe qui, muni d'un marquage, est limite de groupes libres non abéliens marqués, est non abélien.

En utilisant respectivement les assertions (1), (4), (2), (6), (6), (6), et la preuve de l'implication (1) \Rightarrow (6) du théorème précédant, le corollaire suivant est alors immédiat.

- COROLLAIRE 3.2. — (1) *L'ensemble des groupes limites marqués est un fermé de \mathbf{GM} (donc son intersection avec \mathbf{GM}_n est compacte).*
- (2) *Tout sous-groupe de type fini d'un groupe limite est un groupe limite.*
- (3) *Un produit libre (fini) de groupes de type fini est un groupe limite si et seulement si chaque facteur est un groupe limite.*
- (4) *Un groupe limite est sans torsion.*
- (5) *Un groupe limite est commutatif-transitif.*
- (6) *Les sous-groupes abéliens maximaux d'un groupe limite sont malnormaux.*
- (7) *Un sous-groupe d'un groupe limite engendré par deux éléments est libre sur ces deux éléments ou abélien.* ▷

Nous terminons ce chapitre par des exemples explicites de groupes limites. Par une *surface* nous entendrons (sauf mention contraire) une surface réelle compacte connexe à bord (éventuellement vide), de classe C^∞ ou munie d'une structure de CW-complexe. Une surface est *fermée* si son bord est vide. Un *groupe de surface* est un groupe isomorphe au groupe fondamental d'une surface.

PROPOSITION 3.3. — *Tout groupe de surface, sauf celui du plan projectif, de la bouteille de Klein, et de la somme connexe de trois plans projectifs, est un groupe limite.*

PREUVE. Si le bord d'une surface Σ est non vide, alors le groupe fondamental de Σ est libre, donc un groupe limite. Le groupe fondamental du tore est \mathbf{Z}^2 , qui, muni de sa partie génératrice standard, est la limite de la suite de groupes libres marqués $(\mathbf{Z}, (1, n))$.

Si $\Sigma_{2,+}$ est la somme connexe de deux tores, alors une présentation du groupe fondamental de $\Sigma_{2,+}$ est $\langle a, b, x, y \mid [a, b] = [x, y] \rangle$. Notons $x_n = [a, b]^n a [a, b]^{-n}$ et $y_n = [a, b]^n b [a, b]^{-n}$, qui sont deux éléments du groupe F libre sur a, b , et $S_n = (a, b, x_n, y_n)$, qui est une partie génératrice de F . Alors il est facile de voir que la suite de groupes marqués (F, S_n) converge vers $(\pi_1 \Sigma_{2,+}, (a, b, x, y))$. De même, le groupe fondamental de la somme connexe de quatre plans projectifs $\Sigma_{4,-}$, dont une présentation est $\langle a, b, x, y \mid a^2 b^2 = x^2 y^2 \rangle$, marqué par la partie génératrice (a, b, x, y) , est limite de la suite de groupes libres marqués $(F, (a, b, (a^2 b^2)^n a (a^2 b^2)^{-n}, (a^2 b^2)^n b (a^2 b^2)^{-n}))$.

Comme toute surface fermée de caractéristique d'Euler au plus -2 est un revêtement de $\Sigma_{2,+}$ ou de $\Sigma_{4,-}$, que les surfaces fermées de caractéristique d'Euler au moins -1 sont les sommes connexes d'au plus trois plans projectifs, ainsi que la sphère et le tore, et que tout sous-groupe d'un groupe limite est un groupe limite, ceci montre le résultat.

Le groupe fondamental du plan projectif n'est pas un groupe limite, car il n'est pas sans torsion. Le groupe fondamental de la bouteille de Klein, dont une présentation est $\langle a, b \mid a^2 b^{-2} = 1 \rangle$, n'est pas commutatif-transitif (car a et b commutent avec $a^2 = b^2$), donc n'est pas un groupe limite. Il est montré dans [Lyn1] que l'équation $a^2 b^2 c^2 = 1$ dans un groupe libre implique la commutativité de a, b, c , donc le groupe fondamental de la somme

connexe de trois plans projectifs, dont une présentation est $\langle a, b, c \mid a^2b^2c^2 = 1 \rangle$, n'est pas un groupe limite. \triangleright

4. ACTIONS DE GROUPES SUR DES ARBRES À STABILISATEURS D'ARÊTES ABÉLIENS

4.1. Vocabulaire

Soit G un groupe muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$. Une action (toujours supposée simpliciale, sans inversion) du groupe G sur un arbre simplicial est dite *minimale* s'il n'existe pas de sous-arbre invariant propre, et *cofinie* si le graphe quotient est fini. Un sous-groupe d'un groupe agissant sur un arbre simplicial est dit *elliptique* (pour cette action) s'il fixe un sommet. Une action minimale cofinie est dite *cyclique* si le fixateur de chaque arête est infini cyclique, et *relative* si chacun des \check{G}_j est elliptique. Deux actions d'un groupe G sur des arbres simpliciaux T, T' sont dites *isomorphes* s'il existe un isomorphisme d'arbres G -équivariant de T dans T' .

Un automorphisme de G est dit *relatif* s'il agit par une conjugaison sur chaque \check{G}_j . Le groupe des automorphismes relatifs de G agit par précomposition sur les actions relatives de G sur des arbres simpliciaux. Une action relative de G sur un arbre simplicial est dite *invariante par automorphismes* si elle est isomorphe à sa précomposition par tout automorphisme relatif de G .

Une action relative de G sur un arbre T est dite obtenue par *raffinement* d'une action relative de G sur un arbre T' au-dessus d'un ensemble E' de sommets de T' s'il existe une application $f : T \rightarrow T'$ simpliciale et G -équivariante, qui est une bijection de $T - f^{-1}(GE')$ sur $T' - GE'$ (ce qui implique que $f^{-1}(v')$ est connexe pour tout v' dans E').

Un sommet v d'une action cyclique relative de G sur un arbre T est dit *de type surface relatif* s'il existe une surface à bord simplement connexe $\tilde{\Sigma}$, munie de

- une action de G_v propre et libre, de quotient compact et différent de la sphère, du plan projectif, du disque, de l'anneau, du ruban de Möbius, du pantalon et du ruban de Möbius troué,
- une bijection ([RS, DS] demandent seulement une application, mais cela ne change pas le théorème 4.1) G_v -équivariante $e \mapsto f(e)$ entre l'ensemble des arêtes e issues de v et l'ensemble des composantes connexes du bord de $\tilde{\Sigma}$ telle que le fixateur de e dans T coïncide avec le stabilisateur de $f(e)$ dans $\tilde{\Sigma}$, et l'intersection avec G_v de tout conjugué de \check{G}_j préserve une composante connexe du bord de $\tilde{\Sigma}$.

Un sous-groupe H de G est dit *de type surface relatif* s'il existe un sommet v de type surface relatif d'une action cyclique relative de G sur un arbre tel que $H = G_v$.

Un sous-groupe H de G est dit *relativement librement indécomposable* si H est elliptique dans toute action relative de G sur un arbre simplicial à stabilisateurs d'arêtes triviaux.

Lorsque $H = G$, on dit que G est relativement librement indécomposable. On omet le qualificatif de «relatif» lorsque J est vide. Notons que le groupe Z n'est pas librement indécomposable.

Nous renvoyons à [Ser] (voir aussi [Pau2]) pour la correspondance entre actions de groupes sur les arbres et graphes de groupes. Les définitions ci-dessus se transportent en des définitions correspondantes pour les graphes de groupes. Pour rappeler les notations, si un groupe G agit sur un arbre T , nous noterons $G \backslash T$ un graphe de groupes quotient, bien défini à isomorphisme près (voir [Bas]). Si (X, G_\bullet) est un graphe de groupes, nous noterons VX l'ensemble de ses sommets ; EX l'ensemble de ses arêtes ; $o(e)$ le sommet origine, $t(e)$ le sommet terminal et \bar{e} l'arête opposée d'une arête e ; G_v le groupe du sommet v ; $G_e = G_{\bar{e}}$ le groupe et $\rho_e : G_e \rightarrow G_{t(e)}$ le morphisme injectif associé à l'arête e ; et $\pi_1(X, G_\bullet; v_*)$ le groupe fondamental de (X, G_\bullet) de sommet de base v_* , formé, outre l'identité, des éléments non triviaux (dits *en forme réduite*) $g_0 t_{e_1} g_1 t_{e_2} g_2 \dots t_{e_n} g_n$ avec (e_1, \dots, e_n) un chemin d'arêtes dans X d'origine et d'extrémité v_* , g_0 dans G_{v_*} , g_i dans $G_{t(e_i)}$, $g_0 \neq 1$ si $n = 0$, et $g_i \notin G_{e_i}$ si $e_{i+1} = \bar{e}_i$ pour tout $i = 1, \dots, n-1$.

Soit (X, G_\bullet) un graphe de groupes, de groupes d'arête abéliens, et v_* un sommet fixé de X . On définit un sous-groupe $\text{Tw}(X, G_\bullet)$ de $\text{Aut } \pi_1(X, G_\bullet; v_*)$, appelé *groupe des twists de Dehn* de (X, G_\bullet) , de la manière suivante.

Notons $Z_{(X, G_\bullet)}$ le groupe abélien $(\prod_{e \in EX} G_e) / \sim$ où \sim est la relation d'équivalence engendrée par $(c_e)_{e \in EX} \sim (d_e)_{e \in EX}$ si $d_{\bar{e}} = c_e^{-1}$ pour tout e dans EX . On note $[c_e]$ la classe d'équivalence de $(c_e)_{e \in EX}$. Si $g = g_0 t_{e_1} g_1 t_{e_2} g_2 \dots t_{e_n} g_n$ est un élément du sous-groupe $\pi_1(X, G_\bullet; v_*)$ de $\pi_1(X, G_\bullet)$ en forme réduite, alors

$$[c_e] \cdot g = g_0 t_{e_1} \rho_{e_1}(c_{e_1}) g_1 t_{e_2} \rho_{e_2}(c_{e_2}) g_2 \dots t_{e_n} \rho_{e_n}(c_{e_n}) g_n$$

est encore la forme réduite d'un élément de $\pi_1(X, G_\bullet; v_*)$. On vérifie facilement que ceci définit une action de $Z_{(X, G_\bullet)}$ par automorphismes sur $\pi_1(X, G_\bullet; v_*)$. On note alors $\text{Tw}(X, G_\bullet)$ l'image de $Z_{(X, G_\bullet)}$ dans $\text{Aut } \pi_1(X, G_\bullet; v_*)$.

Pour tout sommet v de X de type surface ou tel que G_v est abélien, notons $\text{Aut}(G_v, \partial)$ le sous-groupe du groupe des automorphismes φ de G_v qui induisent sur le sous-groupe $\rho_e(G_e)$, pour toute arête e d'extrémité v , une conjugaison par un certain élément γ_e de G_v , avec $\gamma_e = 1$ si G_v est abélien. Pour $v' \neq v$ et e' toute arête d'extrémité v' , posons $\varphi = \text{id} : G_{v'} \rightarrow G_{v'}$ et $\gamma_{e'} = 1$. D'après [Lev], tout élément de $\text{Aut}(G_v, \partial)$ s'étend en un automorphisme de $\pi_1(X, G_\bullet; v_*)$, en posant, en utilisant les formes réduites,

$$g_0 t_{e_1} g_1 t_{e_2} g_2 \dots t_{e_n} g_n \mapsto \varphi(g_0) \gamma_{\bar{e}_1} t_{e_1} \gamma_{e_1}^{-1} \varphi(g_1) \gamma_{\bar{e}_2} t_{e_2} \gamma_{e_2}^{-1} \varphi(g_2) \dots t_{e_n} \gamma_{e_n}^{-1} \varphi(g_n).$$

Cette extension dépend du choix des γ_e , mais deux choix donnent deux extensions qui diffèrent par un twist de Dehn (en utilisant le fait que, si v est un sommet de type surface, alors le centralisateur dans G_v d'un groupe d'arête $\rho_e(G_e)$ d'extrémité v est réduit à $\rho_e(G_e)$). Voir [Lev] pour des compléments.

4.2. Décomposition de Grushko relative

Soit G un groupe de type fini, muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$. Il est bien connu lorsque J est vide, et la preuve s'adapte facilement si J est non vide (voir [LS]), qu'il existe p dans \mathbb{N} , des sous-groupes de type fini G_1, \dots, G_p de G non triviaux, relativement librement indécomposables dans G , bien définis modulo permutation et conjugaison, et un sous-groupe libre de type fini F de G , bien défini à isomorphisme près, tels que

- (1) tout \check{G}_j est conjugué à un sous-groupe d'un G_i ,
- (2) si G_i est libre, alors G_i contient un conjugué d'un \check{G}_j non trivial,
- (3) le morphisme naturel du produit libre $G_1 * \dots * G_p * F$ dans G est un isomorphisme.

Nous appellerons cet isomorphisme une *décomposition de Grushko relative* de G , et les G_i et F ses *facteurs de Grushko relatifs*. On omet le terme relatif lorsque J est vide.

Une décomposition de Grushko d'un groupe G permet de décrire toutes les décompositions en produit libre de G : si A_1, \dots, A_q sont des sous-groupes de G tels que le morphisme $A_1 * \dots * A_q \rightarrow G$ est un isomorphisme, alors il existe des éléments g_1, \dots, g_p dans G , des sous-groupes libres F_1, \dots, F_q de G , et une partition $\coprod_{j=1}^q I_j$ de $\{1, \dots, p\}$ tels que l'application canonique de $F_j * (*_{i \in I_j} g_i G_i g_i^{-1})$ dans G ait pour image A_j .

Une décomposition de Grushko permet de ramener l'étude des morphismes d'un groupe de type fini G dans un groupe donné H à l'étude des morphismes d'un groupe de type fini librement indécomposable dans H : si $G_1 * \dots * G_p * G_{p+1}$ est une décomposition de Grushko de G (avec G_{p+1} libre), alors, par la propriété universelle des produits libres, l'application

$$\mathrm{Hom}(G, H) \rightarrow \prod_{j=1}^{p+1} \mathrm{Hom}(G_j, H)$$

définie par $\varphi \mapsto (\varphi|_{G_j})_{j=1, \dots, p+1}$ est un isomorphisme de groupes (l'inverse est donné par $(\varphi_j)_{j=1, \dots, p+1} \mapsto \varphi$ avec $\varphi(g_{i_1} \dots g_{i_k}) = \varphi_{i_1}(g_{i_1}) \dots \varphi_{i_k}(g_{i_k})$ pour tous $g_{i_1} \in G_{i_1}, \dots, g_{i_k} \in G_{i_k}$). Nous noterons $\varphi = \varphi_1 * \dots * \varphi_{p+1}$ dans la suite.

Les groupes libres jouent un rôle particulier, car ils admettent (sauf le cas trivial ou infini cyclique) « beaucoup » de décompositions en produit libre de facteurs non triviaux.

Voici un moyen, indiqué par T. Delzant, de fixer un choix de décomposition de Grushko. Si (G, S) est un groupe marqué et φ un automorphisme de G , appelons *dilatation* de φ l'entier

$$\max\{ \|\varphi^\epsilon(s)\|_S / s \in S \cup S^{-1}, \epsilon = \pm \},$$

avec $\|g\|_S$ la longueur minimale d'une écriture de g comme mot dans $S \cup S^{-1}$. Rappelons (voir [LS, page 121]) que pour toute décomposition en produit libre $G = A_1 * \dots * A_p$ de G , il existe au moins un automorphisme φ du groupe libre sur S tel que la suite $\varphi(S)$ admette des sous-suites disjointes S_1, \dots, S_p telles que (l'image dans G de) S_i soit contenue dans A_i et engendre A_i pour $i = 1, \dots, p$. On définit la complexité d'une telle décomposition comme la borne inférieure des dilatations des tels φ . Une décomposition de Grushko de

G est dite S -minimale si sa complexité est minimale parmi toutes les décompositions de Grushko de G . L'ensemble des décompositions de Grushko S -minimales de G est fini.

4.3. Décomposition JSJ cyclique relative

Rips et Sela [RS] (voir aussi [Sel1, Bow, DS, FP]) ont associé à chaque groupe de type fini G , muni d'une famille finie de sous-groupes $(\check{G}_j)_{j \in J}$, une action cyclique de G sur un arbre, qui, pour les actions cycliques relatives, joue le rôle de la décomposition de Grushko relative pour les actions relatives à stabilisateurs d'arête triviaux. Pour des motivations venant de la topologie de petite dimension (voir [JS, Joh]), elle porte, ainsi que son graphe de groupes quotient, le nom de *décomposition JSJ cyclique relative*. Elle permet de décrire toutes les décompositions cycliques relatives de G et, dans les bons cas, est unique à isomorphisme près et invariante par les automorphismes relatifs de G . On omet le terme relatif si J est vide. Ces bons cas incluent celui des groupes hyperboliques ([Bow]), mais voir [For] pour des phénomènes de non unicité.

De la même manière que les groupes libres pour les décompositions en produit libre, les groupes de surface jouent un rôle particulier pour les décompositions cycliques (car ils admettent de « nombreuses » actions cycliques lorsque la caractéristique d'Euler de la surface est grande en valeur absolue). On peut en fait décrire de manière géométrique toutes les actions cycliques d'un groupe de surface de la manière suivante.

Soit Σ une surface (au sens du paragraphe précédant la proposition 3.3). Soit $\tilde{\Sigma} \rightarrow \Sigma$ un revêtement universel de Σ , soit Γ son groupe de revêtement, muni de la famille $(\Gamma_j)_{j \in J}$ des stabilisateurs dans Γ des composantes connexes du bord de $\tilde{\Sigma}$ (finie modulo conjugaison).

Une *courbe décomposante* de Σ est une courbe fermée simple c contenue dans l'intérieur de Σ , localement séparante (i.e. séparant en deux composantes connexes un (tout) voisinage tubulaire), non homotope à zéro, et non parallèle au bord (i.e. aucune adhérence de composante connexe de $\Sigma - c$ n'est un anneau). Un *système de courbes décomposantes* de Σ est une famille finie de courbes décomposantes, deux à deux disjointes et non homotopes. Si $(c_i)_{i \in I}$ est un système de courbes décomposantes, alors on appelle *arbre dual* de $(c_i)_{i \in I}$ le graphe simplicial d'ensemble de sommets l'ensemble des composantes connexes de $\tilde{\Sigma}$ privé des relevés des c_i , d'ensemble d'arêtes les composantes connexes des relevés des c_i , les deux extrémités d'une arête étant les deux sommets dont l'adhérence contient l'arête. Ce graphe est un arbre, muni d'une action induite de Γ . Cette action de Γ est minimale, cofinie, car par exemple $\text{Card } I \leq 3|\chi(S)| + 1$, cyclique et relative. Le *groupe modulaire* de Γ est le sous-groupe du groupe des automorphismes de Γ engendré par les conjugaisons intérieures et les groupes des twists de Dehn de ces telles décompositions cycliques. Son image dans $\text{Out}(\Gamma)$ est plus connue sous le nom de « mapping class group » (relatif), voir par exemple [ZVC].

Un résultat classique (voir [ZVC, HS]) dit que toute action minimale cyclique et relative de Γ est isomorphe à celle sur l'arbre dual d'un système de courbes décomposantes. En

particulier, les seules surfaces Σ telles que Γ n'admet pas d'action minimale cyclique relative non triviale sont la sphère, le plan projectif, le disque, l'anneau, le ruban de Möbius, le pantalon et le ruban de Möbius troué. C'est pour cela que ces surfaces sont exclues dans la définition de sommet de type surface.

THÉORÈME 4.1 (Rips-Sela [RS]). — *Soit G un groupe de présentation finie, sans torsion, muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$, relativement librement indécomposable. Il existe au moins une action cyclique relative de G sur un arbre simplicial T de sommets colorés en blanc ou noir, appelée une décomposition JSJ cyclique relative, telle que:*

- (1) *chaque \check{G}_j fixe un sommet blanc de T ,*
- (2) *tout sommet noir est de type surface relatif dans T ,*
- (3) *le stabilisateur de tout sommet blanc est elliptique dans toute action cyclique relative de G ,*
- (4) *tout sous-groupe de type surface relatif de G fixe un sommet noir dans T ,*
- (5) *pour toute action cyclique relative de G sur un arbre T'' , il existe un raffinement T' de T en des groupes de sommet noir et une application (simpliciale) G -équivariante d'une subdivision de T' dans T'' . ▷*

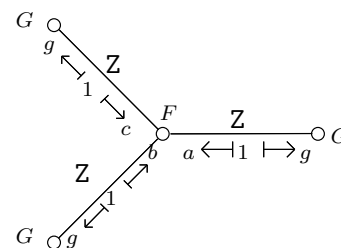
Pour préciser la relation avec les références, l'arbre T défini dans [RS], ainsi que dans [Sel1], vérifie des conditions supplémentaires, non nécessaires pour ce qui suit. L'extension au cas relatif, non traité dans [RS], est immédiat. L'hypothèse de présentation finie est nécessaire [Dun], celle sans torsion facile à contourner (voir par exemple [DS]), et le théorème admet une version en remplaçant cyclique par virtuellement abélien (voir par exemple [DS]). L'une des étapes de la preuve, trop longue pour être rappelée ici, est un théorème de finitude pour la complexité des actions cycliques sur les arbres d'un groupe de présentation finie donné, voir [BF1].

EXEMPLES. 1) Si G est un groupe de surface fermée, alors sa décomposition JSJ cyclique est réduite à un point, car sauf pour la sphère et le plan projectif, G est un sous-groupe de type surface de G . Si G est un groupe abélien libre de type fini, alors sa décomposition JSJ cyclique est réduite à un point, car s'il n'est pas isomorphe à \mathbf{Z}^2 (qui est un groupe de surface fermée), alors toute action cyclique de G sur un arbre admet un point fixe global.

2) Si un groupe G , muni d'une famille de sous-groupes $(G_j)_{j \in J}$, a la propriété (FA) relative de Serre (i.e. toute action relative de G sur un arbre simplicial admet un point fixe), alors sa décomposition JSJ cyclique relative est réduite à un point. Cette hypothèse est par exemple vérifiée si G est un groupe de Kazhdan relatif (i.e. si toute action isométrique affine de G sur un espace de Hilbert, telle que chaque G_j a un point fixe, admet un point fixe global), comme le montre une simple adaptation au cas relatif de la preuve de [HV].

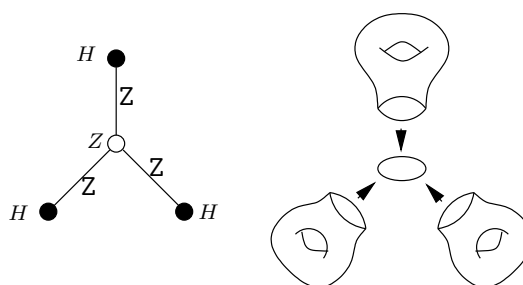
3) Soit $H = SL_3(\mathbf{Z})$, g un élément d'ordre infini de H , et F le groupe de présentation $\langle a, b, c \mid abc = 1 \rangle$, qui est isomorphe au groupe libre \mathbf{L}_2 , et au groupe fondamental du pantalon.

Alors le graphe de groupes ci-contre est une décomposition JSJ cyclique de son groupe fondamental G , mais le sous-groupe F de G est un groupe de sommet blanc (et pas noir) : on a exclu les groupes de pantalon dans la définition de «sommet de type surface relatif», car ils ne contiennent pas de courbe décomposante.



4) Soit H le groupe fondamental du tore troué, et $\{a, b\}$ une partie génératrice libre, de sorte que le groupe fondamental du bord soit engendré par le commutateur $[a, b]$. Soit $Z = \mathbb{Z}$, on considère les morphismes $Z \rightarrow H$ envoyant 1 sur $[a, b]$ et le morphisme identité $Z \rightarrow Z$ (ou plus généralement n'importe quel morphisme injectif).

Le graphe de groupes ci-contre est une décomposition JSJ cyclique de son groupe fondamental G . Pour $Z \rightarrow Z$ l'identité, le graphe de groupes obtenu en écrasant l'une des trois arêtes est aussi une décomposition JSJ cyclique au sens du théorème précédent, non invariante par automorphismes.



Sela obtient un résultat d'unicité faible pour les groupes limites ([Sel1, Theorem 2.7, 3.9]), mais des techniques de Guirardel-Levitt permettent d'obtenir une unicité forte.

PROPOSITION 4.2 (Guirardel-Levitt). — Soit G un groupe de présentation finie, sans torsion, commutatif-transitif, dans lequel tout sous-groupe abélien est abélien libre, et dont les sous-groupes abéliens maximaux sont malnormaux. On suppose que G est muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$, et qu'il est relativement librement indécomposable. Alors G admet une décomposition cyclique relative, notée T_{can} , telle que

- T_{can} est invariante par automorphismes,
- deux sommets de stabilisateurs abéliens ne sont pas reliés par une arête,
- tout stabilisateur de sommet abélien est abélien maximal,
- le centralisateur du stabilisateur d'une arête est contenu dans le groupe de l'un des deux sommets de cette arête,
- T_{can} est une décomposition JSJ cyclique relative, après avoir remplacé chaque groupe de sommet non isolé isomorphe à \mathbb{Z}^2 et dont les groupes d'arête incidents sont contenus dans un facteur infini cyclique, par un lacet de groupes, de groupe de sommet \mathbb{Z} et de groupe d'arête se surjectant sur ce groupe de sommet des deux côtés.

REMARQUE. Nous verrons dans le chapitre suivant (corollaire 5.4) qu'un groupe limite, qui est commutatif-transitif, sans torsion, et dont les sous-groupes abéliens maximaux sont malnormaux par le corollaire 3.2, est de présentation finie, et que tout sous-groupe abélien d'un groupe limite est abélien libre. Ces propriétés sont aussi vérifiées par les groupes hyperboliques sans torsion. Il existe des groupes hyperboliques sans torsion qui ne sont

pas des groupes limites, par exemple ceux (sauf le groupe trivial) qui ont la propriété (T) de Kazhdan, comme les réseaux uniformes de $Sp(n, 1)$.

PREUVE. Notons qu'un sous-groupe de type surface relatif n'est pas abélien (sauf si G est égal à \mathbf{Z}^2 , auquel cas T_{can} réduit à un point convient). Les conditions (3) et (4) du théorème 4.1 impliquent donc que la famille $(H_i)_{i \in V}$ des sous-groupes de G qui sont les groupes de sommet non abéliens d'une décomposition JSJ cyclique relative T , est indépendante de celle-ci. Comme les automorphismes relatifs de G préservent l'ensemble des décompositions JSJ cycliques relatives, cette famille est invariante par automorphismes relatifs. Soit V' l'ensemble des sous-groupes abéliens maximaux de G tels que, pour tout Z dans V' , il existe $i \neq j$ tel que $H_i \cap Z$ et $H_j \cap Z$ sont non triviaux. Soit T' le graphe bipartite de sommets $V \amalg V'$, avec une arête entre $i \in V$ et $Z \in V'$ si et seulement si $H_i \cap Z$ est non trivial. Alors T' est connexe, car T l'est et G est commutatif-transitif, en particulier tout sous-groupe abélien non trivial est contenu dans un unique sous-groupe abélien maximal. Le graphe T' est naturellement muni d'une action de G . Comme les sous-groupes abéliens maximaux de G sont malnormaux, ils sont égaux à leurs normalisateurs, donc le stabilisateur d'un sommet Z dans V' est Z . Donc les stabilisateurs d'arête sont abéliens. Comme H_i est le fixateur d'un unique sommet de T , il est aussi égal à son normalisateur dans G , donc, pour tout i dans V , le fixateur dans G du sommet i de T est H_i .

De plus T' est un arbre. En effet, par l'absurde, supposons qu'il existe un cycle réduit minimal dans T' , de sommets $i_1, Z_1, \dots, i_n, Z_n$ avec Z_j dans V' . Alors $n \geq 2$ et, avec v_{i_j} le sommet de T stabilisé par H_{i_j} , et K le sous-arbre fini de T qui est l'enveloppe convexe des v_{i_j} , on peut supposer que v_{i_2} est un sommet terminal de K . Soient $a \in H_{i_1} \cap Z_1 - \{1\}$, $b \in H_{i_2} \cap Z_1 - \{1\}$, $c \in H_{i_2} \cap Z_2 - \{1\}$ et $d \in H_{i_3} \cap Z_2 - \{1\}$, et e l'arête de T d'origine v_{i_2} commune aux segments $[v_{i_1}, v_{i_2}]$ et $[v_{i_2}, v_{i_3}]$. Comme a fixe v_{i_1} , b fixe v_{i_2} et a, b commutent, le segment $[v_{i_1}, v_{i_2}]$ est contenu dans $\text{Fix}(a) \cup \text{Fix}(b)$. Donc l'arête e est fixée par a ou par b , qui appartiennent à Z_1 , et de même, par c ou par d , qui appartiennent à Z_2 . Comme le stabilisateur de e est infini cyclique, il existe x dans $Z_1 - \{1\}$ et y dans $Z_2 - \{1\}$ qui ont une puissance égale (différente de 1). Comme G est commutatif-transitif, on a donc $Z_1 = Z_2$, ce qui contredit la minimalité.

Pour tout sommet Z dans V' , montrons qu'il existe au plus une arête issue de Z dans T' dont le stabilisateur n'est pas infini cyclique. Sinon, soient $i \neq i'$ dans V des sommets joints par une arête à Z , et a, b dans $Z \cap H_i$ n'appartenant pas à un même sous-groupe infini cyclique de G , et de même pour a', b' dans $Z \cap H_{i'}$. Ceci est possible, car les sous-groupes abéliens de G sont abéliens libres. Comme a, b fixent v_i , a', b' fixent $v_{i'}$ et a, b, a', b' commutent, le segment non trivial $[v_i, v_{i'}]$ est contenu dans $\text{Fix}(a) \cup \text{Fix}(a'), \text{Fix}(a) \cup \text{Fix}(b'), \text{Fix}(b) \cup \text{Fix}(a')$ et $\text{Fix}(b) \cup \text{Fix}(b')$. Ceci implique que a, b fixent l'arête de $[v_i, v_{i'}]$ d'origine v_i ou que a', b' fixent l'arête de $[v_i, v_{i'}]$ d'origine $v_{i'}$. Comme l'action de G sur T est cyclique, ceci est une contradiction.

On considère maintenant l'arbre T_{can} obtenu en écrasant, pour tout sommet Z dans V' dont part une arête e de stabilisateur non infini cyclique, cette arête en un point. Notons que si i est l'extrémité de e , alors Z est contenu dans H_i , car un automorphisme d'un arbre, qui commute avec un groupe d'automorphismes ayant un point fixe unique, fixe aussi ce point.

Si H_i est de type surface dans T , alors il est encore de type surface dans T_{can} . Il est facile de voir que T_{can} vérifie les conditions voulues. \triangleright

Soit G un groupe, muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$, qui est un groupe fondamental d'un graphe de groupes (X, G_\bullet) , tel que chaque \check{G}_j est conjugué à un sous-groupe d'un groupe de sommet. Nous noterons $\text{Mod}(X, G_\bullet)$ le sous-groupe du groupe des automorphismes de G engendré par les éléments valant une conjugaison sur chaque \check{G}_j parmi les suivants (en utilisant les notations et définitions du paragraphe 4.1) :

- les conjugaisons intérieures $\tau_g : h \mapsto ghg^{-1}$,
- les twists de Dehn sur les arêtes du graphe de groupes (X, G_\bullet) ,
- pour tout groupe de sommet v de type surface de (X, G_\bullet) , par l'extension à G des éléments de $\text{Aut}(G_v, \partial)$,
- pour tout groupe de sommet G_v abélien de (X, G_\bullet) , par l'extension à G des éléments de $\text{Aut}(G_v, \partial)$.

Soit G un groupe limite, muni d'une famille de sous-groupes $(\check{G}_j)_{j \in J}$, relativement librement indécomposable, et T_{can} la décomposition cyclique relative de G donnée par la proposition 4.2. Nous noterons $\text{Mod}(G)$, et nous appellerons *groupe modulaire relatif de G* le sous-groupe $\text{Mod}(G \setminus T_{can})$ du groupe des automorphismes de G (qui coïncide avec celui de la décomposition JSJ cyclique relative associée à T_{can}). Il découle du fait que le groupe fondamental d'un graphe de groupes est indépendant «à conjugaison près» des points bases, du fait qu'un graphe de groupes quotient est bien défini à isomorphisme de graphes de groupes près et de la proposition 4.2, que le groupe modulaire relatif de G est bien défini dans $\text{Aut}(G)$. On omet le terme «relatif» si $J = \emptyset$. Le groupe modulaire peut être réduit aux automorphismes intérieurs (par exemple si T_{can} est réduit à un point et G n'est ni abélien ni un groupe de surface). Par exemple, si H est un groupe de surface fermée, alors $\text{Mod}(H)$ correspond au groupe modulaire de H défini avant le théorème 4.1.

5. DIAGRAMME ET RÉOLUTIONS DE MAKANIN-RAZBOROV D'UN GROUPE LIMITE

5.1. Présentation finie des groupes limites

Le but de ce paragraphe est de montrer qu'un groupe limite est de présentation finie. Ce résultat est dû à Kharlampovich-Myasnikov [KM2, Corollary 3] et à Sela [Sel1, Corollary 4.4], mais nous utiliserons l'approche de Gross [Gro] et de Guirardel [Gui] par les actions de

groupes sur les \mathbf{R}^n -arbres, en reprenant essentiellement la présentation de cette dernière référence. Bien que non nécessaire pour la construction du diagramme de Makanin-Razborov (voir la remarque suivant la proposition 5.5), nous l'énonçons en premier, car sa preuve permet aussi d'expliciter la structure des sous-groupes abéliens des groupes limites, ce qui sera utile par la suite.

Soit Λ un groupe abélien ordonné muni de sa valeur absolue $|x| = \sup\{x, -x\}$. Par exemple, $(\mathbf{R}^n, +)$ avec l'ordre lexicographique, ou le groupe ultraproduit ${}^\omega\Lambda'$ d'un groupe abélien ordonné Λ' , muni de l'ordre $[x_i]_\omega \leq [y_i]_\omega$ si et seulement si $x_i \leq y_i$ pour ω -presque tout i , sont des groupes abéliens ordonnés. Notons que la définition d'une distance n'utilise que les axiomes de groupe abélien ordonné du but.

Un Λ -arbre est un ensemble T muni d'une distance d à valeurs dans Λ telle que

- (1) pour tous x, y dans T , il existe a, b dans Λ et $i : [a, b] \rightarrow T$ une isométrie telle que $i(a) = x, i(b) = y$; on note $[x, y]$ l'image d'un tel i , que l'on appelle un *segment* d'extrémités x, y ;
- (2) l'intersection de deux segments ayant une extrémité commune est un segment ;
- (3) si deux segments se rencontrent exactement en une extrémité, alors leur réunion est un segment.

La seconde condition implique l'unicité d'un segment entre deux points. Par exemple, les \mathbf{Z} -arbres sont les ensembles de sommets d'arbres simpliciaux, munis de la distance maximale telle que la distance entre deux sommets d'une arête est 1. L'ensemble ultraproduit ${}^\omega T$ d'un Λ -arbre T est naturellement muni d'une structure de ${}^\omega\Lambda$ -arbre, par $d([x_i]_\omega, [y_i]_\omega) = [d(x_i, y_i)]_\omega$. Voir par exemple [Chi] pour d'autres informations.

En utilisant que le groupe libre \mathbf{L}_2 se plonge dans $\mathrm{SL}_2(\mathbf{Q}_p)$ en agissant librement sur le \mathbf{Z} -arbre de Bruhat-Tits T de $\mathrm{SL}_2(\mathbf{Q}_p)$, qu'un groupe limite se plonge dans ${}^\omega\mathbf{L}_2$, donc dans $\mathrm{SL}_2({}^\omega\mathbf{Q}_p)$ en agissant librement sur ${}^\omega T$, qui est un ${}^\omega\mathbf{Z}$ -arbre, et en utilisant le fait que G est de type fini pour montrer qu'il existe un sous- \mathbf{R}^n -arbre invariant, on obtient le résultat suivant.

PROPOSITION 5.1 (Remeslennikov [Rem], voir aussi Guirardel [Gui]). — *Pour tout groupe limite G , il existe n dans \mathbf{N} tel que G agisse librement sur un \mathbf{R}^n -arbre.* \triangleright

Si $n = 1$, alors la structure de G est bien comprise, par le théorème de Rips suivant (voir [GLP, BF3, Pau2]).

THÉORÈME 5.2 (Rips). — *Un groupe de type fini agissant librement sur un \mathbf{R} -arbre est un produit libre de groupes de surface et de groupes abéliens libres.* \triangleright

Le résultat suivant permet de dévisser la structure des groupes limites par récurrence sur n .

THÉORÈME 5.3 (Gross [Gro], Guirardel [Gui]). — *Soit G un groupe de type fini, librement indécomposable, agissant librement sur un \mathbf{R}^n -arbre, avec $n \geq 2$. Alors G admet une*

action cyclique sur un arbre simplicial, dont les groupes de sommet sont de type fini et agissent librement sur un \mathbf{R}^{n-1} -arbre, et telle que tout sous-groupe abélien non cyclique de G fixe un sommet. \triangleright

COROLLAIRE 5.4 (Kharlampovich-Myasnikov [KM2], Sela [Sel1]). — (1) *Tout groupe limite est de présentation finie.*

(2) *Tout sous-groupe abélien d'un groupe limite est abélien libre de type fini. En particulier, tout élément non trivial d'un groupe limite est puissance d'un élément qui n'est pas une puissance non triviale. Le nombre des classes de conjugaison des sous-groupes abéliens non cycliques d'un groupe limite est fini.*

(3) *Si un groupe limite n'a pas de sous-groupe abélien libre non cyclique, alors il est hyperbolique au sens de Gromov.*

(4) *Si un groupe limite est librement indécomposable, non abélien, non isomorphe à un groupe de surface, alors toute décomposition JSJ cyclique de ce groupe est non triviale.*

PREUVE. Ceci se montre par récurrence, en utilisant, respectivement, le fait que le groupe fondamental d'un graphe fini de groupes de présentation finie est de présentation finie, le fait qu'un sous-groupe abélien non cyclique d'un produit libre est conjugué dans un facteur, le théorème de combinaison de Bestvina-Feighn [BF2], et le fait qu'une décomposition JSJ cyclique d'un groupe librement indécomposable, non isomorphe à un groupe de surface, qui admet une action cyclique non triviale, n'est pas réduite à un sommet. \triangleright

5.2. Un ordre sur les groupes marqués

Nous suivons [CG] dans cette partie. On ordonne \mathbf{GM} par $(G, S) \geq (G', S')$ s'il existe un morphisme de groupes marqués de (G, S) dans (G', S') . L'unicité des morphismes montre que cette relation est bien antisymétrique. De manière équivalente, on a $(G, S) \geq (G', S')$ si et seulement si les éléments de S' vérifient dans G' toutes les relations vérifiées par S dans G : pour tout élément $m(\bar{s}_1, \dots, \bar{s}_k)$ de \mathbf{L} , si le mot $m(s_1, \dots, s_k)$ dans $S \cup S^{-1}$ vaut 1 dans G , alors le mot $m(s'_1, \dots, s'_k)$ dans $S' \cup S'^{-1}$ vaut 1 dans G' .

PROPOSITION 5.5. — *Soit K un compact de \mathbf{GM} , formé de groupes limites marqués. Alors il existe une partie finie F de K , non vide si K l'est, telle que, pour tout (G', S') dans K , il existe (G, S) dans F tel que $(G', S') \leq (G, S)$.*

PREUVE. Un groupe limite est de présentation finie par le corollaire 5.4 (1). Tout groupe de présentation finie (G, S) marqué admet un voisinage formé d'éléments inférieurs ou égaux à (G, S) (par la remarque (3) du chapitre 2). \triangleright

PROPOSITION 5.6. — *Toute suite décroissante de groupes limites marqués est stationnaire.*

PREUVE. Une valeur d'adhérence de cette suite est un groupe limite marqué, inférieur ou égal à tout élément de la suite, et admettant un voisinage formé d'éléments qui lui sont inférieurs ou égaux. \triangleright

COROLLAIRE 5.7. — *Tout groupe limite G est hopfien (i.e. tout morphisme surjectif de G dans lui-même est injectif).*

PREUVE. Soit $f : G \rightarrow G$ un morphisme surjectif non injectif, de noyau N . Soit $\psi : G \rightarrow G$ la composition de la projection canonique $G \rightarrow G/N$ et de l'isomorphisme $G/N \rightarrow G$ induit par f . Alors $G \xrightarrow{\psi} G \xrightarrow{\psi} G \rightarrow \dots$ est une suite décroissante non stationnaire de groupes limites. \triangleright

Nous renvoyons à [Chat, CG], pour une preuve simple des deux propositions précédentes, évitant d'utiliser le fait qu'un groupe limite est de présentation finie.

Nous renvoyons aussi à [BMR, KM1, KM2] pour de très jolis résultats de finitude analogues aux précédents (ainsi qu'au corollaire suivant). Si F est un groupe libre, et X un n -uplet de variables, alors l'ensemble des parties $V(\Sigma)$ de F^p qui sont les ensembles de solutions de systèmes (Σ) (finis ou infinis) d'équations de la forme $m(x_1, \dots, x_p, g_1, \dots, g_q) = 1$ avec x_i dans X , g_i dans F et m un mot à x_i^\pm, g_i^\pm , est l'ensemble des fermés d'une topologie noethérienne. Si $N_{(\Sigma)}$ est le sous-groupe (qui est distingué) de $F * \mathbf{L}(X)$ engendré par les éléments m tels que l'équation $m = 1$ soit conséquence de (Σ) , alors le fermé $V(\Sigma)$ est irréductible (i.e. non réunion de fermés propres) si et seulement si le groupe $F * \mathbf{L}(X)/N_{(\Sigma)}$ est un groupe limite (voir [BMR]).

Soit (G, S) un groupe marqué. On note $\mathbf{GM}(G, S)$ le sous-espace de \mathbf{GM} formé des groupes marqués inférieurs ou égaux à (G, S) . Il est compact (car fermé et contenu dans \mathbf{GM}_k si $S = \{s_i\}_{i \in \mathbb{N}}$ avec $s_i = 1$ pour $i > k$), et ouvert si G est de présentation finie.

L'ensemble $\mathbf{L}(G, S)$ des groupes limites marqués inférieurs à (G, S) est compact. Il est réduit au groupe trivial si et seulement si G ne se surjecte pas sur \mathbf{Z} (par exemple si G est fini, ou est un groupe de Kazhdan). Le résultat suivant découle de la proposition 5.5.

COROLLAIRE 5.8. — *Pour tout groupe marqué (G, S) , l'ensemble $\mathbf{L}_{\max}(G, S)$ des groupes limites, inférieurs à (G, S) , maximaux pour ces propriétés, est fini, non vide.* \triangleright

L'ensemble $\mathbf{GM}(G, S)$ s'identifie avec l'ensemble $\text{Epi}(G, \cdot)$ des classes d'équivalence de morphismes de groupes surjectifs $f : G \rightarrow H$, pour la relation d'équivalence

$$(f : G \rightarrow H) \sim (f' : G \rightarrow H')$$

s'il existe un isomorphisme de groupes $g : H \rightarrow H'$ tel que $g \circ f = f'$, via l'application qui au morphisme surjectif $f : G \rightarrow H$ associe le groupe marqué $(H, f(S))$. Nous noterons de même une classe et un de ses représentants.

En particulier, $\text{Epi}(G, \cdot)$ hérite d'une topologie métrisable compacte, indépendante de S : une suite $(f_n)_{n \in \mathbb{N}}$ converge vers f si et seulement si, pour toute partie finie P de G , il

existe N dans \mathbb{N} tel que pour tout $n \geq N$, et g dans P , on a $f(g) = 1$ si et seulement si $f_n(g) = 1$. L'application $\text{Epi}(G, \cdot) \rightarrow \text{Epi}(G', \cdot)$ obtenue en composant par un morphisme surjectif donné de G' dans G est continue. Il découle du chapitre 3 qu'un groupe G est un groupe limite si et seulement s'il existe une suite $(f_n)_{n \in \mathbb{N}}$ dans $\text{Epi}(G, \cdot)$, d'image un groupe libre, convergeant vers l'identité. De plus, s'il existe une suite de morphismes d'un groupe de type fini G dans un groupe limite qui converge vers l'identité, alors G est un groupe limite.

Le groupe (muni de la topologie discrète) $\text{Aut}(G)$ des automorphismes du groupe G agit continuellement sur $\text{Epi}(G, \cdot)$ par précomposition.

Fixons (G, S) un groupe limite librement indécomposable marqué. Si F est un groupe libre de type fini, et E une famille d'éléments de F , on note $|||E|||$ la borne inférieure, sur toutes les parties génératrices libres A de F du maximum des longueurs des écritures réduites dans $A \cup A^{-1}$ des éléments de E . Un élément $f : G \rightarrow F$ de $\text{Epi}(G, \cdot)$ (et son élément correspondant de $\mathbf{GM}(G, S)$) est dit (G, S) -court si F est un groupe libre, et si

$$|||f(S)||| = \min_{\varphi \in \text{Mod}(G)} |||f \circ \varphi(S)|||.$$

Ceci dépend de S , mais ne dépend pas de la classe d'isomorphisme de f dans $\text{Epi}(G, \cdot)$. Un *quotient raccourcissant* de (G, S) est un groupe marqué, limite de groupes marqués (G, S) -courts.

Par exemple, soit G le groupe fondamental de la surface connexe fermée orientable de genre $g \geq 2$, muni de sa présentation usuelle $\langle a_1, b_1, \dots, a_g, b_g \mid [a_1, b_1] \dots [a_g, b_g] \rangle$, marqué par $S = (a_1, b_1, \dots, a_g, b_g)$. Soit \mathbf{L}_g le groupe libre sur $(\bar{s}_1, \dots, \bar{s}_g)$, marqué par $S' = (\bar{s}_1, 1, \bar{s}_2, 1, \dots, \bar{s}_g, 1)$. Alors (\mathbf{L}_g, S') est un quotient raccourcissant de (G, S) , car (G, S) -court.

L'ensemble $\mathbf{SQ}(G, S)$ des quotients raccourcissants de (G, S) est un compact de $\mathbf{GM}(G, S)$ (car fermé), non vide (car il contient le groupe trivial marqué), et formé de groupes limites. Le résultat suivant découle alors de la proposition 5.5.

COROLLAIRE 5.9. — *L'ensemble $\mathbf{SQ}_{\max}(G, S)$ des quotients raccourcissants maximaux d'un groupe limite librement indécomposable marqué (G, S) est fini, non vide.* \triangleright

Le résultat fondamental sur les quotients raccourcissants est le suivant.

THÉORÈME 5.10 (Claim 5.3 [Sel1]). — *Tout quotient raccourcissant d'un groupe limite librement indécomposable non trivial marqué (G, S) est un quotient strict de (G, S) .* \triangleright

L'idée consiste à considérer une suite de morphismes surjectifs $f_n : G \rightarrow F$ qui sont (G, S) -courts et convergent vers un quotient raccourcissant $f : G \rightarrow G'$ (on peut bien supposer que les groupes libres images des f_n sont constants), et une action libre de F sur un arbre simplicial. On montre que la suite d'actions de G sur T obtenue par précomposition par f_n converge (pour la topologie de Gromov équivariante) vers une

action de G sur un arbre réel, qui factorise par G' . On utilise alors la structure fine des actions de groupes sur les arbres (voir par exemple [BF3, Pau2]) pour empêcher que f vaille l'identité.

5.3. Diagramme de Makanin-Razborov d'un groupe marqué

Dans un arbre enraciné \mathbf{T} , toute arête e sera orientée avec son origine $o(e)$ plus proche de la racine que son extrémité $t(e)$. On notera v_* la racine, et si v est sommet, on notera $\mathbf{T}(v)$ l'ensemble des extrémités des arêtes d'origine v .

Un *arbre enraciné de groupes marqués* $(\mathbf{T}, G_\bullet, S_\bullet)$ est la donnée d'un arbre enraciné \mathbf{T} , pour tout sommet v de \mathbf{T} d'un groupe marqué (G_v, S_v) , et pour toute arête e d'un morphisme de groupes marqués $(G_{o(e)}, S_{o(e)}) \rightarrow (G_{t(e)}, S_{t(e)})$.

Un *diagramme de Makanin-Razborov* d'un groupe marqué (G, S) est l'arbre enraciné de groupes marqués $(\mathbf{T}, G_\bullet, S_\bullet)$, défini récursivement ainsi :

- le groupe marqué de v_* est (G, S) ,
- l'ensemble $\mathbf{T}(v_*)$ est $\mathbf{L}_{\max}(G, S)$ si (G, S) n'est pas un groupe limite marqué,
- pour tout sommet v de \mathbf{T} si (G, S) est un groupe limite marqué, et tout sommet v de $\mathbf{T} - \{v_*\}$ sinon,
 - si G_v est un groupe libre, alors le sommet v est terminal,
 - si G_v n'est pas libre et n'est pas librement indécomposable, alors $\mathbf{T}(v)$ est l'ensemble des facteurs de Grushko des décompositions de Grushko S_v -minimales de G_v , marqués par l'image de S_v par la surjection de G_v (voir le paragraphe 4.2),
 - si G_v est librement indécomposable et non trivial, alors $\mathbf{T}(v)$ est $\mathbf{SQ}_{\max}(G_v, S_v)$;

de plus, le groupe marqué associé à un sommet v dans $\mathbf{T} - \{v_*\}$ est v , le morphisme associé à chaque arête est l'unique morphisme entre les groupes marqués des sommets de l'arête.

Par exemple, le diagramme de Makanin-Razborov d'un groupe libre marqué est réduit à un seul sommet. Le diagramme de Makanin-Razborov d'un groupe non trivial marqué ne se surjectant pas sur \mathbf{Z} est réduit à une arête, et le groupe marqué du sommet non racine est trivial. Par construction, le sous-arbre enraciné de groupes marqués en dessous (au sens évident) d'un sommet v d'un diagramme de Makanin-Razborov est un diagramme de Makanin-Razborov de (G_v, S_v) . Un diagramme de Makanin-Razborov d'un groupe marqué dépend en général du marquage.

PROPOSITION 5.11. — *Un diagramme de Makanin-Razborov d'un groupe limite a un nombre fini de sommets.*

PREUVE. Les corollaires 5.8, 5.9 et le paragraphe 4.2 montrent que l'arbre enraciné de Makanin-Razborov est localement fini. Le théorème 5.10 et le corollaire 5.6 montrent qu'il n'a pas de rayon infini partant de la racine. \triangleright

Un diagramme de Makanin-Razborov $(\mathbf{T}, G_\bullet, S_\bullet)$ d'un groupe marqué (G_0, S_0) permet de «décrire» tous les morphismes de groupes de G_0 dans un groupe libre F_0 , de manière «indépendante» de ce groupe libre, au sens suivant.

Notons que l'ensemble des morphismes d'un groupe libre de type fini F dans un autre F' est facile à comprendre, un tel morphisme étant défini, de manière unique, par une famille $(y_j)_{j=1,\dots,k}$ dans F' si $(x_j)_{j=1,\dots,k}$ est une base de F , par $x_j \mapsto y_j$. Comme tout sous-groupe d'un groupe libre est libre, l'étude des morphismes à valeurs dans un groupe libre se ramène à l'étude des morphismes surjectifs à valeurs dans un groupe libre.

Considérons la donnée $(\mathbf{T}', (f_v), (\varphi_v))$ suivante. Soit \mathbf{T}' un sous-arbre enraciné de \mathbf{T} (contenant la racine de \mathbf{T}) tel que, pour tout sommet v non terminal de \mathbf{T}' , si G_v est un groupe limite non librement indécomposable, alors $\mathbf{T}'(v) = \{v_1, \dots, v_p\}$ avec $G_v = G_{v_1} * \dots * G_{v_p}$ une décomposition de Grushko S_v -minimale de G_v , et si G_v n'est pas un groupe limite, ou est librement indécomposable non trivial, alors $\mathbf{T}'(v)$ contient exactement un sommet v_+ de $\mathbf{T}(v)$. Pour tout sommet v de \mathbf{T}' tel que G_v est librement indécomposable non trivial, soit φ_v dans $\text{Mod}(G_v)$. Pour tout sommet v terminal dans \mathbf{T}' (donc dans \mathbf{T} , donc G_v est libre), soit $f_v : G_v \rightarrow F_0$ un morphisme de groupes.

Construisons par récurrence inverse un morphisme de groupes $f_v : G_v \rightarrow F_0$ pour tout sommet v de \mathbf{T}' . Si v est un sommet terminal, alors f_v est donné. Si v n'est pas un sommet terminal, alors

- si G_v est un groupe limite non librement indécomposable et si $\mathbf{T}'(v) = \{v_1, \dots, v_p\}$, alors on pose $f_v = f_{v_1} * \dots * f_{v_p}$ (avec les notations du paragraphe 4.2),
- si G_v est un groupe limite librement indécomposable non trivial, en notant par $\pi_{v_+} : (G_v, S_v) \rightarrow (G_{v_+}, S_{v_+})$ le morphisme du diagramme de Makanin-Razborov, alors on définit $f_v = f_{v_+} \circ \pi_{v_+} \circ \varphi_v^{-1}$ de sorte que le diagramme suivant commute

$$\begin{array}{ccc} G_v & & \\ \pi_{v_+} \downarrow & \searrow^{f_v \circ \varphi_v} & \\ G_{v_+} & \xrightarrow{f_{v_+}} & F_0 \end{array}$$

- si G_v n'est pas un groupe limite, et si $\pi_{v_+} : (G_v, S_v) \rightarrow (G_{v_+}, S_{v_+})$ est le morphisme du diagramme de Makanin-Razborov, alors on pose $f_v = f_{v_+} \circ \pi_{v_+}$

Posons $f(\mathbf{T}', (f_v), (\varphi_v)) = f_{v_*}$, qui est un morphisme de groupes de G_0 dans F_0 .

PROPOSITION 5.12. — *Soit (G_0, S_0) un groupe marqué, de diagramme de Makanin-Razborov $(\mathbf{T}, G_\bullet, S_\bullet)$. Si f est un morphisme de groupes de G_0 dans un groupe libre F_0 , alors il existe une donnée $(\mathbf{T}', (f_v), (\varphi_v))$ telle que $f = f(\mathbf{T}', (f_v), (\varphi_v))$.*

PREUVE. Par récurrence, on construit une donnée $(\mathbf{T}', (f_v), (\varphi_v))$, ainsi qu'un morphisme de groupes $f_v : G_v \rightarrow F_0$ pour tout sommet v de \mathbf{T}' . On demande que la racine de \mathbf{T} appartienne à \mathbf{T}' , et on pose $f_{v_*} = f$. Supposons construit un sommet v de \mathbf{T}' , et le morphisme $f_v : G_v \rightarrow F_0$.

Si v est terminal dans \mathbf{T} , alors on dit qu'il est terminal dans \mathbf{T}' , et on prend $f_v : G_v \rightarrow F_0$ pour la donnée. Supposons donc que v n'est pas terminal (en particulier G_v n'est pas libre).

Supposons que G_v soit un groupe limite, non librement indécomposable. On choisit une décomposition de Grushko S_v -minimale $G_v = G_{v_1} * \cdots * G_{v_p}$ de G_v . On pose alors $\mathbf{T}'(v) = \{v_1, \dots, v_p\}$, on note f_{v_j} la restriction de f_v à G_{v_j} pour $j = 1 \dots p$. D'après le paragraphe 4.2, on a $f_v = f_{v_1} * \cdots * f_{v_p}$.

Supposons que G_v soit un groupe limite librement indécomposable. Soit φ_v dans $\text{Mod}(G_v)$ tel que

$$|||f_v \circ \varphi_v(S_v)||| = \min_{\psi \in \text{Mod}(G_v)} |||f_v \circ \psi(S_v)|||.$$

Alors le groupe marqué $(F_0, f_v(\varphi_v(S_v)))$ est (G_v, S_v) -court, donc est un quotient raccourcissant de (G_v, S_v) . Il existe donc v_+ dans $\mathbf{T}(v) = \mathbf{SQ}_{\max}(G_v, S_v)$ tel que $(G_{v_+}, S_{v_+}) \geq (F_0, f_v(\varphi_v(S_v)))$. On note alors $\mathbf{T}'(v) = \{v_+\}$, et $f_{v_+} : (G_{v_+}, S_{v_+}) \rightarrow (F_0, f_v(\varphi_v(S_v)))$ l'unique morphisme, qui vérifie $f_v \circ \varphi_v = f_{v_+} \circ \pi_{v_+}$.

Supposons que G_v n'est pas un groupe limite (alors $v = v_*$). Comme $(F_0, f_v(G_v))$ est un groupe libre (donc limite) marqué inférieur à (G_v, S_v) , il existe v_+ dans $\mathbf{T}(v) = \mathbf{L}_{\max}(G_v, S_v)$ tel que $(G_{v_+}, S_{v_+}) \geq (F_0, f_v(G_v))$. On note donc $\mathbf{T}'(v) = \{v_+\}$, et $f_{v_+} : (G_{v_+}, S_{v_+}) \rightarrow (F_0, f_v(G_v))$ l'unique morphisme, qui vérifie $f_v = f_{v_+} \circ \pi_{v_+}$.

Par définition, on a bien $f = f_{v_*} = f(\mathbf{T}', (f_v), (\varphi_v))$. ▷

6. RÉOLUTION DE MAKANIN-RAZBOROV DE GROUPE LIMITE

Le contenu de cette partie est essentiellement repris de l'article de Champetier-Guirardel [CG], et fournit des preuves détaillées de la seconde moitié du chapitre 5 de [Sel1]. Il s'agit de donner une caractérisation des groupes limites par « dévissages » successifs à partir des groupes libres.

Si G_v est le stabilisateur d'un sommet d'une action cyclique d'un groupe G sur un arbre, appelons *voisinage abélien* de G_v le sous-groupe de G engendré par G_v et par les centralisateurs dans G des groupes d'arête d'origine v . Un morphisme de groupes $f : G \rightarrow G'$ est une *résolution de Makanin-Razborov élémentaire* s'il existe une décomposition cyclique en graphe de groupes de G , telle que

- (1) chaque groupe d'arête est abélien maximal dans au moins l'un des groupes de sommet de l'arête ;
- (2) le morphisme f est injectif sur chaque groupe d'arête, et le centralisateur de chaque groupe d'arête est égal au centralisateur de l'une de ses deux images dans ses groupes de sommet ;
- (3) l'image par f de chaque groupe de sommet de type surface est non abélienne ;
- (4) tout groupe de sommet abélien G_v est abélien libre de type fini, f est injective sur le sous-groupe G'_v de G_v engendré par les groupes d'arête d'origine ce sommet ;

- (5) pour tout groupe de sommet G_v qui est non abélien et non de type surface, le morphisme f est injectif sur le voisinage abélien de G_v dans le graphe de groupes obtenu en remplaçant chaque groupe de sommet abélien par son sous-groupe engendré par les groupes d'arête d'origine ce sommet.

Si G est un groupe de type fini, une *résolution de Makanin-Razborov* est une suite de morphismes

$$G_0 \xrightarrow{f_1} G_1 \xrightarrow{f_2} \dots \xrightarrow{f_n} G_n$$

avec $G_0 = G$ et G_n un groupe libre, telle qu'il existe, pour $i = 1, \dots, n$, des décompositions en produit libres $G_{i-1} = G'_1 * \dots * G'_p * F$ et $G_i = G''_1 * \dots * G''_p$, avec p dans \mathbb{N} et F un groupe libre, telles que pour tout $j = 1, \dots, p$, on a $f_i(G'_j) = G''_j$ et $f_i : G'_j \rightarrow G''_j$ est une résolution de Makanin-Razborov élémentaire. Cette définition est une version légèrement corrigée de la notion de «strict MR resolution» dans [Sel1]), car le théorème 5.2 de [Sel1] est légèrement erroné, voir [CG].

THÉORÈME 6.1 (Sela [Sel1] Theorem 5.12). — *Un groupe de type fini G est un groupe limite si et seulement s'il admet une résolution de Makanin-Razborov.*

PREUVE. Supposons tout d'abord que le groupe G admette une résolution de Makanin-Razborov. Par récurrence sur la longueur, comme un groupe libre est un groupe limite, comme un sous-groupe de type fini d'un groupe limite et un produit libre d'un nombre fini de groupes limites sont des groupes limites, il suffit de montrer que si L est un groupe limite et $f : G \rightarrow L$ est une résolution de Makanin-Razborov élémentaire, alors G est un groupe limite. On se fixe une décomposition en graphe de groupes de G comme dans la définition, avec (X, G_*) son graphe de groupes.

Le résultat suivant servira dans plusieurs étapes de la preuve.

LEMME 6.2 (Baumslag). — Soit G' un groupe limite, n un entier non nul, a_0, a_1, \dots, a_n des éléments de G' et c un élément de G' ne commutant avec aucun des éléments a_1, \dots, a_{n-1} . Alors il existe N dans \mathbb{N} tel que pour tous k_1, \dots, k_n dans \mathbb{Z} avec $|k_i| \geq N$, l'élément $a_0 c^{k_1} a_1 c^{k_2} a_2 \dots c^{k_n} a_n$ est non trivial dans G' .

PREUVE. C'est un exercice si G' est un groupe libre (voir par exemple [Bau1, Prop. 1]), et on se ramène à ce cas en considérant une suite de morphismes de G' dans un groupe libre convergeant vers l'identité. \triangleright

Étape 1 : La première étape est de se ramener au cas où les groupes de sommet abéliens de (X, G_*) sont engendrés par les groupes d'arête d'origine ce sommet.

Un groupe G' est obtenu par *extension libre de centralisateurs* à partir d'un groupe G'' s'il existe un groupe abélien libre de type fini A et un élément g de G'' de centralisateur Z dans G'' , tels que G' est (à isomorphisme près) le produit amalgamé $G'' *_Z (Z \times A)$ pour les morphismes $x \mapsto x$ de Z dans G'' et $z \mapsto (z, 0)$ de Z dans $Z \times A$.

PROPOSITION 6.3 (Baumslag). — *Une extension libre de centralisateurs d'un groupe limite est un groupe limite. De plus, avec les notations précédentes, le centralisateur de g dans G' est $Z \times A$.*

PREUVE. Soit $f : G' \rightarrow G''$ le morphisme surjectif valant l'identité sur G'' et trivial sur A (en identifiant G'' et A avec leurs images dans G'). Montrons qu'il existe une suite $(\varphi_n)_{n \in \mathbb{N}}$ d'automorphismes de groupes de $Z \times A$, valant l'identité sur le premier facteur, ce qui permet de l'étendre (par l'identité sur G'') à G' , telle que $f \circ \varphi_n$ converge vers l'identité. Ceci entraînera que G' est un groupe limite. En effet, on peut supposer g non trivial, et il suffit de prendre, avec (a_1, \dots, a_p) une base de A , l'automorphisme $\varphi_n(x, a_1^{k_1} \dots a_p^{k_p}) = (xg^{n(k_1 + \dots + k_p)}, a_1^{k_1} \dots a_p^{k_p})$. Pour vérifier que $f \circ \varphi_n$ converge vers l'identité, on peut supposer par récurrence $p = 1$, et on applique le lemme 6.2.

Une autre manière de montrer le résultat est de dire, en supposant A infini cyclique engendré par a , que si S est une partie génératrice finie de G'' , alors $S_n = S \cup \{g^n\}$ aussi, et la suite de groupes marqués (G'', S_n) converge vers le groupe marqué $(G', S \cup \{a\})$ si $g \neq 1$, par le lemme 6.2. \triangleright

REMARQUE. Cette notion a été beaucoup étudiée (voir les travaux de Lyndon, Baumslag, Kharlampovich, Myasnikov, Remeslennikov). Il est montré dans [KM2] (voir aussi [CG]) qu'un groupe de type fini est un groupe limite si et seulement s'il est isomorphe à un sous-groupe d'un groupe obtenu par un nombre fini d'extensions libres de centralisateurs à partir d'un groupe libre (la proposition ci-dessus montrant le sens réciproque).

La première étape découle par application d'un nombre fini de fois cette proposition, car les conditions (1)-(5) de la définition de résolution de Makanin-Razborov élémentaire sont préservées si on remplace chaque groupe de sommet abélien par son sous-groupe engendré par les groupes d'arête d'origine ce sommet, et $f : G \rightarrow L$ par sa restriction au groupe fondamental du graphe de groupes obtenu par ces remplacements.

Nous supposons donc dans la suite que f est injective sur les groupes de sommet abéliens.

Étape 2 : La seconde étape consiste à montrer que, quitte à admettre les pantalons et les rubans de Möbius troués dans la définition de «sommet de type surface d'une décomposition cyclique» (voir partie 4.1), alors on peut supposer que, pour tout sommet v de type surface de (X, G_*) , le morphisme f est injectif sur le voisinage abélien de G_v (et toujours f injective sur les groupes de sommet abéliens).

Le résultat se déduit des deux lemmes suivants, qui découlent essentiellement de la structure des sous-groupes à deux générateurs d'un groupe limite (voir corollaire 3.2 (7)).

LEMME 6.4 (Sela [Sel1] Lemma 5.13). — Soit H le groupe fondamental d'une surface S compacte connexe à bord (éventuellement vide), de caractéristique d'Euler au plus -1 . Soit L un groupe limite, et $f : H \rightarrow L$ un morphisme de groupes d'image non abélienne,

injective en restriction au groupe fondamental de chaque composante connexe du bord de S . Alors il existe une famille finie de courbes fermées simples disjointes de S , telle que chaque composante connexe C du complémentaire soit un pantalon ou un ruban de Möbius troué, et telle que la restriction de f au groupe fondamental de C soit injective.

▷

LEMME 6.5 (Champetier-Guirardel [CG] Claim 4.5). — Si H est un groupe obtenu par extension libre de centralisateurs des groupes fondamentaux d'au plus deux composantes connexes du bord d'un pantalon ou d'au plus une composante connexe du bord d'un ruban de Möbius troué, et si $f : H \rightarrow L$ est un morphisme de groupes d'image un groupe limite non abélien, injectif sur les centralisateurs des groupes fondamentaux des composantes connexes du bord, alors f est injective. ▷

Étape 3 : Montrons qu'il existe un groupe limite L , un morphisme de groupes $f : G \rightarrow L$ et un graphe de groupes (X', G'_\bullet) de groupe fondamental G , dont les groupes de sommet abéliens et les groupes d'arête sont non triviaux, et égaux à leurs centralisateurs dans G , avec f injective sur les groupes de sommet. Ceci implique que les groupes d'arête sont abéliens maximaux dans les deux groupes de sommet, et que f est injective sur les voisinages abéliens des groupes de sommet.

Soient (X, G_*) et $f : G \rightarrow L$ vérifiant le résultat de l'étape 2. On construit un graphe de groupes (X', G'_*) , avec $X' = X$, de la manière suivante. On remplace chaque groupe de sommet G_v par le sous-groupe G'_v de G engendré par G_v et par les centralisateurs dans G des groupes d'arête d'origine v . On remplace chaque groupe d'arête G_e par son centralisateur dans G . Les morphismes des groupes de sommet dans les groupes d'arête sont ceux qui sont évidents.

Comme f est injective sur les groupes de sommet de (X, G_*) par l'étape 2 et la propriété (5), ceux-ci sont des groupes limites. Par la propriété (2), les groupes d'arête de (X', G'_*) sont des centralisateurs d'éléments non triviaux dans des groupes de sommet de (X, G_*) , donc sont abéliens non triviaux. Ils sont égaux à leurs centralisateurs dans G , car $Z_G(Z_G(H)) = Z_G(H)$ si H et $Z_G(H)$ sont abéliens.

Soit v un sommet de X de groupe G_v . Montrons que G'_v est égal à son centralisateur dans G si G'_v est abélien, et que f est injective sur G'_v . Si G_v n'est pas abélien, alors G'_v non plus et cela découle de l'étape (2) et de la propriété (5). Supposons donc G_v abélien.

S'il existait une arête e qui est une boucle en un sommet v de groupe abélien, alors par la propriété (1), le groupe G_v serait infini cyclique, égal par exemple à $\rho_e(G_e)$. Mais s'il existe un morphisme φ d'un groupe de présentation $\langle a, t \mid tat^{-1} = a^m \rangle$ dans un groupe limite, tel que $\varphi(a) \neq 1$, alors $m = 1$, car l'image doit être abélienne sans torsion (par le corollaire 3.2 (7) et (4)). Donc $\rho_e(G_e)$ ne serait pas égal à son centralisateur dans G , ce qui contredit (2).

Supposons tout d'abord que pour toute arête e d'origine v , on a $Z_G(G_e) = Z_{G_v}(\rho_{\bar{e}}(G_e))$. En particulier, par définition, G'_v est égal à G_v . Si $g \in Z_G(G_v)$, alors $g \in Z_G(G_e)$, donc $g \in G_v$. Donc $G'_v = G_v$ est égal à son centralisateur dans G , et f est injective sur $G'_v = G_v$ par l'étape (2).

Sinon, il existe une arête e d'origine v telle que $Z_G(G_e) \neq Z_{G_v}(\rho_{\bar{e}}(G_e))$. Par les propriétés (2) et (1), on a alors $Z_G(G_e) = Z_{G_{t(e)}}(\rho_e(G_e))$, ce groupe contient proprement $\rho_e(G_e)$ et $G_v = \rho_{\bar{e}}(G_e)$.

Supposons qu'il existe une autre arête e' d'origine v telle que $Z_G(G_{e'}) \neq Z_{G_v}(\rho_{\bar{e'}}(G_{e'}))$. Identifions G_e et $G_{e'}$ avec G_v par $\rho_{\bar{e}}$ et $\rho_{\bar{e'}}$, et G_e avec un sous-groupe de $G_{t(e)}$ par ρ_e . Si $t(e) = t(e')$, alors $G_{e'}$ s'envoie par $\rho_{e'}$ sur un sous-groupe uG_vu^{-1} de $G_{t(e)}$ pour un u dans $G - G_{t(e)}$. Si $t(e) \neq t(e')$, alors on identifie G_e avec un sous-groupe de $G_{t(e')}$ par $\rho_{e'}$. Dans ce second cas, la propriété (2) entraîne l'égalité des centralisateurs $Z_{G_{t(e)}}(G_v)$ et $Z_{G_{t(e')}}(G_v)$, qui contiennent strictement G_v . Il existerait alors un élément de $G - G_v$ fixant v dans un revêtement universel de (X, G_*) où on a relevé $e \cup e'$, ce qui est impossible. Dans le premier cas, on aurait $Z_{G_{t(e)}}(G_v) = uZ_{G_{t(e)}}(G_v)u^{-1}$. Pour a dans $G_v - \{1\}$, l'élément uau^{-1} commute avec a , donc leurs images par f commutent. Par la structure des sous-groupes à deux générateurs d'un groupe limite, $f(u)$ et $f(a)$ commutent, donc $f(uau^{-1}) = f(a)$. Comme f est injective sur $G_{t(e)}$, on en déduit que u appartient au centralisateur de G_v dans G , qui est contenu dans $G_{t(e)}$, ce qui est une contradiction.

Donc pour toute arête e' d'origine v , différente de e , le centralisateur $Z_G(G_{e'})$ est contenu dans G_v , ce qui montre par définition que $G'_v = Z_G(G_e) = Z_{G_{t(e)}}(G_e)$. En particulier, G'_v est égal à son centralisateur dans G . Si $G_{t(e)}$ n'est pas abélien, alors par l'étape (2), le morphisme f est injectif sur G'_v . Si $G_{t(e)}$ est abélien, alors $G'_v = G_{t(e)}$, et f est aussi injective sur G'_v par l'étape (1).

Étape 4 : Montrons que si G est un groupe de type fini, L un groupe limite, $f : G \rightarrow L$ un morphisme de groupes et (X', G'_\bullet) un graphe de groupes de groupe fondamental G , à groupes d'arête abéliens non triviaux et égaux à leurs centralisateurs dans G (donc abéliens maximaux dans les deux groupes de sommet), avec f injective sur les voisinages abéliens des groupes de sommet, et dont les groupes de sommet abéliens sont égaux à leurs centralisateurs dans G , alors il existe une suite de twists de Dehn φ_n sur (X', G'_\bullet) telle que $f \circ \varphi_n$ converge vers l'identité.

On procède par récurrence sur le nombre n d'arêtes de X' (que l'on peut supposer non nul). Si $n = 1$, alors G est un produit amalgamé $A *_C B$ ou une extension HNN $A *_C$. Soit c un élément non trivial de C .

Supposons tout d'abord que $G = A *_C B$ avec $C = A \cap B$. Notons φ_n le twist de Dehn qui est la conjugaison par c^n sur A et l'identité sur B . Soit g un élément non trivial de G , que l'on peut supposer n'appartenant pas à B , car φ_n est l'identité sur B et f injective sur B . Considérons son écriture réduite $g = a_1b_1 \dots a_pb_p$ avec $p \geq 1$, a_i dans $A - C$ et b_i dans $B - C$ (sauf peut-être $a_1 = 1$ auquel cas $p \geq 2$, ainsi que $b_p = 1$). Pour n assez

grand, l'élément $f \circ \varphi_n(g) = f(c)^n f(a_1) f(c)^{-n} f(b_1) \dots f(c)^n f(a_p) f(c)^{-n} f(b_p)$ du groupe limite L est non trivial par le lemme 6.2. En effet, si $f(c)$ et $f(a_i)$ commutent, comme f est injective sur A , alors a_i et c commutent, ce qui contredit le fait que a_i n'appartient pas à C , qui est le centralisateur de c dans le groupe limite A , car abélien maximal. De même, $f(c)$ et $f(b_i)$ ne commutent pas.

Supposons maintenant $G = \langle A, t \mid tct^{-1} = \theta(c), c \in C \rangle$, avec $C \subset A$ et $\theta : C \rightarrow A$ un morphisme injectif. S'il existe a dans A tel que $f(at)$ commute avec $f(c)$, alors $f(a^{-1}ca) = f(tct^{-1})$, donc par injectivité de $f|_A$, comme tct^{-1} est dans A , on a $(at)^{-1}cat = c$. Donc $f(at)^{-1}f(C)f(at) \cap f(C)$ est non trivial, et comme L est un groupe limite (voir le corollaire 3.2 (6)), l'élément $f(at)$ commute avec tout $f(C)$, et donc comme ci-dessus, at commute avec C . Comme un groupe de sommet abélien est égal à son centralisateur, le groupe A est non abélien. Donc f est injective sur le sous-groupe engendré par A et par at , donc est injectif sur G , et $\varphi_n = id$ convient (notons pour usage ultérieur que G est une extension libre de centralisateurs de A). Supposons donc qu'il n'existe pas de a dans A tel que $f(at)$ commute avec $f(c)$ (et de même pas de a dans A tel que $f(t^{-1}a)$ commute avec $f(c)$). Notons φ_n le twist de Dehn qui vaut l'identité sur A et envoie t sur tc^n . Soit g dans $G - A$ (car si g est dans $A - \{1\}$, alors $f \circ \varphi_n(g) = f(g) \neq 1$), d'écriture réduite (voir [LS, page 181]) $g = a_0 t^{\epsilon_1} a_1 t^{\epsilon_2} a_2 \dots t^{\epsilon_p} a_p$ avec $p \geq 1$, a_i dans A , et pour $i = 1, \dots, n-1$, $a_i \notin C$ si $\epsilon_i = -\epsilon_{i+1} = 1$, et $a_i \notin \theta(C)$ si $\epsilon_i = -\epsilon_{i+1} = -1$. Alors pour n assez grand, l'élément $f \circ \varphi_n(g) = f(a_0) f(tc^n)^{\epsilon_1} f(a_1) f(tc^n)^{\epsilon_2} f(a_2) \dots f(tc^n)^{\epsilon_p} f(a_p)$ du groupe limite L est non trivial par le lemme 6.2. En effet, $f(c)$ ne commute pas avec $f(a_i t)$, $f(t^{-1}a_i)$, ni avec $f(a_i)$ si $\epsilon_i = -\epsilon_{i+1} = 1$ (car C est abélien maximal dans A), ni avec $f(t^{-1}a_i t)$ si $\epsilon_i = -\epsilon_{i+1} = -1$ (car $\theta(C)$ est abélien maximal dans A).

Supposons maintenant $n \geq 2$. Soit e une arête de X' , qui induit une décomposition en produit amalgamé $A *_{G'_e} B$ de G si e sépare X' , et une décomposition en extension HNN $A *_{G'_e}$ de G' sinon. Par récurrence, il existe une suite $(\varphi_n)_{n \in \mathbb{N}}$ de twists de Dehn sur les arêtes de $X - \{e, \bar{e}\}$ telle que $f \circ \varphi_n|_A$ et $f \circ \varphi_n|_B$ convergent vers l'identité. Par compacité, quitte à extraire, $f \circ \varphi_n$ converge vers un morphisme f' de G dans un groupe limite. Ce morphisme f' est injectif sur les groupes de sommet du produit amalgamé ou de l'extension HNN, qui coïncident avec leurs voisinages abéliens. Si e sépare, on peut supposer que les injections de G'_e dans A et B sont propres, donc que ni A ni B ne sont abéliens. Si e ne sépare pas, et si A est abélien, alors G a pour présentation $\langle A, t \mid tat^{-1} = \varphi(a) \forall a \in A \rangle$, avec $\varphi : A \rightarrow A$ un isomorphisme. Par un argument déjà vu, comme $f'(tat^{-1})$ et $f'(a)$ commutent, $f'(t)$ et $f'(a)$ commutent, et comme f' est injective sur A , l'isomorphisme φ est l'identité. Donc t centralise A , ce qui contredit le fait que A est égal à son centralisateur dans G . On conclut alors par le cas $n = 1$: si φ'_n est une suite de twists de Dehn sur e telle que $f' \circ \varphi'_n$ converge vers l'identité, alors $f \circ \varphi_n \circ \varphi'_n$ aussi.

L'étape 4 implique que G est un groupe limite et conclut la preuve de la partie directe du théorème 6.1.

Montrons maintenant la partie réciproque du théorème 6.1. Soit G_0 un groupe limite, et S_0 une partie génératrice fixée de G_0 . Montrons que G_0 admet une résolution de Makanin-Razborov. En utilisant des décompositions de Grushko et le résultat de finitude 5.6 en conjonction avec le théorème 5.10, il suffit donc de montrer que si (G, S) est un groupe limite marqué librement indécomposable, alors il existe un quotient raccourcissant maximal (G'', S'') de (G, S) tel que le morphisme de groupes $G \rightarrow G''$ est une résolution de Makanin-Razborov élémentaire.

On utilise l'identification $\mathbf{GM}(G, S) = \text{Epi}(G, \cdot)$, voir la partie 5.2. Soit F un groupe libre et $(f_n : G \rightarrow F)_{n \in \mathbb{N}}$ une suite de morphismes surjectifs qui converge vers l'identité (qui existe comme G est un groupe limite). Pour tout n dans \mathbb{N} , soit φ_n dans $\text{Mod}(G)$ tel que $f_n \circ \varphi_n$ est (G, S) -court. Alors, quitte à extraire, la suite $f_n \circ \varphi_n$ converge vers un quotient raccourcissant $f' : G \rightarrow G'$. Celui-ci factorise à travers un quotient raccourcissant maximal $f'' : G \rightarrow G''$. Montrons que f'' est une résolution de Makanin-Razborov élémentaire, en vérifiant les propriétés (1)-(5) de leur définition. Remarquons que si φ_n vaut une conjugaison sur un élément h non trivial de G pour tout n , alors $f_n(h) \neq 1$ pour n assez grand, donc $f_n \circ \varphi_n(h) \neq 1$, et donc $f'(h) \neq 1$, d'où $f''(h) \neq 1$.

Soit (X, G_*) la décomposition cyclique de G donnée par la proposition 4.2, qui vérifie que le centralisateur dans G d'un groupe d'arête est contenu dans l'un de ses groupes de sommet. Comme G est commutatif-transitif (corollaire 3.2), la propriété (1) est vérifiée. Comme φ_n vaut une conjugaison sur chaque groupe d'arête H , sur chaque groupe de sommet H non abélien, non de type surface, ainsi que sur chaque sous-groupe H de groupe de sommet abélien engendré par les stabilisateurs d'arêtes d'origine ce sommet, le morphisme f'' est injectif sur ces groupes H , et les propriétés (2) et (4) sont vérifiées.

Soit G_v un groupe de sommet non abélien, non de type surface, et e une arête d'origine v . Soit h_n dans G tel que φ_n agit par la conjugaison par h_n sur G_v . Notons (X', G'_*) le graphe de groupes obtenu en remplaçant chaque groupe de sommet abélien par son sous-groupe engendré par les groupes d'arête d'origine ce sommet, et G' son groupe fondamental (d'origine v). Montrons que φ_n agit aussi par la conjugaison par h_n sur le centralisateur Z_e dans $G'_{t(e)}$ de $\rho_e(G_e)$. On peut supposer que le centralisateur de $\rho_e(G_e)$ dans G n'est pas contenu dans G_v , donc par les propriétés (1) et (2), on a $Z_G(G_e) = Z_{G'_{t(e)}}(\rho_e(G_e))$. Si $G'_{t(e)}$ est de type surface, alors $Z_{G'_{t(e)}}(\rho_e(G_e))$ est réduit à $\rho_e(G_e)$, donc il n'y a rien à montrer. Si $G'_{t(e)}$ est abélien, alors $Z_e = G'_{t(e)}$, et φ_n agit par la conjugaison par un élément h'_n sur Z_e . De même, si $G'_{t(e)}$ n'est ni abélien, ni de type surface, alors φ_n agit par la conjugaison par un élément h'_n sur $G'_{t(e)}$, donc sur Z_e (que e soit une boucle en v ou non). Maintenant, si a est un élément non trivial de G_e , alors $h_n^{-1}h'_n$ centralise a , donc appartient à $G'_{t(e)}$. Donc φ_n agit aussi par la conjugaison par h_n sur Z_e , et ceci étant

vrai pour tout e , φ_n agit par la conjugaison par h_n sur le voisinage abélien de G_v dans (X', G'_*) . Ceci montre que f'' est injectif sur ce voisinage abélien.

Par définition, un groupe de sommet de type surface est non abélien, sauf s'il y a un unique sommet dans X de groupe \mathbf{Z}^2 , auquel cas le résultat est clair. Comme une limite de groupes marqués non abéliens est non abélien, l'image par f' (donc par f'') d'un groupe de sommet de type surface est non abélien, ce qui montre l'assertion (3). \triangleright

Les résultats suivants découlent alors de la preuve du théorème 6.1.

COROLLAIRE 6.6 (Kharlampovich-Myasnikov-Remeslennikov [KM2] Theorem 6). — *Un groupe est un groupe limite si et seulement s'il appartient à la collection \mathbf{C} de groupes définie récursivement par*

- *un groupe libre de type fini est dans \mathbf{C} ;*
- *un produit libre de deux groupes dans \mathbf{C} est dans \mathbf{C} ;*
- *une extension libre de centralisateurs d'un élément de \mathbf{C} est dans \mathbf{C} ;*
- *un produit amalgamé de deux groupes dans \mathbf{C} , au-dessus d'un groupe cyclique qui est abélien maximal dans au moins l'un de ces deux groupes, muni d'un morphisme à valeurs dans un élément de \mathbf{C} , injectif sur le voisinage abélien des groupes de sommet, est dans \mathbf{C} ;*
- *une extension HNN acylindrique (i.e. de la forme $A*_C$ avec $\rho_e(C)a^{-1} \cap \rho_{\bar{e}}(C) = \{1\}$ pour tout a dans A), d'un groupe A dans \mathbf{C} au-dessus d'un groupe abélien dont l'une des deux images dans A est abélienne maximale, munie d'un morphisme à valeurs dans un élément de \mathbf{C} , injectif sur le voisinage abélien du groupe de sommet, est dans \mathbf{C} .* \triangleright

COROLLAIRE 6.7. — *Pour tout groupe limite librement indécomposable marqué (G, S) , il existe un quotient raccourcissant $f : (G, S) \rightarrow (G', S')$ et une suite $(\varphi_n)_{n \in \mathbf{N}}$ dans $\text{Mod}(G)$ telle que $f \circ \varphi_n$ converge vers l'identité.*

6.1. Tour multi-résiduellement libre hyperbolique

Rappelons que si X, Y sont des CW-complexes, A un sous-CW-complexe de X , et $f : A \rightarrow Y$ une application cellulaire, on appelle *recollement de X sur Y par f* le CW-complexe quotient $X \amalg_f Y = (X \amalg Y) / \sim$, avec \sim la relation d'équivalence engendrée par $x \sim f(x)$ pour tout x dans A . La projection canonique $Y \rightarrow X \amalg_f Y$ est un homéomorphisme sur son image, et on identifie Y avec celle-ci (voir par exemple [Hat]). On parle de *somme pointée* de X et Y lorsque A est un singleton.

Un CW-complexe X est dit *multi-résiduellement libre de niveau $\leq n$* si $n = 0$ et X est un point, ou si $n \geq 1$ et si X est un CW-complexe obtenu, à partir d'un CW-complexe Y multi-résiduellement libre de niveau $\leq n - 1$, par l'une des opérations suivantes:

- (1) (extension libre) X est une somme pointée de Y et d'un bouquet de cercles ;

- (2) (extension de type surface) X est un recollement $S \amalg_f Y$ sur Y d'une surface S compacte connexe à bord non vide, de caractéristique d'Euler au plus -2 ou égale au tore troué ou à la bouteille de Klein trouée, avec $f : \partial S \rightarrow Y$ une application cellulaire, envoyant chaque composante connexe de ∂S sur un lacet non homotope à zéro, et telle qu'il existe une rétraction de X sur Y , dont l'application induite sur les groupes fondamentaux envoie $\pi_1 S$ sur un groupe non abélien ;
- (3) (extension abélienne) X est un recollement $(\mathbf{T}^n \times \mathbf{T}^m) \amalg_f Y$, avec $\mathbf{T}^k = \mathbf{R}^k / \mathbf{Z}^k$ le tore de dimension k , ayant 0 pour sommet, et $f : \{0\} \times \mathbf{T}^m \rightarrow Y$ une application cellulaire, induisant sur les groupes fondamentaux un isomorphisme sur un sous-groupe abélien maximal de $\pi_1 Y$.

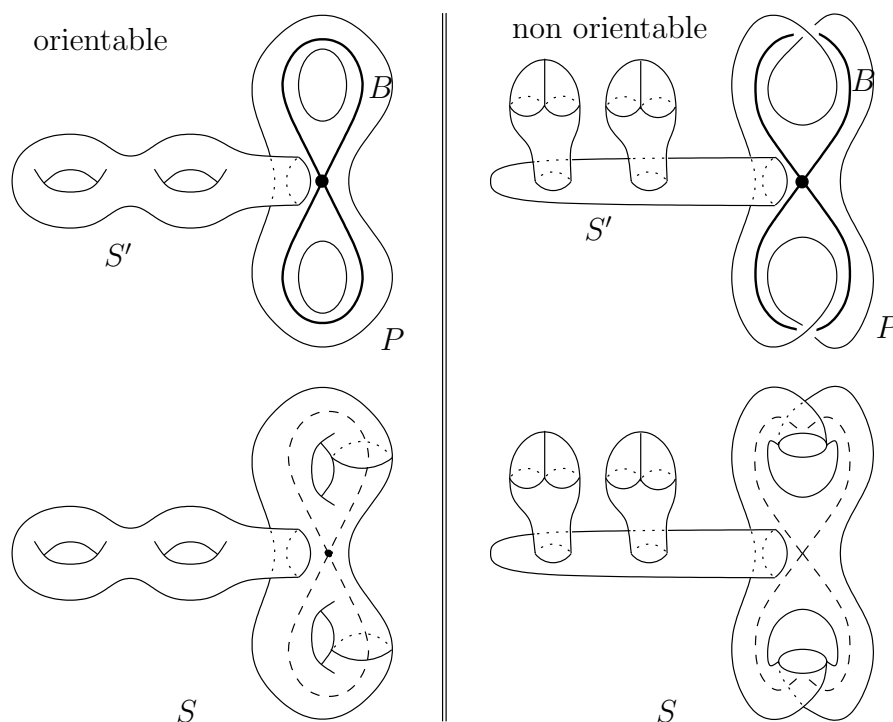
Un CW-complexe X est dit *multi-résiduellement libre* s'il existe n dans \mathbf{N} tel que X est multi-résiduellement libre de niveau $\leq n$. Une *tour de Sela* est le groupe fondamental d'un CW-complexe multi-résiduellement libre.

Il est facile de voir qu'une tour de Sela admet une résolution de Makanin-Razborov, donc est un groupe limite (voir le théorème 6.1). Une tour de Sela est *hyperbolique* si elle est construite sans aucune extension abélienne, ou, de manière équivalente, si elle n'a pas de sous-groupe abélien non cyclique. Une tour de Sela hyperbolique est hyperbolique au sens de Gromov (voir le corollaire 5.4). Voir les «regular NTQ groups» de [KM1]-[KM5].

Par exemple, un groupe de surface, non homéomorphe à la somme connexe de 1, 2 ou 3 plans projectifs, est une tour hyperbolique. En effet, soit S une surface compacte connexe de genre g , avec $g \geq 2$ si S est orientable, et $g \geq 4$ sinon. On considère une surface compacte connexe S' qui est

- orientable de genre $g - 2$ si S est orientable,
- une sphère si S est non orientable et $g = 4$,
- non orientable de genre $g - 4$ sinon.

Soit P un pantalon si S est orientable, et la bouteille de Klein trouée sinon. Soit S'' la somme connexe $S' \# P$. Alors il existe une rétraction $f : S'' \rightarrow B$ dans le bouquet de deux cercles B , plongé dans P . L'espace topologique X recollement de S'' sur B par la restriction de f au bord de S'' , admet une structure de CW-complexe multi-résiduellement libre de niveau ≤ 2 (extension de type surface d'une extension libre), et X est homéomorphe à S . La rétraction f de S'' sur B induit par passage au quotient une rétraction r de X sur B . Au niveau des groupes fondamentaux, l'application r induit une surjection de $\pi_1 S''$ sur $\pi_1 B$, qui est un groupe libre de rang 2 (voir dessin ci-dessous).



Dans le cas orientable, on peut aussi considérer, pour tout entier $g \geq 2$, une surface compacte connexe planaire P' de caractéristique d'Euler $1 - g$ (i.e. homéomorphe à un disque avec g disques ouverts enlevés). Alors P' se rétracte sur un bouquet B' de g cercles plongé dans P' . Le double de P' le long de son bord est une surface compacte connexe orientable de genre g , qui se rétracte évidemment sur le bouquet de cercles B' contenu dans l'une des deux copies de P' .

La partie facile suivante de la caractérisation des groupes de type fini élémentairement équivalents à un groupe libre de type fini non cyclique découle des techniques utilisées dans la construction des résolutions de Makanin-Razborov. En effet, un groupe de type fini G élémentairement équivalent à un groupe libre de type fini non cyclique est en particulier un groupe limite, donc admet une résolution de Makanin-Razborov. En utilisant des formules closes $\forall\exists$, il n'est pas très difficile d'améliorer une telle résolution en une tour hyperbolique, voir [Sel6].

THÉORÈME 6.8 (Sela [Sel6] Prop. 6). — *Soit G un groupe de type fini élémentairement équivalent à un groupe libre de type fini non cyclique. Alors G est isomorphe à une tour de Sela hyperbolique.* ▷

Mais la réciproque (qui contient la solution du problème de Tarski) est beaucoup plus difficile, et nécessite de nombreux autres outils.

BIBLIOGRAPHIE

- [Bas] H. Bass, *Covering theory for graphs of groups*, J. Pure Appl. Math. **89** (1993) 3-47.
- [Bau1] G. Baumslag, *On generalized free products*, Math. Z. **78** (1962) 423-438.
- [Bau2] G. Baumslag, *Topics in combinatorial group theory*, Lect. in Math. Birkhäuser, 1993.
- [BMR] G. Baumslag, A. Myasnikov, V. N. Remeslennikov, *Algebraic geometry over groups I. Algebraic sets and ideal theory.*, J. Algebra **219** (1999) 16-79.
- [BF1] M. Bestvina, M. Feighn, *Bounding the complexity of simplicial group actions on trees*, Inv. Math. **103** (1991) 449-469.
- [BF2] M. Bestvina, M. Feighn, *A combination theorem for negatively curved groups*, J. Diff. Geom. **35** (1992), 85-101, Addendum **43** (1996), 783-788.
- [BF3] M. Bestvina, M. Feighn, *Stable actions of groups on real trees*, Inv. Math. **121** (1995) 287-321.
- [Bou] N. Bourbaki, *Topologie générale*, chap. 1 à 4, Hermann, Paris, 1971.
- [Bow] B. Bowditch, *Cut points and canonical splittings of hyperbolic groups*, Acta Math. **180** (1998), 145-186.
- [Cham] C. Champetier, *L'espace des groupes de type fini*, Topology **39** (2000) 657-680.
- [CG] C. Champetier, V. Guirardel, *Limit groups as limits of free groups: compactifying the set of free groups*, en préparation.
- [CK] C.C. Chang, H.J. Keisler, *Model theory*, Stud. Log. **73**, North-Holand, 1973.
- [Chat] Z. Chatzidakis, *Limit groups, viewed by a logician*, Notes d'exposés, 2001 (<http://www.logique.jussieu.fr/~zoe>).
- [Chi] I. Chiswell, *Introduction to Λ -trees*, World Scientific, 2001.
- [Dun] M.J. Dunwoody, *Groups acting on protrees*, J. Lond. Math. Soc., **56** (1997) 125-136.
- [DS] M. J. Dunwoody, M. E. Sageev, *JSJ-splittings for finitely presented groups over slender groups*, Inv. Math. **135** (1999) 25-44.
- [For] M. Forester, *On uniqueness of JSJ decompositions of finitely generated groups*, à paraître dans Comm. Math. Helv.
- [FP] K. Fujiwara, P. Papasoglu, *JSJ-splittings and complexes of groups*, prépublication Orsay, 1998.
- [GLP] D. Gaboriau, G. Levitt, F. Paulin, *Pseudogroups of isometries of \mathbb{R} and Rips' theorem on free actions on \mathbb{R} -trees*, Israel J. Math. **87** (1994), 403-428.
- [Ghy1] É. Ghys, *Les groupes hyperboliques*, Sémin. Bourbaki (1989/90), Exp. n° 722, Astérisque **189-190** (1990) 203-238.
- [Ghy2] É. Ghys, *Les groupes aléatoires [d'après Misha Gromov, ...]*, Sémin. Bourbaki (2002/03), mars 2003, Exp. n° 916.
- [Gri] R. I. Grigorchuk, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk. SSSR **48** (1984) 939-985.
- [Gro] S. Gross, PhD Thesis, The Hebrew University, 2001.
- [Gui] V. Guirardel, *Limit groups and groups acting on \mathbb{R}^n -trees*, prépublication Univ. Toulouse (2003).
- [HV] P. de la Harpe, A. Valette, *La propriété T de Kazhdan pour les groupes localement compacts*, Astérisque **175**, Soc. Math. France (1989).
- [Hat] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, 2002, <http://www.math.cornell.edu/~hatcher>.
- [HS] H. Hendriks, A. Shastri, *A splitting theorem for surfaces*, dans "Topological structures, II" (Proc. Symp. Topo. Geom., Amsterdam, 1978), Part 1, pp. 117-121, Math. Centre Tracts **115**, Math. Centrum, Amsterdam, 1979.

- [JS] W. H. Jaco, P. B. Shalen, *Seifert fibered spaces in 3-manifolds*, Memoirs **220**, Amer. Math. Soc. 1979.
- [Joh] K. Johannson, *Homotopy equivalences of 3-manifolds with boundaries*, Lect. Notes in Math. **761**, Springer Verlag, 1979.
- [KM0] O. Kharlampovich, A. Myasnikov, *Tarski's problem about the elementary theory of free groups has a positive solution*, Elec. Res. Ann. Amer. Math. Soc. **4** (1998) 101-108.
- [KM1] O. Kharlampovich, A. Myasnikov, *Irreducible affine varieties over a free group. I. Irreducibility of quadratic equations and Nullstellensatz*, J. Algebra **200** (1998) 472-516.
- [KM2] O. Kharlampovich, A. Myasnikov, *Irreducible affine varieties over a free group. II. Systems in triangular quasi-quadratic form and description of residually free groups*, J. Algebra **200** (1998) 517-570.
- [KM3] O. Kharlampovich, A. Myasnikov, *Implicit function theorem over free groups*, prépublication 1999, révisée 2003, <http://www.math.mcgill.ca/~olga>.
- [KM4] O. Kharlampovich, *Equations over fully residually free groups*, prépublication 1999, révisée 2003, <http://www.math.mcgill.ca/~olga>.
- [KM5] O. Kharlampovich, A. Myasnikov, *Elementary theory of free nonabelian groups*, prépublication 1999, révisée 2003, <http://www.math.mcgill.ca/~olga>.
- [Lev] G. Levitt, *Automorphisms of hyperbolic groups and graphs of groups*, prépublication Univ. Toulouse, Oct. 2002.
- [Lyn1] R. C. Lyndon, *The equation $a^2b^2 = c^2$ in free groups*, Mich. Math. J. **6** (1959) 89-95.
- [Lyn2] R. C. Lyndon, *Equations in free groups*, Trans. Amer. Math. Soc. **96** (1960) 445-457.
- [LS] R. C. Lyndon, P. E. Schupp, *Combinatorial group theory*, Ergeb. Math. Grenz. **89**, Springer Verlag 1977.
- [Mak1] G. S. Makanin, *Equations in a free group*, Math. USSR Izv. **21** (1983) 449-469.
- [Mak2] G. S. Makanin, *Decidability of the universal and positive theories of a free group*, Math. USSR Izv. **25** (1985) 75-88.
- [Mer] Yu. I. Merzlyakov, *Positive formula on free groups*, Algebra i Logika **5** (1966) 25-42.
- [Pau1] F. Paulin, *Topologie de Gromov équivariante, structures hyperboliques et arbres réels*, Inv. Math. **94** (1988), 53-80.
- [Pau2] F. Paulin, *Actions de groupes sur les arbres*, Sémin. Bourbaki (1995/96), Exp. n° 808, Astérisque **241** (1997) 97-137.
- [Raz] A. A. Razborov, *On systems of equations in a free group*, Math. USSR Izv. **25** (1985) 115-162.
- [Rem] V. N. Remeslennikov, *\exists -free groups*, Siberian Math. J. **30** (1989) 998-1001.
- [RS] E. Rips, Z. Sela, *Cyclic splittings of finitely presented groups and the canonical JSJ decomposition*, Ann. Math. **146** (1997) 53-104.
- [Sac] G. S. Sacerdote, *Elementary properties of free groups*, Trans. Amer. Math. Soc. **178** (1973) 127-138.
- [Sel1] Z. Sela, *Diophantine geometry over groups I: Makanin-Razborov diagrams*, Pub. Math. IHÉS **93** (2001) 31-105.
- [Sel2] Z. Sela, *Diophantine geometry over groups II: completions, closures and formal solutions*, Israel J. Math. **134** (2003), 173-254.
- [Sel3] Z. Sela, *Diophantine geometry over groups III: rigid and solid solutions*, prépublication 2001, 49 pages, <http://www.ma.huji.ac.il/~zlil>.
- [Sel4] Z. Sela, *Diophantine geometry over groups IV: an iterative procedure for validation of a sentence*, prépublication Jul. 2001, 106 pages, <http://www.ma.huji.ac.il/~zlil>.

- [Sel5] Z. Sela, *Diophantine geometry over groups V: quantifier elimination*, prépublication Oct. 2001, 223 pages, <http://www.ma.huji.ac.il/~zlil>.
- [Sel6] Z. Sela, *Diophantine geometry over groups VI: the elementary theory of a free group*, prépublication Oct. 2001, 14 pages, <http://www.ma.huji.ac.il/~zlil>.
- [Sel7] Z. Sela, *Diophantine geometry over groups VII: the elementary theory of a hyperbolic group*, prépublication Mars 2002, 65 pages, <http://www.ma.huji.ac.il/~zlil>.
- [Ser] J.-P. Serre, *Arbres, amalgames, SL_2* , Astérisque **46**, Soc. Math. France (1983).
- [Tar] A. Tarski, *Some notions and methods on the borderline of algebra and metamathematics*, dans “Proc. I.C.M. Cambridge 1950”, vol. 1, pp. 705–720, Amer. Math. Soc. (1952).
- [ZVC] H. Zieschang, E. Vogt, H. Coldewey, *Surfaces and planar discontinuous groups*, Lect. Notes **835**, Springer Verlag (1980).

Frédéric PAULIN

DMA, UMR 8553 CNRS

Ecole Normale Supérieure

45 rue d’Ulm

75230 PARIS Cedex 05, FRANCE

e-mail: Frederic.Paulin@ens.fr

GROUPES DE GALOIS DE CORPS DE TYPE FINI
[d'après Pop]

par **Tamás SZAMUELY**

Le but de cet exposé est de montrer que le groupe de Galois absolu d'un corps de fonctions est un invariant très fin : d'une part, il détermine le corps de fonctions à isomorphisme près ; d'autre part, il est possible d'en extraire des informations très riches et variées sur l'arithmétique et la géométrie du corps en question.

1. ÉNONCÉS

Les deux résultats principaux qui nous intéressent ici sont les suivants. Ici, et dans la suite, pour un corps F on fixe toujours une clôture séparable F^s , et on note G_F le groupe de Galois absolu $\text{Gal}(F^s|F)$.

THÉORÈME 1.1 (Florian Pop, [30], [31]). — *Soient K, L deux corps infinis, de type fini sur le corps premier. Supposons qu'il existe un isomorphisme $\Phi : G_K \xrightarrow{\sim} G_L$ de groupes profinis. Alors il existe des extensions purement inséparables $K'|K, L'|L$ avec $K' \cong L'$. De plus, il existe un isomorphisme $\phi : L^s \xrightarrow{\sim} K'^s$ de clôtures séparables tel que pour tout élément $g \in G_{K'}$ on ait $\Phi(g) = \phi^{-1} \circ g \circ \phi$.*

REMARQUES 1.2. —

- (1) En particulier, tout corps de type fini sur \mathbf{Q} est déterminé à isomorphisme près par son groupe de Galois absolu. Un tel énoncé ne vaut pas en caractéristique positive, car si $K'|K$ est une extension purement inséparable de corps, alors l'homomorphisme de restriction $G_K \rightarrow G_{K'}$ est un isomorphisme. La formulation ci-dessus tient compte de ce contre-exemple évident.
- (2) Pop démontre en fait un énoncé plus précis : pour L, K comme ci-dessus, considérons la règle qui associe à tout isomorphisme $\phi : L^i \xrightarrow{\sim} K^i$ entre leurs clôtures parfaites l'isomorphisme $\Phi : G_K \xrightarrow{\sim} G_L$ donné par la formule $\Phi(g) = \phi^{-1} \circ g \circ \phi$. En caractéristique 0, c'est une bijection entre l'ensemble des isomorphismes de corps $L \xrightarrow{\sim} K$ et le *quotient* de l'ensemble des isomorphismes de groupes profinis $G_K \xrightarrow{\sim} G_L$ par l'action intérieure naturelle de G_L . En caractéristique positive on a une bijection

analogue, mais il faut y identifier les isomorphismes $L^i \xrightarrow{\sim} K^i$ qui sont les mêmes « quitte à tordre par un automorphisme de Frobenius ».

Dans un travail en préparation, Pop obtient l'amplification remarquable que voici. Pour un groupe profini G et un nombre premier ℓ , notons G^ℓ le pro- ℓ -quotient maximal de G .

THÉORÈME 1.3 (Florian Pop, [32], [33]). — *Soient ℓ un nombre premier, K et L deux corps de type fini sur la clôture algébrique du corps premier dont le degré de transcendance est au moins 2 et dont la caractéristique est différente de ℓ . Supposons qu'il existe un isomorphisme $\Phi : G_K^\ell \xrightarrow{\sim} G_L^\ell$ de groupes profinis. Alors il existe des extensions purement inséparables $K'|K, L'|L$ avec $K' \cong L'$. De plus, il existe un isomorphisme $\phi : L'^s \xrightarrow{\sim} K'^s$ de clôtures séparables tel que pour tout élément $g \in G_K^\ell$ on ait $\Phi(g) = \phi^{-1} \circ g \circ \phi$.*

REMARQUES 1.4. —

- (1) Bien entendu, on a encore un énoncé plus précis comme dans la remarque ci-dessus.
- (2) L'énoncé est *faux* pour les corps de degré de transcendance un. En effet, si K est le corps de fonctions de n'importe quelle courbe propre lisse X de genre $g > 0$ sur un corps algébriquement clos, et ℓ est premier à la caractéristique de K , alors G_K^ℓ est la limite projective des groupes fondamentaux ℓ -complétés $\pi_1(U, u)^\ell$ pour les sous-schémas ouverts $U \subset X$. Or la présentation d'un tel $\pi_1(U, u)^\ell$ est bien connue d'après les travaux de Riemann, Poincaré et Grothendieck (voir [15]) : elle est de la forme $\langle a_1, b_1, \dots, a_g, b_g, i_1, \dots, i_r \mid [a_1, b_1] \dots [a_g, b_g] i_1 \dots i_r = 1 \rangle$, où r est le nombre des points de $X \setminus U$. Ainsi, G_K^ℓ ne dépend que de g et du cardinal du corps de base.
- (3) Le th. 1.1 ne semble pas découler de manière évidente du th. 1.3. Toutefois, après avoir fait une bonne partie de la démonstration du th. 1.1, on peut conclure par une application de 1.3 (voir la remarque 8.6).

2. HISTORIQUE

Bien que la caractérisation galoisienne des corps réels clos par Artin et Schreier [2] puisse être regardée comme un précurseur des résultats ci-dessus, l'histoire commence réellement avec des travaux de Neukirch à la fin des années 1960. Celui-ci a été amené par son étude de la structure du groupe de Galois absolu d'un corps p -adique (un sujet de recherche florissant à l'époque) à une caractérisation cohomologique des sous-groupes de décomposition associés aux idéaux premiers dans une extension galoisienne de corps de nombres. De là il a pu déduire, par une application du théorème de densité de Chebotarev, le th. 1.1 dans le cas où L et K sont des extensions *finies galoisiennes* de \mathbf{Q} (voir [24]). En plus, il a remarqué que l'on pourrait étendre son résultat au cas de corps de nombres arbitraires via la solution d'un « problème de plongement » en théorie de Galois inverse.

Ce travail a été mené à bien par Iwasawa (non publié) et Uchida [39] (voir aussi [26], [27]). Ce dernier a également étendu dans [40] la méthode de Neukirch au cas des corps de fonctions à une variable sur un corps fini.

L'histoire semblait donc arriver à une conclusion satisfaisante, mais l'intérêt pour ce genre de questions a été réalimenté vers le milieu des années 1980 par la « vision anabélienne » de Grothendieck, exposée dans sa célèbre lettre à Faltings [17]. Ce programme, qui vise la reconstitution de certains schémas définis sur des corps de type fini (dits *schémas anabéliens*) ainsi que leurs morphismes dominants à partir de leur groupe fondamental géométrique muni de l'action extérieure du groupe de Galois du corps de base, a donné lieu à de spectaculaires développements ces dernières années dont la plupart ont été rapportés dans ce séminaire par Faltings [13] (voir aussi les survols [23], [27], [29], [38] et [18]). Grothendieck énonce dans sa lettre le th. 1.1 comme une variante birationnelle de ses conjectures sur les groupes fondamentaux, avec toutefois une différence notable : il considère les groupes de Galois des corps de fonctions munis d'augmentations naturelles vers le groupe de Galois absolu d'un corps de base k (qui peut être le corps premier, supposé le même pour K et L , ou plus généralement un corps de type fini sur le corps premier contenu dans chacun de ces deux corps) et par suite les k -isomorphismes entre K et L . En revanche, il prédit que les k -homomorphismes $K \rightarrow L$ devraient également être reconstituables à partir des G_k -morphisms induits sur les groupes de Galois. Or le th. 1.1 et surtout le th. 1.3 montrent que du moins en ce qui concerne la reconstruction des isomorphismes, l'action du groupe de Galois d'un corps de base n'est pas essentielle. Les travaux récents de Tamagawa, Raynaud et autres sur les groupes fondamentaux des courbes sur un corps algébriquement clos de caractéristique positive témoignent d'ailleurs du même phénomène.

Concernant la conjecture birationnelle de Grothendieck, le premier pas est franchi par Pop [28], où le th. 1.1 est démontré dans le cas des corps de fonctions à une variable sur un corps de nombres, en combinant la stratégie de Neukirch avec des considérations en théorie des valuations utilisant des techniques de logique. Spiess [37] a ensuite trouvé une démonstration purement arithmético-géométrique encore plus proche de l'esprit original de Neukirch. En combinant ces idées avec des arguments de géométrie birationnelle, Pop a ensuite réussi à démontrer le th. 1.1.

Entretiens, dans ses articles [4], [5], Bogomolov a esquissé un programme pour démontrer un résultat similaire au th. 1.3, mais pour des corps de type fini au-dessus d'un corps algébriquement clos quelconque. D'après son point de vue, le corps de fonctions K devrait être déterminé par le deuxième quotient $G_K^\ell / [[G_K^\ell, G_K^\ell], G_K^\ell]$ de la suite centrale descendante de G_K^ℓ . Dans leurs articles récents [6], [7], Bogomolov et Tschinkel fournissent des détails de cette démonstration dans le cas des corps de fonctions de certaines surfaces au-dessus de $\bar{\mathbf{F}}_p$. Certaines idées de Bogomolov ont inspiré Pop dans sa démonstration du théorème 1.3. Pour un complément intéressant, voir [34].

Cet aperçu ne serait pas complet si on ne mentionnait pas un résultat très important de Mochizuki [22] : il y est démontré que si K et L sont deux corps de type fini sur un corps k qui peut lui-même être plongé dans une extension de type fini de \mathbf{Q}_p (Mochizuki appelle un tel corps *sous- p -adique*), alors tout G_k -morphisme $G_K \rightarrow G_L$ de groupes profinis à image ouverte est induit par un k -morphisme $L \rightarrow K$ de corps. Mochizuki dérive ce théorème de son résultat anabélien fondamental d'après lequel tout G_k -morphisme $\pi_1(X) \rightarrow \pi_1(Y)$ à image ouverte entre les groupes fondamentaux de deux k -courbes hyperboliques X et Y est induit par un k -morphisme dominant $X \rightarrow Y$ (voir l'exposé [13] de Faltings). La stratégie est alors la suivante (pour les détails, cf. [22], §§15, 16) : on montre d'abord que si K est le corps de fonctions de X , alors tout G_k -morphisme $G_K \rightarrow \pi_1(Y)$ se factorise à travers $\pi_1(X)$, d'où on conclut aisément en utilisant le résultat précédent qu'il est induit par un morphisme $\text{Spec } K \rightarrow Y$. Un argument (non trivial) de fibration en courbes permet alors d'en déduire un résultat similaire pour K corps de type fini quelconque sur k . Ensuite, étant donné un G_k -morphisme $G_K \rightarrow G_L$, où L est le corps de fonctions de Y , on peut le composer par le G_k -morphisme naturel $G_L \rightarrow \pi_1(Y)$ et ainsi récupérer par ce qui précède le morphisme dominant $\text{Spec } K \rightarrow Y$; il se factorise à travers $\text{Spec } L$. Enfin, le résultat pour L général s'en suit à nouveau par un argument de fibration (facile cette fois).

Noter que dans cette preuve l'action de G_k joue un rôle essentiel ; la variante « absolue » du théorème, ainsi que son analogue en caractéristique positive, ne sont pas connues à ce jour.

3. STRATÉGIE

Nous donnons maintenant un survol parallèle des démonstrations des théorèmes 1.1 et 1.3, laissant les détails aux chapitres suivants. Nous ignorons ici les complications causées par les extensions inséparables en caractéristique positive.

On va appeler la situation du th. 1.1 le *cas arithmétique* et celle du th. 1.3 le *cas géométrique*. Le corps de base k est le corps premier de K dans le premier cas et sa clôture algébrique dans le second ; de même pour l et L .

Première étape : Correspondance locale. Soit X un k -schéma normal intègre de type fini ayant K pour corps de fonctions, et soit P un point de codimension 1 de X . Notons D_P un sous-groupe de décomposition de G_K (resp. G_K^ℓ) associé à P , i.e. le stabilisateur d'un point \bar{P} au-dessus de P du normalisé de X dans la clôture algébrique \bar{K} de K . Alors on montre que $D_Q := \Phi(D_P)$ est un groupe de décomposition associé à un point Q de codimension 1 d'un k -schéma normal intègre de type fini Y ayant L pour corps de fonctions. En plus, on obtient ainsi une bijection entre les sous-groupes de décomposition de G_K et de G_L (resp. leurs pro- ℓ -quotients) du type décrit ci-dessus.

On démontre ensuite que Φ transforme le sous-groupe d'inertie de D_P en celui de D_Q , d'où un isomorphisme $\Phi_P : G_{k(P)} \rightarrow G_{k(Q)}$ entre les groupes de Galois des corps résiduels de P et de Q (resp. leurs pro- ℓ -quotients).

Les résultats locaux esquissés ci-dessus permettent déjà de reconstituer beaucoup d'invariants arithmétiques et géométriques des corps K et L . En particulier, dans le cas arithmétique, l'utilisation des résultats précédents permet d'établir que les corps de base $k \subset K$ et $l \subset L$ sont isomorphes, et que Φ transforme la projection canonique $G_K \rightarrow G_k$ en la projection $G_L \rightarrow G_l$; par le théorème de Neukirch/Uchida, on a donc $k \cong l$. Dans le cas géométrique ces deux projections sont bien entendu triviales.

Une fois la projection $G_K \rightarrow G_k$ (ou même la projection $G_K^\ell \rightarrow G_k^\ell$) connue, il est possible de déduire le cas arithmétique du cas géométrique (cf. la remarque 8.6). On va opter pour ce procédé en caractéristique positive; en revanche, en caractéristique 0 l'approche directe au th. 1.1 est plus simple.

Deuxième étape : Correspondance kummérienne. Soit maintenant n un entier positif que l'on suppose toujours premier à la caractéristique de K et de L ; en plus, dans le cas géométrique, on suppose que n est une puissance de ℓ .

Dans le cas arithmétique, le groupe G_K agit sur le groupe μ_n des racines n -ièmes de l'unité à travers son quotient G_k qui est préservé par Φ par ce qui précède; dans le cas géométrique cette action est triviale. Donc, par application de la théorie de Kummer, on obtient une suite d'isomorphismes

$$K^\times / K^{\times n} \cong H^1(G_K, \mu_n) \cong H^1(G_L, \mu_n) \cong L^\times / L^{\times n}$$

où K^\times (resp. L^\times) est le groupe multiplicatif de K (resp. L), les deux groupes au milieu sont des groupes de cohomologie galoisienne et leur isomorphisme est induit par Φ . Notons ici que Φ est également applicable dans le cas géométrique : comme l'action de G_K sur μ_n est triviale, les éléments de $H^1(G_K, \mu_n)$ sont alors des homomorphismes $G_K \rightarrow \mu_n$; ils se factorisent à travers G_K^ℓ car on a supposé que n est une puissance de ℓ .

Par passage à la limite projective sur les puissances de ℓ , on obtient également un isomorphisme

$$\Phi^{(\ell)} : K^{\times(\ell)} \cong L^{\times(\ell)}$$

entre les complétés ℓ -adiques des groupes multiplicatifs des deux corps.

Le noyau du morphisme $K^\times \rightarrow K^{(\ell)}$ est le sous-groupe ℓ -divisible maximal de K^\times . Dans le cas arithmétique on peut montrer que ce sous-groupe est trivial; en revanche, dans le cas géométrique ce sous-groupe est égal à k^\times .

Troisième étape : Récurrence. Dans le cas arithmétique, on dispose des résultats de Neukirch et Uchida sur les corps globaux classiques. Il est donc possible de les utiliser comme point de départ d'une récurrence sur le degré de transcendance commun de K et L (en effet, ce degré est lisible sur la dimension cohomologique de $G_K \cong G_L$). En particulier, on peut supposer que les corps résiduels de deux points de codimension 1 P

et Q sur des modèles de K et L qui se correspondent par la bijection de la première étape sont isomorphes.

La combinaison de ces deux types d'information, à savoir les « isomorphismes modulo P » ainsi que l'isomorphisme des groupes multiplicatifs modulo n établi à l'étape précédente permet alors de construire un isomorphisme entre K et L par « recollement », du moins en caractéristique zéro. Notons toutefois que l'argument contient plusieurs détails délicats, comme par exemple l'établissement d'une correspondance canonique entre les modèles géométriques suffisamment petits de K et de L (cf. le chap. 5).

En revanche, dans le cas géométrique une pareille récurrence n'est *a priori* possible qu'à partir du cas de degré de transcendance 2 (cf. la remarque 1.4 (2)), que l'on ne connaît pas à l'avance. Toutefois, la théorie locale fournit suffisamment d'informations sur les corps résiduels de points de codimension 1 dans ce cas pour que l'on puisse commencer une récurrence et caractériser, sinon l'image de K^\times/k^\times dans $K^{\times(\ell)}$, du moins son tensorisé par $\mathbf{Z}_{(\ell)}$, l'anneau des fractions rationnelles à dénominateur non divisible par ℓ .

Fin de la preuve dans le cas géométrique. Finalement, par des arguments géométriques astucieux (cf. le chap. 8), on arrive à caractériser K^\times/k^\times en tant que réseau entier dans son tensorisé par $\mathbf{Z}_{(\ell)}$. Pour terminer la preuve, regardons K^\times/k^\times comme l'espace projectif $\mathbf{P}(K)$ associé au k -espace vectoriel K (et de même pour L). On vérifie alors que la bijection ensembliste $\mathbf{P}(\Phi) : \mathbf{P}(K) \cong \mathbf{P}(L)$ dont on dispose *préserve les droites*. Une fois cela acquis, on peut appliquer le théorème fondamental de la géométrie projective qui dit que $\mathbf{P}(\Phi)$ est induit par un isomorphisme $K \rightarrow L$ des espaces vectoriels sous-jacents. C'est alors un exercice facile de vérifier que cet isomorphisme est également multiplicatif.

Enfin, dans les deux cas, on vérifie que l'isomorphisme de corps ainsi obtenu s'étend à un isomorphisme de clôtures séparables induisant bien Φ sur les groupes de Galois, tel qu'il est écrit dans les th. 1.1 et 1.3.

4. VALUATIONS

Soit K un corps de type fini sur un corps parfait k . Appelons *anneau de valuation divisoriel* tout anneau de valuation discrète $A \supset k$ de corps de fractions K et de corps résiduel κ satisfaisant à $\text{tr.deg}(\kappa|k) = \text{tr.deg}(K|k) - 1$. De même, appelons *valuation divisorielle* toute valuation discrète de K associée à un anneau de valuation divisoriel. Nous expliquons dans ce chapitre que dans la situation des th. 1.1 et 1.3 l'isomorphisme Φ induit une bijection entre les anneaux de valuation divisoriels de K et de L , ainsi que des isomorphismes sur les sous-groupes de décomposition et d'inertie associés.

Remarquons d'abord que les prolongements des valuations divisorielles de K à une clôture séparable (resp. à une pro- ℓ -extension maximale) correspondent bijectivement aux sous-groupes de décomposition qui leur sont associés dans G_K (resp. G_K^ℓ). Cela résulte

d'un théorème classique de F. K. Schmidt [35] d'après lequel un corps non-séparablement clos ne peut être muni que d'une seule valuation discrète hensélienne (voir [25], Lemma 8 pour la variante « pro- ℓ »).

Or, dans le cas des corps globaux classiques, Neukirch a obtenu la caractérisation suivante des sous-groupes de décomposition comme ci-dessus : un sous-groupe D de G_K est groupe de décomposition si et seulement si étant donné un nombre premier p , le groupe de cohomologie galoisienne $H^2(G, \mathbf{Z}/p\mathbf{Z})$ est isomorphe à $\mathbf{Z}/p\mathbf{Z}$ ou 0 pour tout sous-groupe ouvert G de D , selon la présence ou non des racines p -ièmes de l'unité dans le corps fixé par G . Cela se déduit après passage à la limite de la suite exacte d'Albert-Brauer-Hasse-Noether en théorie du corps de classes global. Dans son article [37], Michael Spiess démontre un analogue de ce critère pour K le corps de fonctions d'une courbe sur un corps de nombres, en se basant sur un principe de Hasse cohomologique de K. Kato.

Utilisant le principe de Hasse cohomologique de U. Jannsen (non publié à l'heure actuelle) pour une variété projective lisse sur un corps de nombres, il devrait être possible de prouver un critère similaire dans le cas d'un corps de type fini quelconque sur \mathbf{Q} . Mais en fait, la caractérisation est possible en utilisant des outils moins élaborés.

Avant d'expliquer comment, rappelons quelques propriétés des valuations générales (comme références de base, voir [8] ou [10]). Par définition, une *valuation (de Krull)* sur un corps K est une application $v : K \rightarrow \Gamma_v \cup \{\infty\}$, où Γ_v est un groupe abélien totalement ordonné, satisfaisant aux axiomes habituels : (1) $v(x) = \infty$ si et seulement si $x = 0$; (2) $v(xy) = v(x) + v(y)$; et (3) $v(x + y) \geq \min(v(x), v(y))$ pour tout $x, y \in K$. L'ensemble des éléments à valuation non négative est un sous-anneau A_v de K , appelé *l'anneau de valuation de v* . C'est un anneau local de corps de fractions K ; notons M_v son idéal maximal et $\kappa(v)$ son corps résiduel. La dimension $\text{rr}(\Gamma_v)$ (éventuellement infinie) du \mathbf{Q} -espace vectoriel $\Gamma_v \otimes \mathbf{Q}$ est appelée le *rang (rationnel)* de v . Dans le cas où K contient un sous-corps k sur lequel v est triviale, on a l'inégalité fondamentale :

$$\text{tr.deg}(K|k) \geq \text{rr}(\Gamma_v) + \text{tr.deg}(\kappa(v)|k).$$

Les anneaux de valuation sont caractérisés par la propriété suivante : un anneau A de corps de fractions K est l'anneau de valuation d'une valuation de Krull si et seulement si tout $x \in K \setminus A$ satisfait à $x^{-1} \in A$. De cette propriété on déduit immédiatement deux énoncés très utiles pour la suite. Le premier dit que si $A \subset B$ sont des anneaux de valuation ayant même corps de fractions et si B domine A , alors $A = B$. Le deuxième est le suivant : si A est un anneau de valuation de corps de fractions K et corps résiduel κ , l'image réciproque dans A de tout anneau de valuation de κ est un anneau de valuation de K .

Comme dans le cas des valuations discrètes, si $N|K$ est une extension galoisienne (finie ou infinie) et si w est une valuation prolongeant la valuation v , on définit le *sous-groupe de décomposition* $D_w \subset \text{Gal}(N|K)$ associé à w comme le stabilisateur de w .

L'extension des corps résiduels $\kappa(w)|\kappa(v)$ est quasi-galoisienne et le morphisme naturel $D_w \rightarrow \text{Gal}(\kappa(w)|\kappa(v))$ est surjectif; on définit le *sous-groupe d'inertie* I_w comme son noyau. Le groupe Γ_w/Γ_v est de torsion; posons $X_w := \text{Hom}(\Gamma_w/\Gamma_v, \kappa(w)^\times)$. On dispose d'un morphisme naturel $\chi : I_w \rightarrow X_w$ défini comme suit : on associe à $\sigma \in I_w$ le caractère χ_σ défini par $\chi_\sigma(\alpha) := \sigma(x)/x \bmod M_w$, avec $x \in N^\times$ un élément arbitraire de valuation α . La caractéristique χ_σ est bien défini, car on vérifie immédiatement que $\sigma(u)/u = 1 \bmod M_w$ si u est une unité de A_w ou un élément de K . On montre alors que le noyau P_w de $\chi : I_w \rightarrow X_w$ est l'unique sous-groupe de (pro)- p -Sylow de I_w , où p est la caractéristique de $\kappa(w)$; il est trivial pour $p = 0$. On appelle P_w le *sous-groupe d'inertie sauvage* et le quotient $I_w^{\text{mod}} := I_w/P_w$ le *groupe d'inertie modéré*. Le groupe I_w^{mod} est donc abélien, isomorphe à $\text{Hom}(\Gamma_w/\Gamma_v, \kappa(w)^\times)$. En particulier :

LEMME 4.1. — Dans le cas où N est la pro- ℓ -extension maximale de K pour un $\ell \neq p$, le groupe $I_w = I_w^{\text{mod}}$ est isomorphe à \mathbf{Z}_ℓ si et seulement si $\Gamma_w/\ell\Gamma_w \cong \mathbf{Z}/\ell\mathbf{Z}$.

Passons maintenant à la construction promise. Elle est basée sur les deux théorèmes remarquables que voici.

THÉORÈME 4.2. — Soit $\ell \neq 2$ un nombre premier et F un corps satisfaisant à la condition $[F^\times : F^{\times\ell}] \geq \ell^2$. Alors un sous-groupe $D \subset G_F^\ell$ est le sous-groupe de décomposition associé à une valuation (de Krull) w , avec ℓ premier à la caractéristique de $\kappa(w)$ et $\Gamma_w/\ell\Gamma_w \neq 0$, si et seulement si D admet un sous-groupe normal abélien non trivial.

Par la discussion précédente, la condition du théorème est nécessaire. La suffisance a été démontrée par Engler-Koenigsmann [11]. L'énoncé vaut également pour $\ell = 2$ avec toutefois un cas exceptionnel : voir Efrat [9] et Engler-Nogueira [12]. Les démonstrations utilisent des constructions algébriques complètement explicites et élémentaires.

L'autre ingrédient essentiel est le suivant :

THÉORÈME 4.3. — Soit K un corps de type fini sur un corps parfait k . Supposons $k = \mathbf{Q}$ où $\text{tr.deg}(K|k) > 1$. Alors G_K n'admet pas de sous-groupe normal fermé pro-résoluble non trivial. De même, G_K^ℓ n'admet pas de sous-groupe normal fermé abélien non trivial.

C'est une application du théorème d'irréductibilité de Hilbert. Pour le premier énoncé, voir [14], Theorem 15.10. Le deuxième découle du premier (cf. [28], Lemma 1.15).

Nous pouvons maintenant procéder au

THÉORÈME 4.4. — Soient K, L et Φ comme le th. 1.1 ou le th. 1.3. Dans la situation du th. 1.1, on suppose de plus que K et L ne sont pas des corps globaux classiques. Alors Φ induit une bijection entre les valuations divisorielles de K et de L .

Nous ne donnons la preuve qu'en caractéristique positive; pour des remarques concernant le cas de caractéristique 0, voir plus bas.

La démonstration va utiliser à un moment le th. 4.2 que nous n'avons énoncé que pour $\ell \neq 2$; on peut montrer en se basant sur les résultats de Efrat/Engler-Nogueira cités ci-dessus qu'il est également applicable pour $\ell = 2$.

Démonstration. — Dans la situation du th. 1.1, choisissons un nombre premier ℓ premier à la caractéristique (que l'on suppose désormais positive), et passons à l'isomorphisme $G_K^\ell \cong G_L^\ell$ induit par Φ , que l'on va abusivement noter de la même façon.

Soit v une valuation divisorielle de K , et soit $D_w \subset G_K^\ell$ le groupe de décomposition associé à un prolongement w de v . D'après les rappels sur les valuations (en particulier le lemme 4.1), D_w est extension de $G_{\kappa(v)}^\ell$ par le sous-groupe d'inertie (modérée) $I_w \cong \mathbf{Z}_\ell$. D'après le th. 4.3, I_w est donc l'unique sous-groupe normal abélien fermé maximal de D_w . Posons $D := \Phi(D_w)$. Il admet $\Phi(I_w)$ comme sous-groupe normal abélien non trivial, donc il est égal au groupe de décomposition $D_{w'}$ d'une valuation de Krull w' d'après le th. 4.2. Comme $\Phi(I_w) \cong \mathbf{Z}_\ell$ est le sous-groupe normal abélien maximal de $D_{w'}$, il ne peut être que le sous-groupe d'inertie; toujours d'après le lemme 4.1, la restriction v' de w' à L satisfait à $\Gamma_{v'}/\ell\Gamma_{v'} \cong \mathbf{Z}/\ell\mathbf{Z}$. Ensuite, notons que v' est nécessairement triviale sur la clôture algébrique l du corps premier (fini) dans L . On a alors $\text{tr.deg}(L|l) = 1 + \text{tr.deg}(\kappa(v')|l)$. Cela se vérifie en utilisant les théorèmes de transition de Tate en cohomologie galoisienne (*cf.* [36], §II.4) : ils nous apprennent en effet que les degrés de transcendance de L (resp. $\kappa(v')$) se lisent sur la *dimension cohomologique* de G_L^ℓ (resp. $G_{\kappa(v')}^\ell$) et que ces dimensions cohomologiques diffèrent de 1. Enfin, par comparaison avec l'inégalité fondamentale $\text{tr.deg}(L|l) \geq \text{rr}(v') + \text{tr.deg}(\kappa(v')|l)$, on obtient $\text{rr}(v') = 1$. Comme d'autre part $\Gamma_{v'}/\ell\Gamma_{v'} \cong \mathbf{Z}/\ell\mathbf{Z}$, v' est discrète. \square

Notons que la preuve donnée ci-dessus montre également que Φ induit un isomorphisme entre les sous-groupes de décomposition et d'inertie associés à w (resp. w') dans G_K^ℓ (resp. G_L^ℓ), d'où également un isomorphisme $G_{\kappa(v)}^\ell \cong G_{\kappa(v')}^\ell$.

En caractéristique 0, la situation est plus compliquée, car la valuation v' construite ci-dessus pourrait être non triviale sur le corps de base l , et de caractéristique résiduelle positive. Dans la situation du th. 1, Pop surmonte cette difficulté (voir [28], pp. 163–170 pour les détails) en considérant simultanément tous les ℓ , en montrant que la valuation v' obtenue ci-dessus ne dépend pas de ℓ et que ses sous-groupes de décomposition et d'inertie dans G_L tout entier sont isomorphes à ceux de v via Φ . Ainsi, $G_{\kappa(v)} \cong G_{\kappa(v')}$; en particulier, la ℓ -dimension cohomologique de $\kappa(v')$ est > 1 pour tout nombre premier ℓ , donc $\kappa(v')$ ne peut être que de caractéristique 0. On montre alors de la même façon que v' est divisorielle. Le cas de caractéristique 0 du th. 1.3 est encore plus compliqué et utilise la théorie des éléments d'inertie (*cf.* le chap. 9); nous ne le discutons pas ici.

Voici une application, annoncée au chapitre précédent.

PROPOSITION 4.5. — *Soient K, L deux corps de type fini sur \mathbf{Q} , soit $\Phi : G_K \xrightarrow{\sim} G_L$ un isomorphisme comme dans le th. 1.1, et soit k (resp. l) la clôture algébrique du corps*

premier dans K (resp. L). Alors Φ transforme la projection $G_K \rightarrow G_k$ en la projection $G_L \rightarrow G_l$.

Démonstration. — Soit X une k -variété projective lisse géométriquement intègre ayant K comme corps de fonctions (une telle X existe d'après le théorème de désingularisation de Hironaka), et soit H_X le sous-groupe fermé de G_K engendré par les sous-groupes d'inertie associés aux points de codimension 1 de X . Comme X est régulier, la description du groupe fondamental d'un schéma normal jointe au théorème de Zariski sur la pureté du lieu de ramification (cf. [16], exposé X) montrent que le quotient G_K/H_X est isomorphe au groupe fondamental de X . Mais ce dernier groupe est extension de G_k par le groupe fondamental de $X \times_k \bar{k}$, où $\bar{k} \supset k$ est une clôture algébrique ; on sait que ce dernier groupe est topologiquement de type fini (voir [15]). Ainsi, si maintenant H_K est le sous-groupe fermé de K engendré par les sous-groupes d'inertie de toutes les valuations divisorielles, on obtient que $\Pi_K := G_K/H_K$ est extension de G_k par un sous-groupe fermé F_K topologiquement de type fini. Procédant de façon analogue pour L , on construit des groupes H_L , Π_L et F_L . D'après le th. 4.4, Φ induit des isomorphismes $H_K \cong H_L$ et $\Pi_K \cong \Pi_L$. Mais $\Phi(F_K) \subset F_L$, car G_l n'admet pas de sous-groupes normaux fermés non triviaux qui sont topologiquement de type fini (à nouveau par le théorème d'irréductibilité de Hilbert, cf. [14], th. 15.10). D'où $\Phi(F_K) = F_L$ par symétrie, et la proposition est établie. \square

Nous disposons donc un isomorphisme $G_k \cong G_l$; le théorème de Neukirch-Uchida cité au chap. 2 permet d'en déduire l'existence d'un isomorphisme $k \cong l$.

Conformément à la stratégie du chapitre précédent, nous allons utiliser la proposition ci-dessus dans la preuve du th. 1.1 en caractéristique 0. En revanche, en caractéristique positive nous ne démontrons ici que l'énoncé plus faible ci-dessous, ce qui est déjà suffisant pour déduire le th. 1.1 du th. 1.3 (voir la remarque 8.6).

PROPOSITION 4.6. — Soient K, L deux corps de type fini sur un corps fini, k et l comme dans la proposition précédente, et soit $\Phi : G_K^\ell \rightarrow G_L^\ell$ un isomorphisme avec ℓ premier à la caractéristique de K et L . Alors Φ transforme la projection $G_K^\ell \rightarrow G_k^\ell$ en la projection $G_L^\ell \rightarrow G_l^\ell$.

Démonstration. — Prenons pour X une k -variété lisse (non nécessairement projective), géométriquement intègre, de corps de fonctions K . Comme ci-dessus, le pro- ℓ -quotient maximal $\pi_1^\ell(X)$ du groupe fondamental de X est le quotient de G_K^ℓ par un sous-groupe normal fermé H_X . Or d'après un théorème de Katz et Lang [20], l'abélianisé $\pi_1^{\text{ab},\ell}(X)$ est extension de $G_k^\ell \cong \mathbf{Z}_\ell$ par un ℓ -groupe fini. Le noyau de la projection $\pi_1^{\text{ab},\ell}(X) \rightarrow G_k^\ell$ n'est donc autre que le sous-groupe de torsion de $\pi_1^{\text{ab},\ell}(X)$. On termine la preuve de manière similaire au cas précédent, en remarquant que la preuve du th. 4.4 a permis de repérer les sous-groupes de décomposition des points de codimension 1 déjà dans G_K^ℓ . \square

On peut montrer que k et l ont même cardinalité par un argument de récurrence sur la dimension se basant sur les estimations de Hasse-Weil sur le nombre de points des courbes sur un corps fini (voir [32], chap. 2) ; nous n'aurons pas besoin de ce résultat.

5. MODÈLES

Soient k un corps parfait et K une extension de type fini de k . Appelons *modèle* de $K|k$ toute k -variété (i.e. k -schéma intègre de type fini) X dont le corps de fonctions est isomorphe à K . Dans la suite, on identifie souvent les anneaux locaux de X à des sous-anneaux de K contenant k .

On a vu au chapitre 4 que si K est comme dans le th 1.1 (resp. le th. 1.3), l'ensemble des anneaux de valuation divisoriels ayant K pour corps de fractions est encodé dans G_K (resp. dans son pro- ℓ -quotient maximal pour ℓ premier à la caractéristique de K). Évidemment, l'anneau local d'un modèle normal de X en un point de codimension 1 est un tel anneau. Réciproquement, étant donné un anneau de valuation divisoriel $A \subset K$, c'est un exercice facile de construire un modèle normal de K dont A est un anneau local.

Lorsque $\text{tr.deg}(K|k) = 1$, tout anneau de valuation divisoriel est un anneau local de l'unique modèle propre et normal de $K|k$. Mais en dimension supérieure, un tel modèle n'est pas unique.

La tâche suivante est donc la caractérisation des ensembles d'anneaux de valuation divisoriels provenant d'*un seul* modèle normal X de K lorsque $\text{tr.deg}(K|k) > 1$. L'exemple type d'un anneau de valuation divisoriel ne provenant pas de X s'obtient comme suit : éclatons un point P de codimension > 1 du k -schéma X et prenons l'anneau local du schéma éclaté à un point générique du diviseur exceptionnel. Heureusement, lorsque P est un point régulier, le corps résiduel de cet anneau possède une propriété géométrique particulière qui va nous permettre de le distinguer : il est le corps de fonctions du diviseur exceptionnel qui est un fibré projectif dans ce cas – en particulier, c'est une variété qui contient beaucoup de courbes *rationnelles* (i.e. birationnelles, après passage à la clôture algébrique du corps de base, à la droite projective). Pop considère alors la notion suivante :

DÉFINITION 5.1. — *Une variété X est dite antiréglée si l'ensemble des points des courbes rationnelles contenues dans X n'est pas Zariski dense.*

REMARQUE 5.2. — On a la relation suivante à des classes de variétés plus souvent étudiées en géométrie algébrique. Rappelons (cf. [21], Proposition IV.1.3) qu'une variété est *uniréglée* si, après changement de base à un corps algébriquement clos non dénombrable, elle admet un *ouvert* dense dont chaque point est contenu dans une courbe rationnelle. Mais, comme nous fait remarquer J.-L. Colliot-Thélène, il existe des variétés non uniréglées qui ne sont pourtant pas antiréglées : prendre, par exemple, une surface K3 complexe admettant une fibration au-dessus de la droite projective dont la fibre générique est

une courbe elliptique sur $\mathbf{C}(t)$ ayant une infinité de points rationnels ; la surface contient alors une infinité de courbes rationnelles, mais pas d'ouvert dense couvert par ces courbes. Ainsi, être antiréglé est une notion plus forte que d'être non uniréglé.

On vérifie immédiatement, en se servant du théorème de Lüroth, la propriété suivante.

LEMME 5.3. — Soit $Y \xrightarrow{g} K X$ une application rationnelle dominante et génériquement finie de k -variétés. Alors si X est antiréglée, il en est de même pour Y . En particulier, être antiréglé est une propriété birationnelle.

Le deuxième énoncé du lemme permet en particulier d'introduire les notions suivantes.

DÉFINITION 5.4. — Un corps de type fini sur k est dit antiréglé s'il est le corps de fonctions d'une k -variété antiréglée. L'anneau local d'un point d'un k -schéma de type fini est antiréglé si son corps résiduel l'est.

Nous aurons également besoin du lemme suivant :

LEMME 5.5. — Soit K un corps de type fini sur k . Alors il existe une extension finie $K'|K$, avec K' un corps antiréglé. On peut même supposer que le degré $[K' : K]$ est une puissance d'un nombre premier ℓ donné.

Démonstration. — Si K est non antiréglé, une construction simple est la suivante. Soit X un modèle de $K|k$. On peut supposer X affine, équipé d'un morphisme fini $X \rightarrow \mathbf{A}^n$ par le lemme de normalisation de Noether. Or \mathbf{A}^n admet beaucoup de revêtements finis antiréglés de degré une puissance de ℓ : prendre, par exemple, un produit $Z = C_1 \times \cdots \times C_n$ de n courbes lisses, affines, munies de projections $C_i \rightarrow \mathbf{A}^1$ de degré une puissance de ℓ , et dont le compactifié est de genre positif ; on vérifie en utilisant le théorème de Lüroth que Z est antiréglé (en fait, il ne contient point de courbe rationnelle). On peut alors prendre pour K' le corps de fonctions d'une composante du produit fibré $X \times_{\mathbf{A}^n} Z$ qui domine Z ; celle-ci est antiréglée d'après le lemme 5.3. \square

Après ces préparations, on peut procéder à la caractérisation promise. D'abord une terminologie : on dit que deux ensembles S et T sont *presque égaux* si les différences $S \setminus (S \cap T)$ et $T \setminus (S \cap T)$ sont finies.

DÉFINITION 5.6. — Soit K un corps de type fini sur le corps parfait k . Notons \mathcal{D}_K l'ensemble des anneaux de valuation divisoriels dont le corps de fractions est K . On appelle un sous-ensemble $S \subset \mathcal{D}_K$ géométrique s'il est presque égal à l'ensemble des anneaux locaux des points de codimension 1 d'un modèle normal de $K|k$.

REMARQUE 5.7. — Dans la définition ci-dessus, on peut même supposer que le modèle normal en question est *propre*. En effet, si U est un modèle normal de $K|k$, on peut le réaliser comme sous-schéma ouvert d'un modèle normal propre X (normaliser une compactification de Nagata). Le complément $X \setminus U$ est un fermé propre, donc ne contient qu'un nombre fini de points de codimension 1.

On a le lemme suivant :

LEMME 5.8. — Si $K'|K$ est une extension finie de K , un sous-ensemble $U \subset \mathcal{D}_K$ est géométrique si et seulement si l'ensemble U' formé des éléments $A' \in \mathcal{D}_{K'}$ satisfaisant à $A' \cap K \in U$ l'est.

C'est évident (si U est géométrique, normaliser un modèle normal correspondant à U dans K' ; la réciproque est similaire).

Vient alors le point clef :

PROPOSITION 5.9. — Soit K un corps de type fini antiréglé sur le corps parfait k , avec $\text{tr.deg}(K|k) > 1$. Alors un sous-ensemble $S \subset \mathcal{D}_K$ est géométrique si et seulement si il est presque égal au sous-ensemble de \mathcal{D}_K constitué par les anneaux de valuation divisoriels antiréglés.

Démonstration. — Soit X un modèle normal de $K|k$. Remarquons d'abord que comme X est antiréglé, il n'admet qu'un nombre fini de points de codimension 1 dont l'anneau local n'est pas antiréglé. Ceci est dû au fait que par hypothèse l'adhérence des courbes rationnelles dans X est un sous-schéma fermé propre.

Ainsi, tenant compte de la remarque 5.7, il suffit de démontrer que si X est propre et A est un élément antiréglé de \mathcal{D}_K , alors A est l'anneau local d'un point de codimension 1 de X . Pour ce faire, traitons d'abord le cas où X est régulier. Comme X est également supposé propre, A domine l'anneau local $\mathcal{O}_{X,x}$ d'un point x de X par le critère valuatif de propreté. Si x est de codimension 1, $\mathcal{O}_{X,x}$ est un anneau de valuation discrète et on a nécessairement $A = \mathcal{O}_{X,x}$. Sinon, éclatons x ; alors A domine l'anneau local d'un point \tilde{x} du schéma éclaté \tilde{X} contenu dans le diviseur exceptionnel. D'après un lemme classique de Zariski (*cf.* [41], Theorem 10 pour le cas typique de dimension 2), en répétant ce procédé un nombre fini de fois, on arrive à un point de codimension 1. Mais il est alors point générique du diviseur exceptionnel provenant de l'éclatement d'un point régulier. Son anneau local est donc non antiréglé (en fait uniréglé), ce qui contredit le choix de A .

Dans le cas général, on peut trouver, d'après le théorème d'altération de de Jong ([19], [3]), une variété projective lisse Y équipée d'un morphisme propre dominant génériquement fini $\pi : Y \rightarrow X$. Or Y est antiréglée en vertu du lemme 5.3, donc le cas précédent s'y applique. On conclut par le lemme 5.8 et par la remarque que π étant dominant, son image contient les points de codimension 1 de X sauf pour un nombre fini. \square

Nous arrivons maintenant à la caractérisation des ensembles géométriques.

THÉORÈME 5.10. — Soient $K|k, L|l$ et Φ comme dans les théorèmes 1.1 ou 1.3. Alors Φ induit une bijection entre les sous-ensembles géométriques de \mathcal{D}_K et \mathcal{D}_L .

Démonstration. — L'énoncé est évident lorsque $\text{tr.deg}(K|k) = \text{tr.deg}(L|l) = 1$: les ensembles géométriques sont les sous-ensembles cofinis de \mathcal{D}_K .

Supposons donc $\text{tr.deg}(K|k) = \text{tr.deg}(L|l) > 1$. D'après le lemme 5.8, il suffit de démontrer l'énoncé après passage à un sous-groupe ouvert dans $G_K \cong G_L$ (resp. $G_K^\ell \cong G_L^\ell$), i.e. à des extensions finies de K et L . En vertu des lemmes 5.3 et 5.5, on peut supposer K et L antirégles, et le théorème résulte alors de la proposition précédente et du fait que la bijection du th. 4.4 envoie les éléments antirégles de \mathcal{D}_K sur ceux de \mathcal{D}_L . Dans la situation du th. 1.3, ce dernier fait sera vérifié pour $\text{tr.deg}(K|k) = \text{tr.deg}(L|l) = 2$ dans la remarque 9.5 (2). Dans tous les autres cas, on peut supposer par récurrence sur le degré de transcendance que Φ induit un isomorphisme sur les corps résiduels, d'où le résultat. \square

6. RÉCURRENCE

Une bonne partie des démonstrations des th. 1.1 et 1.3 procède par récurrence sur le degré de transcendance commun des corps K et L . Dans la situation du th. 1.1, le point de départ est le cas des corps globaux classiques, où le résultat est connu d'après les travaux de Neukirch et Uchida cités au chap. 2. Dans la situation du th. 1.3, la récurrence part du cas de degré de transcendance 2.

L'idée de la récurrence (déjà utilisée dans la dernière preuve ci-dessus) est la suivante : si A est un élément de \mathcal{D}_K et $A' \in \mathcal{D}_L$ l'anneau que lui correspond par la bijection établie au chap. 4, on peut supposer (sauf dans le cas exceptionnel $\text{tr.deg}(K|k) = \text{tr.deg}(L|l) = 2$ du th. 1.3) par récurrence que leurs corps résiduels sont isomorphes, car ils ont un degré de transcendance inférieur à celui de K et L . On peut alors essayer de relever et recoller ces isomorphismes en un isomorphisme entre K et L . Dans le cas où K et L sont de type fini sur \mathbf{Q} , c'est possible ; dans les autres cas on n'obtient qu'un résultat partiel.

Pour construire des relèvements, on exploite les résultats géométriques du chapitre précédent. Nous adaptons ici des arguments de l'article de Michael Spieß [37].

Soit donc K un corps de type fini sur un corps parfait k ; on peut supposer k algébriquement clos dans K . Par ce qui précède, on peut se borner aux cas où $\text{tr.deg}(K|k) > 1$ ainsi qu'au cas où $\text{tr.deg}(K|k) = 1$ et k est un corps de nombres.

LEMME 6.1. —

- (1) Soit U un modèle normal de $K|k$. Alors $\mathcal{O}^\times(U)/k^\times$ est un groupe abélien libre de type fini, où $\mathcal{O}^\times(U)$ est l'anneau des fonctions régulières inversibles sur U .
- (2) Il existe un ensemble fini P_1, \dots, P_m de points de codimension 1 de U tel que le morphisme de réduction naturel vers les corps résiduels

$$\mathcal{O}^\times(U) \rightarrow \bigoplus_{i=1}^m \kappa(P_i)^\times$$

soit injectif.

Démonstration. — Soit X un modèle propre normal de $K|k$ dont U est un sous-schéma ouvert. Alors on dispose d'une suite exacte

$$0 \rightarrow \mathcal{O}^\times(X) \rightarrow \mathcal{O}^\times(U) \rightarrow \text{Div}_{X \setminus U}(X)$$

où $\text{Div}_{X \setminus U}(X)$ est le groupe abélien (libre de type fini) des diviseurs de Weil concentrés sur $X \setminus U$. Comme X est propre, et également géométriquement connexe d'après l'hypothèse sur $K|k$, on a $\mathcal{O}^\times(X) \cong k^\times$, d'où le premier énoncé.

Pour le second, soit d'abord P_1 un point de codimension 1 quelconque de U . Notons C_1 le noyau de la réduction $\mathcal{O}^\times(U) \rightarrow \kappa(P_1)^\times$. Évidemment $C_1 \cap k^\times = \{1\}$, donc C_1 est libre de type fini. Soit x un générateur de C_1 . Choisissons le point $P_2 \neq P_1$ de sorte que l'image \bar{x} de x dans $\kappa(P_2)$ soit différente d'une racine de l'unité (dans le cas où $\text{tr.deg}(\kappa(P_2)|k) > 0$ on peut même prendre \bar{x} transcendant sur k). Alors le noyau C_2 du morphisme $\mathcal{O}^\times(U) \rightarrow \kappa(P_1)^\times \oplus \kappa(P_2)^\times$ a un rang plus petit que celui de C_1 . En continuant ce procédé, on arrive à un noyau trivial en temps fini. \square

Nous pouvons maintenant démontrer le th. 1.1 en caractéristique 0.

THÉORÈME 6.2. — *Soient K, L deux corps de type fini sur \mathbf{Q} . Supposons qu'il existe un isomorphisme $\Phi : G_K \xrightarrow{\sim} G_L$ de groupes profinis. Alors il existe un isomorphisme $\phi : L \xrightarrow{\sim} K$ de corps induisant Φ .*

Démonstration. — Si $\text{tr.deg}(K|\mathbf{Q}) = \text{tr.deg}(L|\mathbf{Q}) = 0$, c'est le théorème de Neukirch-Uchida cité au chap. 2. Supposons donc le théorème connu pour les corps de degré de transcendance plus petit que $\text{tr.deg}(K|\mathbf{Q})$.

Soit U un modèle normal de $K|k$. D'après le th. 5.10, quitte à remplacer U par un ouvert, on peut supposer que Φ induit une bijection entre les points de codimension 1 de U et ceux d'un modèle normal U' de L . Soit P_1, \dots, P_m un ensemble de points de U comme dans le lemme précédent et soit P'_i ($i = 1, \dots, m$) le point de U' correspondant à P_i par la bijection induite par Φ . Quitte à agrandir l'ensemble des P_i (ce qui n'affecte pas l'injectivité du morphisme $\mathcal{O}^\times(U) \rightarrow \bigoplus \kappa(P_i)^\times$), on peut supposer que le morphisme $\mathcal{O}^\times(U') \rightarrow \bigoplus \kappa(P'_i)^\times$ est également injectif.

Nous voulons maintenant définir un morphisme $\phi_U : \mathcal{O}^\times(U) \rightarrow \mathcal{O}^\times(U')$ rendant le diagramme

$$\begin{array}{ccc} \mathcal{O}^\times(U) & \xrightarrow{\phi_U} & \mathcal{O}^\times(U') \\ r \downarrow & & \downarrow r' \\ \bigoplus_{i=1}^m \kappa(P_i)^\times & \xrightarrow{\cong} & \bigoplus_{i=1}^m \kappa(P'_i)^\times \end{array}$$

commutatif, où le morphisme du bas est un isomorphisme par l'hypothèse de récurrence. Comme les morphismes verticaux sont injectifs, ϕ_U est alors nécessairement un isomorphisme. Pour l'existence de ϕ_U , il suffit de montrer que le morphisme naturel $\psi_U : \mathcal{O}^\times(U) \rightarrow \text{coker}(r')$ induit par les morphismes connus du diagramme est trivial.

Or d'après l'argument de théorie de Kummer expliqué au chap. 3, on dispose pour tout $n > 0$ des isomorphismes compatibles $K^\times/K^{\times n} \cong L^\times/L^{\times n}$. D'autre part, la suite exacte

$$0 \rightarrow \mathcal{O}^\times(U) \rightarrow K^\times \rightarrow \text{Div}(U)$$

implique que le morphisme naturel $\mathcal{O}^\times(U)/n \rightarrow K^\times/n$ est injectif (car $\text{Div}(U)$ est sans torsion). Ainsi, par functorialité des isomorphismes de Kummer, on dispose de morphismes compatibles $\phi_{U,n} : \mathcal{O}^\times(U)/n \rightarrow \mathcal{O}^\times(U')/n$ rendant le diagramme

$$\begin{array}{ccc} \mathcal{O}^\times(U)/n & \xrightarrow{\phi_{U,n}} & \mathcal{O}^\times(U')/n \\ r_n \downarrow & & \downarrow r'_n \\ \bigoplus_{i=1}^m \kappa(P_i)^\times/n & \xrightarrow{\cong} & \bigoplus_{i=1}^m \kappa(P'_i)^\times/n \end{array}$$

commutatif. Par conséquent, l'image du morphisme ψ_U défini ci-dessus est contenu dans le sous-groupe divisible de coker (r'). Mais ce dernier sous-groupe est trivial, car coker (r'), étant le conoyau du morphisme $\mathcal{O}^\times(U')/k^\times \rightarrow \bigoplus \kappa(P'_i)^\times/k^\times$, est le quotient d'un groupe libre (somme directe des groupes des diviseurs principaux sur des modèles propres des $\kappa(P'_i)$) par un sous-groupe libre de type fini – il est donc extension d'un groupe libre par un groupe fini.

Ergo habemus ϕ_U . On vérifie immédiatement qu'en prenant U de plus en plus petit, les isomorphismes ϕ_U ainsi obtenus sont compatibles, d'où un isomorphisme $K^\times \xrightarrow{\sim} L^\times$ car tout élément de K^\times définit une fonction régulière sur U suffisamment petit. Enfin, en considérant l'image d'éléments $x, y, x + y \in K$ arbitrairement choisis dans une infinité de $\kappa(P)$, on vérifie (en utilisant l'hypothèse de récurrence) qu'en rajoutant les éléments 0 on obtient une bijection additive. \square

En examinant la preuve donnée ci-dessus, on voit que le seul obstacle qui l'empêche de s'appliquer à la situation du th. 1.3 est que là, on ne connaît la correspondance de Kummer que pour n une puissance du nombre premier ℓ . (Il y a un problème similaire dans le cas du th. 1.1 en caractéristique positive, où n doit être supposé premier à la caractéristique.) Or le conoyau de r' peut très bien contenir des éléments ℓ -divisibles non triviaux, à savoir les éléments d'un sous-groupe fini d'ordre premier à ℓ . Mais l'on peut tuer ce sous-groupe en tensorisant avec $\mathbf{Z}_{(\ell)}$, le localisé de \mathbf{Z} en (ℓ) . Ainsi, dans la situation du th. 1.3, en supposant $\text{tr.deg}(K|k) > 2$ et le résultat connu pour les corps de degré de transcendance plus petit, on peut au moins déduire l'existence d'un isomorphisme de $\mathbf{Z}_{(\ell)}$ -modules $\phi_\ell : K^\times \otimes \mathbf{Z}_{(\ell)} \xrightarrow{\sim} L^\times \otimes \mathbf{Z}_{(\ell)}$.

Mais en fait, pour mener à bien la preuve du th. 1.3, on peut se contenter d'un énoncé un peu plus faible. Au chapitre 9, nous allons donner des indications sur la caractérisation de l'image de $K^\times/k^\times \otimes \mathbf{Z}_{(\ell)}$ dans $K^{\times(\ell)}$ dans le cas $\text{tr.deg}(K|k) = 2$. Admettons pour le moment que l'on connaît ce résultat. Alors on peut mettre en marche la récurrence précédente et l'on obtient :

PROPOSITION 6.3. — Soient K, L, Φ comme dans le théorème 1.3. Alors (quitte à passer à des extensions purement inséparables en caractéristique positive) il existe un isomorphisme de $\mathbf{Z}(\ell)$ -modules

$$\phi_\ell : (K^\times/k^\times) \otimes \mathbf{Z}(\ell) \xrightarrow{\sim} (L^\times/l^\times) \otimes \mathbf{Z}(\ell)$$

s'insérant dans le diagramme commutatif :

$$\begin{array}{ccc} (K^\times/k^\times) \otimes \mathbf{Z}(\ell) & \xrightarrow{\cong} & (L^\times/l^\times) \otimes \mathbf{Z}(\ell) \\ \downarrow & & \downarrow \\ K^{\times(\ell)} & \xrightarrow{\cong} & L^{\times(\ell)} \end{array}$$

7. PROJECTIONS

Dans ce chapitre et le suivant, nous complétons la démonstration du th. 1.3 (modulo le résultat pour $\text{tr.deg}(K|k) = 2$ admis au chapitre précédent). Cette partie de la démonstration utilise plusieurs idées remontant au travail [5] de Bogomolov ; outre les textes de Pop, la consultation de [7] nous a été très utile lors de la rédaction de ces deux chapitres.

Dans la démonstration, un outil de première importance est la considération des quotients de G_K^ℓ et G_L^ℓ correspondant à des morphismes d'un modèle propre normal de K (resp. L) vers la droite projective. Dans ce chapitre, nous étudions ces quotients.

Appelons une fonction $x \in K$ *générale* si le sous-corps $k(x)$ est algébriquement clos dans K . La signification géométrique de cette définition est la suivante. Soit $X|k$ un modèle projectif normal de $K|k$. La fonction x induit une application rationnelle $\pi_x : X \rightarrow \mathbf{P}_k^1$ définie en codimension 1 ; quitte à éclater X , on peut donc supposer que π_x est un morphisme.

LEMME 7.1. — Dans la situation ci-dessus, supposons en plus que le morphisme π_x est séparable. Alors x est une fonction générale si et seulement si les fibres de π_x sont connexes.

Démonstration. — Comme X est supposé projectif, le morphisme $\pi_x : X \rightarrow \mathbf{P}_k^1$ se factorise en un composé $X \xrightarrow{c_x} C \xrightarrow{f_x} \mathbf{P}_k^1$, où f_x est fini et les fibres de c_x sont connexes, d'après le théorème de factorisation de Stein. Le morphisme séparable f_x est un isomorphisme si et seulement si x est une fonction générale, d'où le lemme. \square

Toute projection $\pi_x : X \rightarrow \mathbf{P}_k^1$ suffisamment générale (au sens de la géométrie algébrique) est à fibres connexes, ce qui explique la terminologie. D'autre part, utilisant cette remarque on montre facilement que les fonctions générales engendrent l'extension $K|k$.

Le groupe de Galois $G_{k(x)}^\ell$ est de façon naturelle quotient du groupe G_K^ℓ (rappeler que nous sous-entendons qu'une clôture séparable de K a été fixée). Notre objectif suivant est de montrer que l'isomorphisme Φ du th. 1.3 respecte ce type de quotients.

Commençons par montrer que Φ induit une bijection entre les quotients de G_K^ℓ et G_L^ℓ provenant de sous-corps algébriquement clos dans K (resp. L) de degré de transcendance 1.

PROPOSITION 7.2. — Soit t une fonction (non nécessairement générale) de K , et soit T la clôture algébrique de $k(t)$ dans K . Notons N_T le noyau de la projection $G_K^\ell \rightarrow G_T^\ell$. Alors il existe un sous-corps $T' \subset L$, de degré de transcendance 1 sur le corps des constantes $l \subset L$, donnant lieu à un diagramme commutatif :

$$\begin{array}{ccc} G_L^\ell & \longrightarrow & G_L^\ell / \Phi(N_T) \\ \downarrow = & & \downarrow \cong \\ G_L^\ell & \longrightarrow & G_{T'}^\ell \end{array}$$

Avant de démontrer la proposition, rappelons quelques faits bien connus sur le *symbole modéré* : pour un corps F muni d'une valuation discrète v , c'est le morphisme de groupes $\partial_v : F^\times \times F^\times \rightarrow \kappa(v)^\times$ qui associe à un couple (f, g) l'image de $(-1)^{v(f)v(g)} f^{v(g)} / g^{v(f)}$ dans $\kappa(v)^\times$. De plus, on a un diagramme commutatif

$$\begin{array}{ccc} F^{\times(\ell)} \times F^{\times(\ell)} & \xrightarrow{\partial_v} & \kappa(v)^{\times(\ell)} \\ \downarrow & & \downarrow \\ H^2(G_F^\ell, \mathbf{Z}_\ell) & \longrightarrow & H^1(G_{\kappa(v)}^\ell, \mathbf{Z}_\ell) \end{array}$$

où les morphismes verticaux proviennent de la théorie de Kummer (suivie d'un cup-produit pour le morphisme de gauche) et le morphisme horizontal du bas est un morphisme *résidu* en cohomologie galoisienne (voir l'appendice à la partie II de [36]). Les groupes de cohomologie à coefficients dans \mathbf{Z}_ℓ sont définis ici par passage à la limite projective à partir de ceux à coefficients $\mathbf{Z}/\ell^m \mathbf{Z}$.

Démonstration. — Le groupe $H^1(G_T^\ell, \mathbf{Z}_\ell)$ s'identifie à un sous-groupe de $H^1(G_K^\ell, \mathbf{Z}_\ell)$. D'après le théorème de Tsen, G_T^ℓ est de dimension cohomologique 1, donc le produit $H^1(G_T^\ell, \mathbf{Z}_\ell) \times H^1(G_T^\ell, \mathbf{Z}_\ell)$ est annulé par le cup-produit

$$H^1(G_K^\ell, \mathbf{Z}_\ell) \times H^1(G_K^\ell, \mathbf{Z}_\ell) \rightarrow H^2(G_K^\ell, \mathbf{Z}_\ell).$$

Il en est alors de même pour le sous-groupe $H^1(G_K^\ell / \Phi(N_T), \mathbf{Z}_\ell) \subset H^1(G_L^\ell, \mathbf{Z}_\ell) \cong L^{\times(\ell)}$. Considérons les éléments de $\Phi(K^\times \otimes \mathbf{Z}_{(\ell)}) = L^\times \otimes \mathbf{Z}_{(\ell)} \subset L^{\times(\ell)}$ contenus dans ce sous-groupe. Il s'agit de montrer qu'ils proviennent tous d'éléments de L^\times qui sont algébriquement dépendants sur l . Or, si f, g sont deux éléments non constants de L qui sont algébriquement indépendants, alors leurs images dans $H^1(G_L^\ell, \mathbf{Z}_\ell)$ ont un cup-produit non trivial. Ce dernier fait se vérifie, par exemple, en utilisant le symbole modéré. En effet, on voit aisément qu'il existe une valuation discrète v de L telle que $\partial_v(f, g) \neq 1$ (on peut prendre pour v la valuation divisorielle provenant d'un point générique du diviseur de f sur un modèle propre et normal de L ; alors $v(f) > 0$, mais $v(g) = v(g-1) = 0$, cf. [7], Lemma 10.1). On conclut par le diagramme ci-dessus que le cup-produit des classes de f et g dans $H^1(G_L^\ell, \mathbf{Z}_\ell)$ a un résidu non trivial, donc il est lui-même non trivial. \square

Soit maintenant $T \subset K$ un sous-corps de degré de transcendance 1 comme ci-dessus. Prenons des modèles projectifs normaux X et C de $K|k$ et de $T|k$, respectivement. Quitte

à éclater X , on peut supposer que l'inclusion $T \subset K$ est induite par un morphisme (nécessairement surjectif) $\pi : X \rightarrow C$. En particulier, tout point fermé de C est l'image d'un point de codimension 1 de X . Ces points donnent lieu à des sous-groupes d'inertie *canoniques* dans les abélianisés $G_T^{\text{ab},\ell}$ et $G_K^{\text{ab},\ell}$, respectivement, et en plus, tout sous-groupe d'inertie dans $G_K^{\text{ab},\ell}$ associé à un point P de C est l'image du sous-groupe d'inertie associé à un point de codimension 1 dans $\pi^{-1}(P)$. D'après la proposition précédente, T correspond par l'isomorphisme Φ à un sous-corps $T' \subset L$, de degré de transcendance 1. Une application du th. 4.4 (aux corps K et L) et la discussion précédente montrent alors que Φ envoie isomorphiquement les sous-groupes d'inertie contenus dans $G_T^{\text{ab},\ell}$ sur ceux de $G_{T'}^{\text{ab},\ell}$.

Soit I_T le sous-groupe fermé de $G_T^{\text{ab},\ell}$ engendré par les sous-groupes d'inertie. Le quotient $G_T^{\text{ab},\ell}/I_T$ est isomorphe au pro- ℓ -quotient abélien maximal $\pi_1^{\text{ab},\ell}(C)$ du groupe fondamental de la courbe C . On sait (cf. la remarque 1.4 (2)) que c'est un pro- ℓ -groupe abélien libre dont le rang est le double du genre g de C . Ainsi, si C' est la courbe propre lisse correspondant au corps T' défini ci-dessus, la discussion précédente implique que C et C' ont même genre ; en particulier, C est rationnelle si et seulement si C' l'est. Donc nous avons finalement démontré :

COROLLAIRE 7.3. — *Soit x une fonction générale de K définissant un sous-corps $k(x)$ algébriquement clos dans K . Alors il existe une fonction générale $y \in L$ telle que l'on ait $\Phi(G_{k(x)}^\ell) = G_{l(y)}^\ell$.*

Concentrons-nous maintenant sur le sous-groupe I_T . Soit $(i_1, \dots, i_r) \subset I_T$ une famille de générateurs de sous-groupes d'inertie associés à des points fermés P_1, \dots, P_r de C . Notons U la sous-courbe ouverte de C obtenue en enlevant les points fermés P_j et I_U le sous-groupe fermé de I_T engendré par les i_j . Alors I_U est exactement le noyau de la projection canonique $\pi_1^{\text{ab},\ell}(U) \rightarrow \pi_1^{\text{ab},\ell}(C)$, donc, par la structure de $\pi_1(U)$ rappelée dans la remarque 1.4 (2), il admet la présentation $\langle i_1, \dots, i_r \mid \sum i_j = 0 \rangle$ comme pro- ℓ -groupe abélien (où nous avons employé la notation additive dans $\pi_1^{\text{ab},\ell}(U)$). De là on déduit que le pro- ℓ -groupe abélien I_T admet la présentation $I_T = \langle i_P \mid \sum i_P = 0 \rangle$, où P parcourt l'ensemble des points fermés de C . Cette présentation implique que si (j_P) est une autre famille d'éléments de I_T indexée par les points fermés de C telle que $\sum j_P = 0$, on a nécessairement $i_P = \varepsilon j_P$ pour tout P avec une unité ℓ -adique ε ne dépendant pas de P . Ceci s'applique en particulier à toute famille de générateurs des sous-groupes d'inertie associés aux points fermés de C .

Soit $\text{Div}^0(C)$ le groupe des diviseurs de degré 0 sur C . Toute famille $\mathcal{I} = (i_P)$ de générateurs de sous-groupes d'inertie définit un plongement naturel

$$j_{\mathcal{I}} : \text{Div}^0(C) \rightarrow \text{Hom}_{\mathbf{Z}_\ell}(I_T, \mathbf{Z}_\ell)$$

de la façon suivante : soit d'abord \mathcal{F} le pro- ℓ -groupe abélien libre engendré par les i_P . On définit un morphisme $\text{Div}(C) \rightarrow \text{Hom}_{\mathbf{Z}_\ell}(\mathcal{F}, \mathbf{Z}_\ell)$ en associant à un point fermé P le morphisme qui envoie i_P sur 1 et les autres générateurs sur 0. Il suffit alors de noter que

$D \in \text{Div}(C)$ s'applique à un morphisme $\mathcal{F} \rightarrow \mathbf{Z}_\ell$ qui passe au quotient dans I_T si et seulement si $D \in \text{Div}^0(C)$. Par ce qui précède, le morphisme $j_{\mathcal{I}}$ dépend de \mathcal{I} mais les images de deux tels morphismes sont les mêmes à multiplication par une unité ℓ -adique près.

Supposons maintenant que $T = k(x)$ est le corps de fonctions rationnelles engendré par une fonction générale x . Alors on a les identifications $I_T = G_K^{\text{ab}, \ell}$, $C \cong \mathbf{P}^1$ et $\text{Div}^0(C) \cong K^\times/k^\times$ (car tout diviseur de degré 0 sur \mathbf{P}^1 est principal). D'autre part, l'on dispose d'un morphisme canonique

$$\partial_T : \text{Div}^0(C) \cong T^\times/k^\times \rightarrow H^1(G_T^{\text{ab}, \ell}, \mathbf{Z}_\ell) = H^1(I_T, \mathbf{Z}_\ell) = \text{Hom}_{\mathbf{Z}_\ell}(I_T, \mathbf{Z}_\ell)$$

induit par la théorie de Kummer. On vérifie alors sans difficulté qu'il existe une famille \mathcal{I} de générateurs d'inertie telle que l'on ait $\partial_T = j_{\mathcal{I}}$. On peut tirer de cette discussion le corollaire suivant :

COROLLAIRE 7.4. — *Conservons les notations du corollaire précédent, et notons C (resp. C') la droite projective ayant $k(x)$ (resp. $l(y)$) comme corps de fonctions. Identifions $\text{Div}^0(C)$ (resp. $\text{Div}^0(C')$) à leurs images dans $\text{Hom}(I_{k(x)}, \mathbf{Z}_\ell)$ (resp. $\text{Hom}(I_{l(y)}, \mathbf{Z}_\ell)$) via les morphismes de Kummer $\partial_{k(x)}$, $\partial_{l(y)}$. Alors l'isomorphisme $I_{k(x)} \cong I_{l(y)}$ induit par Φ envoie le groupe $\text{Div}^0(C)$ sur $\varepsilon \cdot \text{Div}^0(C')$, avec ε une unité de \mathbf{Z}_ℓ .*

8. RÉSEAUX

Nous pouvons maintenant terminer la démonstration du th. 1.3 pour $\text{tr.deg}(K|k) > 2$. Soient donc K, L des corps de type fini sur un corps k qui est la clôture algébrique de son corps premier. D'après la prop. 6.3, nous disposons déjà d'un isomorphisme de $\mathbf{Z}_{(\ell)}$ -modules $\phi_\ell : (K^\times/k^\times) \otimes \mathbf{Z}_{(\ell)} \xrightarrow{\sim} (L^\times/l^\times) \otimes \mathbf{Z}_{(\ell)}$. Ces $\mathbf{Z}_{(\ell)}$ -modules sont en fait *libres*, car K^\times/k^\times (resp. L^\times/l^\times) est isomorphe au groupe abélien libre des diviseurs principaux sur un modèle propre normal de $K|k$ (resp. $L|k$). Ainsi, K^\times/k^\times est un *réseau* dans $(K^\times/k^\times) \otimes \mathbf{Z}_{(\ell)}$, i.e. un \mathbf{Z} -module libre sur une base du $\mathbf{Z}_{(\ell)}$ -module correspondant, et de même pour L^\times/l^\times . Pour alléger la notation, notons \mathcal{L}_K (resp. \mathcal{L}_L) ces deux réseaux dans la suite.

Bien sûr, *a priori* il n'y a aucune raison pour que ϕ_ℓ envoie \mathcal{L}_K sur \mathcal{L}_L . Mais l'on va voir que l'information accumulée jusqu'ici sur les fonctions générales permet tout de même de les raccrocher.

Si x est une fonction générale de K , le quotient $k(x)^\times/k^\times$ définit un sous-groupe \mathcal{L}_x de \mathcal{L}_K , lui-même un réseau dans $\mathcal{L}_x \otimes \mathbf{Z}_{(\ell)}$. Le point est que les réseaux \mathcal{L}_x sont extrêmement rigides pour la correspondance induite par ϕ_ℓ . En effet, soit $l(y) \subset L$ le sous-corps correspondant à $k(x)$ par le cor. 7.3, et notons \mathcal{L}_y le réseau $l(y)^\times/l^\times$.

LEMME 8.1. — Dans la situation ci-dessus, l'image de $\phi_\ell(\mathcal{L}_x)$ dans $\mathcal{L}_L \otimes \mathbf{Z}_{(\ell)}$ est $\varepsilon \cdot \mathcal{L}_y$, avec ε une unité de $\mathbf{Z}_{(\ell)}$.

Démonstration. — Avec les notations du cor. 7.4, on a $\mathcal{L}_x \cong \text{Div}^0(C)$ et $\mathcal{L}_y \cong \text{Div}^0(C')$. Donc le corollaire s'applique et fournit une unité $\varepsilon \in \mathbf{Z}_\ell$ comme dans l'énoncé. Mais comme ϕ_ℓ identifie $\phi_\ell(\mathcal{L}_x) \otimes \mathbf{Z}_{(\ell)}$ à $\mathcal{L}_y \otimes \mathbf{Z}_{(\ell)}$, on a nécessairement $\varepsilon \in \mathbf{Z}_{(\ell)}$. \square

Fixons maintenant une fonction générale $x \in K$. D'après le lemme, quitte à remplacer ϕ_ℓ par un composé avec la multiplication par une unité de $\mathbf{Z}_{(\ell)}$, on peut supposer que $\phi_\ell(\mathcal{L}_x) = \mathcal{L}_y$. Le miracle est que cette seule normalisation entraîne que ϕ_ℓ raccroche \mathcal{L}_K et \mathcal{L}_L . Notons en effet $\tilde{\mathcal{L}}_K := \mathcal{L}_K \cap \phi_\ell^{-1}(\mathcal{L}_L)$ (resp. $\tilde{\mathcal{L}}_L := \mathcal{L}_L \cap \phi_\ell(\mathcal{L}_K)$). Ces groupes sont des sous-quotients de K^\times (resp. L^\times); ils sont non triviaux par ce qui précède. De plus, ϕ_ℓ envoie $\tilde{\mathcal{L}}_K$ isomorphiquement sur $\tilde{\mathcal{L}}_L$.

Enfin, notons \tilde{K}^\times l'image réciproque de $\tilde{\mathcal{L}}_K$ dans K^\times ; le groupe \tilde{L}^\times est défini de façon analogue. Avec ces notations, on tire le corollaire suivant du lemme ci-dessus :

COROLLAIRE 8.2. — *Pour une fonction générale $t \in K$ contenue dans \tilde{K}^\times , on a $\mathcal{L}_t \subset \tilde{\mathcal{L}}_K$.*

Bien entendu, il y a un corollaire analogue pour les éléments de \tilde{L}^\times .

Démonstration. — Une application du lemme précédent montre qu'il existe un entier positif m (le dénominateur de l'unité ε qui y figure) tel que l'on ait $\mathcal{L}_x \cap \tilde{\mathcal{L}}_K = m\mathcal{L}_x$, l'intersection étant prise dans \mathcal{L}_K . En particulier, l'image de x dans \mathcal{L}_x est contenu dans $m\mathcal{L}_x$. Mais cela veut dire que le diviseur de x sur la droite projective associée à $k(x)$ a des coefficients divisibles par m , ce qui n'est possible que pour $m = 1$. \square

PROPOSITION 8.3. — *Le sous-ensemble $\tilde{K} := \tilde{K}^\times \cup \{0\} \subset K$ (resp. $\tilde{L} := \tilde{L}^\times \cup \{0\} \subset L$) est un corps.*

Démonstration. — Il suffit de faire la démonstration pour \tilde{K} . Notons d'abord que comme \tilde{K}^\times est par construction un sous-groupe de K^\times , notre tâche est de voir que $\tilde{K} \subset K$ est un sous-groupe additif. Pour cela il suffit de montrer, en se servant de la multiplicativité, que pour tout $t \in \tilde{K}$ on a $t + 1 \in \tilde{K}$.

Soient donc t et x comme ci-dessus. S'ils sont algébriquement dépendants, on a $t \in k(x)$ car $k(x)$ est algébriquement clos dans K . Mais alors $k(t) \subset k(x) \subset \tilde{K}$ par le choix de x , donc en particulier $t + 1 \in \tilde{K}$.

Donc on peut supposer x et t algébriquement indépendants. Prenons $\kappa \in k^\times$ et considérons les éléments

$$w := (x + \kappa)/t \quad \text{et} \quad z := (t + x + \kappa)/(x + \kappa - 1).$$

En se servant de l'indépendance de t et x on vérifie que l'on peut choisir κ tel que w et z soient des fonctions générales de K .

Comme $t \in \tilde{K}$ et $k(x) \subset \tilde{K}$, on a $w \in \tilde{K}$, d'où $k(w) \subset \tilde{K}$ par le corollaire précédent. En particulier, $w + 1 \in \tilde{K}$, d'où $z = t(w + 1)/(x + \kappa - 1) \in \tilde{K}$ par multiplicativité. De nouveau par le corollaire, on a $k(z) \subset \tilde{K}$, donc en particulier $t + 1 = (z - 1)(x + \kappa - 1) \in \tilde{K}$. \square

PROPOSITION 8.4. — *L'extension $K|\tilde{K}$ est purement inséparable.*

Démonstration. — Soit x une fonction générale dans $K \setminus \tilde{K}$. D'après le lemme 8.1, il existe un entier positif m (le dénominateur de l'unité ε du lemme) tel que $m \cdot \mathcal{L}_x \subset \tilde{\mathcal{L}}_K$. Mais alors $k(x)$ est une extension de $k(x) \cap \tilde{K}$ tel que $k(x)^m \subset k(x) \cap \tilde{K}$. Comme $x \notin \tilde{K}$, ceci n'est possible que si la caractéristique est positive et l'extension est purement inséparable. A fortiori, l'extension $\tilde{K}(x)|\tilde{K}$ est purement inséparable. La proposition résulte alors du fait que les fonctions générales engendrent K au-dessus de k . \square

Enfin, le th. 1.3 résulte de :

PROPOSITION 8.5. — *Il existe un unique isomorphisme de corps $\phi : \tilde{K} \rightarrow \tilde{L}$ induisant l'isomorphisme $\tilde{\mathcal{L}}_K \rightarrow \tilde{\mathcal{L}}_L$ fourni par ϕ_ℓ .*

Démonstration. — Notons que par construction $\tilde{\mathcal{L}}_K$ est exactement l'espace projectif $\mathcal{P}(\tilde{K})$ associé au k -espace vectoriel \tilde{K} , et de même pour $\tilde{\mathcal{L}}_L$. D'après le théorème fondamental de la géométrie projective ([1], Theorem 2.26), si l'on sait montrer que $\phi_\ell : \mathcal{P}(\tilde{K}) \xrightarrow{\sim} \mathcal{P}(\tilde{L})$ préserve les droites, il s'en suit que $k \cong l$ et ϕ_ℓ est induit par un isomorphisme de k -espaces vectoriels $\tilde{K} \xrightarrow{\sim} \tilde{L}$. Or on sait déjà d'après le cor. 7.3 que Φ préserve de nombreuses droites, à savoir celles provenant des sous-corps $k(x)$ engendrés par une fonction générale x . Un argument astucieux de géométrie projective axiomatique (voir [7], §3 ainsi que la prop. 2.10) montre alors que la connaissance de ces droites permet déjà de reconstituer toutes les droites dans $\mathcal{P}(\tilde{K})$ et $\mathcal{P}(\tilde{L})$.

Enfin, la multiplicativité de l'application $\tilde{K} \rightarrow \tilde{L}$ s'en suit par un calcul facile (*loc. cit.*, th. 3.7 ou [31], fin du §5). \square

REMARQUE 8.6. — Notons enfin que l'argument ci-dessus fournit également la preuve du th. 1.1 en caractéristique positive. En effet, soit K un corps de type fini sur un corps fini k . Remarquons que si ℓ est premier à la caractéristique de K , le morphisme naturel $K^\times \rightarrow K^{\times(\ell)}$ est injectif car K^\times , étant extension d'un groupe libre par un groupe fini, n'admet pas d'élément ℓ -divisible non trivial. D'autre part, le corps $K\bar{k}$ obtenu en prenant le composé de K avec la clôture algébrique du corps premier k dans une clôture séparable de K satisfait aux hypothèses du th. 1.3. On a obtenu ci-dessus une caractérisation de l'image du morphisme naturel $(K\bar{k})^\times \rightarrow (K\bar{k})^{\times(\ell)}$. Ce morphisme est équivariant pour l'action du groupe de Galois G_k de k , que l'on sait caractériser comme quotient de G_K par la prop. 4.6. L'image de K^\times dans $K^{\times(\ell)}$ s'obtient alors en prenant les points fixes sous G_k .

9. SURFACES

Dans ce dernier chapitre, nous donnons des indications sur la démonstration du th. 1.3 dans le cas où $K|k$ est une extension de type fini de degré de transcendance 2. Comme déjà mentionné *supra*, dans ce cas on ne dispose pas des informations obtenues par la récurrence du chap. 6, donc il faut modifier la stratégie.

L'idée de Pop est de considérer d'abord les notions suivantes.

DÉFINITION 9.1. — *Un élément $g \in G_K^\ell$ est dit élément d'inertie s'il est contenu dans le sous-groupe d'inertie I_w associée à une prolongation w d'une valuation v sur K . L'élément d'inertie g est divisoriel si v est une valuation divisorielle de K . On note In_K l'ensemble des éléments d'inertie contenus dans G_K^ℓ .*

Supposons maintenant de plus que k est la clôture algébrique d'un corps fini; pour des remarques concernant le cas $k = \bar{\mathbf{Q}}$, voir la fin du chapitre. Le grand avantage de ce cas est que toute valuation de K est nécessairement triviale sur k .

On a alors les deux résultats suivants.

PROPOSITION 9.2. — *L'ensemble In_K est fermé dans G_K^ℓ .*

La démonstration consiste à construire explicitement, pour tout g dans l'adhérence de In_k , une valuation contenant g dans son sous-groupe d'inertie. Nous omettons les détails; voir [32].

En revanche, nous donnons la preuve de la proposition clef suivante :

PROPOSITION 9.3. — *L'ensemble In_K est l'adhérence exacte du sous-ensemble des éléments d'inertie divisoriels dans G_K^ℓ .*

Démonstration. — Soit $g \in \text{In}_K$. Notons K^ℓ une pro- ℓ -extension maximale de K sur laquelle g agit, w la valuation de K^ℓ associée à g et $\Lambda \subset K^\ell$ le sous-corps fixé par g . D'après la théorie de Kummer, il existe un élément $\beta \in \Lambda$ qui n'est pas une puissance ℓ -ième et un système compatible (α_i) de racines ℓ^i -ièmes de β tel que K^ℓ soit la réunion des corps $\Lambda(\alpha_i)$. Alors la famille des corps de la forme $F(\alpha_i)$, avec $F \subset \Lambda$ extension galoisienne finie de K contenant β , est cofinale dans la famille des sous-extensions galoisiennes finies de $K^\ell|K$, il suffit donc de montrer que la restriction de g à un tel corps est contenue dans le sous-groupe d'inertie d'une valuation divisorielle.

Soit X un modèle propre régulier de $F|k$; un tel modèle existe comme $\text{tr.deg}(K|k) = 2$ d'après Abhyankar. Pour v la restriction de w à F , l'anneau de valuation A_v domine un anneau local $\mathcal{O}_{X,P}$ de X ; cet anneau est régulier, donc factoriel. Ainsi, β s'écrit comme un produit $\beta = u\pi_1^{\nu_1} \dots \pi_r^{\nu_r}$, avec une unité u et des éléments premiers π_j de $\mathcal{O}_{X,P}$. On montre alors qu'il existe un $1 \leq j \leq r$ pour lequel $\ell \nmid \nu_j$. (En effet, comme $\text{Gal}(K^\ell|\Lambda) \cong \mathbf{Z}_\ell$, l'anneau de valuation $A_{w,\Lambda}$ de la restriction de w à Λ est *pro- ℓ -hensélien*, donc son groupe d'unités est ℓ -divisible. Ainsi, si $\ell \mid \nu_j$ pour tout j , alors x est une puissance ℓ -ième dans $A_{w,\Lambda} \subset \Lambda$, contradiction.) Soit v_j la valuation divisorielle associée à π_j et w_j une extension de v_j à $F(\alpha_i)$. Alors l'on a $\nu_j = v_j(\beta) = w_j(\beta) = \ell^i w_j(\alpha_i)$, donc $\ell \nmid \nu_j$ implique que l'extension $F(\alpha_i)|F$ est non ramifiée pour w_j , i.e. le sous-groupe d'inertie de w_j est égal à $\text{Gal}(F(\alpha_i)|F)$ tout entier. \square

Soit maintenant v_P une valuation divisorielle de K ; elle provient donc d'un point de codimension 1 sur une k -surface normale de corps de fonctions K . Le corps résiduel $\kappa(P)$

est le corps de fonctions d'une k -courbe propre lisse C_P , donc toute valuation v_Q sur $\kappa(P)$ est la valuation discrète associée à un point fermé Q de la courbe C_P . L'image réciproque de l'anneau de valuation de v_Q (i.e. de l'anneau local de Q) dans l'anneau de valuation A_P de v_P est un anneau de valuation provenant d'une valuation v de K (de rang 2). Réciproquement, toute valuation de K dont l'anneau de valuation est contenu dans A_P induit une valuation sur $\kappa(P)$.

LEMME 9.4. — Soit $v_{P'}$ la valuation divisorielle correspondant à P via Φ . Alors Φ envoie les éléments d'inertie de $G_{\kappa(P)}$ sur ceux de $G_{\kappa(P')}$.

Démonstration. — Fixons un groupe de décomposition $D_P \subset G_K^\ell$ d'un prolongement de v_P ; il s'envoie surjectivement sur $G_{\kappa(P)}^\ell$. On vérifie alors sans peine que tout sous-groupe d'inertie I_Q associé à une prolongation de v_Q dans $G_{\kappa(P)}^\ell$ est quotient d'un sous-groupe d'inertie $I_w \subset D_P$ associé à une prolongation w de v dans G_K^ℓ . Donc on obtient finalement que $\text{In}_{\kappa(P)}$ est l'image exacte de l'ensemble $\text{In}_K \cap D_P$. D'après la proposition ci-dessus, $\text{In}_K \cap D_P$ est l'intersection de D_P avec l'adhérence des éléments d'inertie divisoriels. Pour terminer la preuve, il suffit de noter que d'après le th. 4.4 (et sa preuve), l'isomorphisme Φ envoie D_P sur $D_{P'}$ et respecte les sous-groupes d'inertie divisoriels. \square

REMARQUES 9.5. —

- (1) Les sous-groupes d'inertie de $G_{\kappa(P)}^\ell$ sont tous procycliques. On peut les caractériser de la façon suivante : ils sont les sous-groupes procycliques *maximaux* engendrés par un élément d'inertie. Ainsi, d'après le lemme, ils sont respectés par Φ .
- (2) Le lemme implique, comme dans la discussion précédant le cor. 7.3, que les courbes propres et lisses correspondant à $\kappa(P)$ et $\kappa(P')$ ont même genre.

Nous pouvons maintenant donner une *esquisse de la démonstration du th. 1.3* pour $\text{tr.deg}(K|k) = \text{tr.deg}(L|l) = 2$ et k, l des clôtures algébriques d'un corps fini; pour les détails, voir [33].

Soit X un modèle quasi-projectif lisse de $K|k$, et soit \mathcal{D}_X l'ensemble géométrique de valuations divisorielles provenant de points de codimension 1 de X . Utilisant le th. 5.10 et le lemme de Chow, on voit que quitte à rapetisser X , on peut supposer que Φ envoie \mathcal{D}_X à l'ensemble des valuations provenant de points de codimension 1 d'un modèle quasi-projectif lisse Y de $L|l$.

Prenons maintenant un système $\mathcal{I} = (i_P)$ de générateurs de sous-groupes d'inertie associés aux points de codimension 1 de X dans $G_K^{\ell, \text{ab}}$ et posons $\widehat{\text{Div}}(X) := \text{Hom}_{\mathbf{Z}_\ell}(\mathcal{F}_{\mathcal{I}}, \mathbf{Z}_\ell)$, où $\mathcal{F}_{\mathcal{I}}$ est le pro- ℓ -groupe abélien libre de base \mathcal{I} . Comme dans le cas des courbes discuté au chap. 7, le groupe $\widehat{\text{Div}}(X)$ est bien défini à multiplication par une unité de \mathbf{Z}_ℓ près, et l'on dispose d'un morphisme naturel $K^{\times(\ell)} \cong \text{Hom}_{\mathbf{Z}_\ell}(G_K^{\ell, \text{ab}}, \mathbf{Z}_\ell) \rightarrow \widehat{\text{Div}}(X)$ (dépendant de \mathcal{I}). Soit maintenant $D_P \subset G_K^{\ell, \text{ab}}$ le sous-groupe de décomposition d'un point P de codimension 1 sur X . Alors $\kappa(P)$ est le corps de fonctions d'une courbe propre lisse C_P pour laquelle

on dispose d'un morphisme analogue $\kappa(P)^{\times(\ell)} \cong \text{Hom}_{\mathbf{Z}_\ell}(G_{\kappa(P)}^{\ell, \text{ab}}, \mathbf{Z}_\ell) \rightarrow \widehat{\text{Div}}(C_P)$ discuté au chap. 7. Tout élément dans le noyau du morphisme $\text{Hom}_{\mathbf{Z}_\ell}(D_P, \mathbf{Z}_\ell) \rightarrow \text{Hom}_{\mathbf{Z}_\ell}(I_P, \mathbf{Z}_\ell)$ provient d'un élément de $\kappa(P)^{\times(\ell)} \cong \text{Hom}_{\mathbf{Z}_\ell}(G_{\kappa(P)}^{\ell, \text{ab}}, \mathbf{Z}_\ell)$.

Notons \mathcal{K}_X l'image de $K^{\times(\ell)}$ dans $\widehat{\text{Div}}(X)$; on définit \mathcal{K}_{C_P} de façon analogue pour tout P . Ces sous-groupes se projettent surjectivement sur chaque composante de $\widehat{\text{Div}}(X)$ (resp. $\widehat{\text{Div}}(C_P)$) engendrée par un élément de la base canonique. Considérons $\text{Div}(X) \otimes \mathbf{Z}_\ell$ (resp. $\text{Div}(C_P) \otimes \mathbf{Z}_\ell$) comme \mathbf{Z}_ℓ -réseaux dans $\widehat{\text{Div}}(X)$ (resp. $\widehat{\text{Div}}(C_P)$), et notons \mathcal{L}_X (resp. \mathcal{L}_{C_P}) leurs intersections avec \mathcal{K}_X (resp. \mathcal{K}_{C_P}). Ces sous-groupes déterminent $\text{Div}(X) \otimes \mathbf{Z}_\ell$ et $\text{Div}(C_P) \otimes \mathbf{Z}_\ell$ comme \mathbf{Z}_ℓ -réseaux par ce qui précède. Alors il n'est pas difficile à démontrer la caractérisation suivante de $\text{Div}(X) \otimes \mathbf{Z}_\ell$: il est l'unique \mathbf{Z}_ℓ -réseau qui soit compatible, à multiplication par une unité de \mathbf{Z}_ℓ près, avec les \mathbf{Z}_ℓ -réseaux résiduels $\text{Div}(C_P) \otimes \mathbf{Z}_\ell$. L'énoncé précis est le suivant : $\text{Div}(X) \otimes \mathbf{Z}_\ell$ est l'unique \mathbf{Z}_ℓ -réseau, à multiplication par une unité de \mathbf{Z}_ℓ près, tel que pour tout P et tout sous- \mathbf{Z}_ℓ -module de type fini Δ de \mathcal{K}_X dont l'image Δ_P dans $\text{Hom}_{\mathbf{Z}_\ell}(D_P, \mathbf{Z}_\ell)$ est contenue dans \mathcal{K}_{C_P} , l'image dans $\text{Hom}_{\mathbf{Z}_\ell}(D_P, \mathbf{Z}_\ell)$ de la partie de \mathcal{L}_X contenue dans Δ soit égale à la partie de \mathcal{L}_{C_P} contenue dans Δ_P , à multiplication par une unité de \mathbf{Z}_ℓ près.

On peut faire une construction analogue pour le modèle Y de $L|l$ et obtenir, en particulier, un \mathbf{Z}_ℓ -module \mathcal{L}_Y de la même manière que \mathcal{L}_X . D'après le th. 4.4 et le lemme 9.4, l'isomorphisme Φ respecte ces constructions, et donc, par la caractérisation ci-dessus, quitte à multiplier Φ par une unité de \mathbf{Z}_ℓ , on peut supposer qu'il envoie \mathcal{L}_X sur \mathcal{L}_Y . Le problème est alors que ces \mathbf{Z}_ℓ -modules sont en général plus grands que $K^\times/k^\times \otimes \mathbf{Z}_\ell$ (resp. $L^\times/l^\times \otimes \mathbf{Z}_\ell$). Mais on peut surmonter cette difficulté en exploitant que le quotient $Q_X := \mathcal{L}_X / (K^\times/k^\times \otimes \mathbf{Z}_\ell)$ est de torsion. Pour se convaincre de ce dernier fait, on peut d'abord remplacer X par une compactification lisse (elle existe d'après Abhyankar), car ce ne fait qu'agrandir le groupe Q_X . Ensuite, observons que par sa définition même, l'image de \mathcal{L}_X dans le groupe $(\text{Pic}(X)/\ell^n) \otimes \mathbf{Z}_\ell$ est triviale pour tout n ; a fortiori, il en est de même dans $(NS(X)/\ell^n) \otimes \mathbf{Z}_\ell$, où $NS(X)$ est le groupe de Néron-Severi de X . Mais $NS(X) \otimes \mathbf{Z}_\ell$ est de type fini sur \mathbf{Z}_ℓ , donc ne contient pas d'élément ℓ -divisible. Il en résulte que Q_X est quotient du groupe $\text{Pic}^0(X) \otimes \mathbf{Z}_\ell$ qui est effectivement de torsion, X étant projective et lisse sur la clôture algébrique d'un corps fini.

De plus, on peut montrer que \mathcal{L}_X est le plus grand sous- \mathbf{Z}_ℓ -module de \mathcal{K}_X contenant $K^\times/k^\times \otimes \mathbf{Z}_\ell$ tel que le quotient soit de torsion. Nous avons bien sûr une caractérisation analogue de \mathcal{L}_Y . Cette propriété de torsion permet alors de terminer la démonstration par une légère adaptation des arguments présentés aux chapitres 7 et 8, où l'on remplace l'utilisation de $K^\times/k^\times \otimes \mathbf{Z}_\ell$ et $L^\times/l^\times \otimes \mathbf{Z}_\ell$ par celle de \mathcal{L}_X et \mathcal{L}_Y . En particulier, la prop. 7.2 et le lemme 8.1 doivent être adaptés à ce contexte; nous renvoyons à [33] pour les détails techniques.

Terminons par un mot sur le cas où $k = \bar{\mathbf{Q}}$. Le point le plus délicat est alors, comme on l'a déjà mentionné au chap. 4, la distinction entre les valuations divisorielles et les

valuations discrètes non triviales sur $\bar{\mathbf{Q}}$; la théorie des éléments d'inertie joue ici un rôle crucial. Ensuite, on travaille avec des modèles entiers sur $\text{Spec } \bar{\mathbf{Z}}$ des surfaces ayant K et L pour corps de fonctions, et on établit le résultat par un argument de spécialisation modulo p , exploitant le cas $k = \bar{\mathbf{F}}_p$ que l'on vient de voir.

Remerciements. — Lors de la préparation de cet exposé, j'ai bénéficié d'une invitation de Florian Pop à l'université de Bonn qui m'a permis d'avoir de nombreuses discussions avec lui sur ses travaux en préparation. Je lui suis très reconnaissant pour sa générosité et sa disponibilité. Un grand merci également à Jean-Louis Colliot-Thélène, Philippe Gille, János Kollár et Jakob Stix pour des corrections apportées *in extremis*.

RÉFÉRENCES

- [1] E. ARTIN – *Geometric Algebra*, Interscience, New York/London, 1957.
- [2] E. ARTIN, O. SCHREIER – Eine Kennzeichnung der reell abgeschlossenen Körper, *Abh. Math. Sem. Univ. Hamburg* 5 (1927), 225-231.
- [3] P. BERTHELOT – Altérations de variétés algébriques (d'après A. J. de Jong), *Séminaire Bourbaki, année 1995/96*, Astérisque 241 (1997), exp. 815, 273–311.
- [4] F. BOGOMOLOV – Sous-groupes abéliens de groupes de Galois (en russe), *Izv. Akad. Nauk SSSR Ser. Mat.* 55 (1991), 32–67.
- [5] F. BOGOMOLOV – On two conjectures in birational algebraic geometry, in *Algebraic geometry and analytic geometry (Tokyo, 1990)*, ICM-90 Satell. Conf. Proc., Springer, Tokyo, 1991, p. 26–52.
- [6] F. BOGOMOLOV, Yu. TSCHINKEL – Commuting elements in Galois groups of function fields, in *Motives, Polylogarithms and Hodge Theory* (F. Bogomolov, L. Katzarkov, eds.), International Press, 2002, p. 75–120.
- [7] F. BOGOMOLOV, Yu. TSCHINKEL – Reconstruction of function fields, prépublication [math.AG/0303075](https://arxiv.org/abs/math/0303075), 2003.
- [8] N. BOURBAKI – *Algèbre commutative*, Hermann, Paris, 1964.
- [9] I. EFRAT – Abelian subgroups of pro-2 Galois groups, *Proc. Amer. Math. Soc.* 123 (1995), 1031–1035.
- [10] O. ENDLER – *Valuation Theory*, Springer-Verlag, Berlin, 1972.
- [11] A. J. ENGLER, J. KOENIGSMANN – Abelian subgroups of pro- p Galois groups, *Trans. Amer. Math. Soc.* 350 (1998), 2473–2485.
- [12] A. J. ENGLER, J. B. NOGUEIRA, Maximal abelian normal subgroups of Galois pro-2-groups, *J. Algebra* 166 (1994), 481–505.

- [13] G. FALTINGS – Curves and their fundamental groups (following Grothendieck, Tamagawa and Mochizuki), *Séminaire Bourbaki, année 1997/98*, Astérisque 252 (1998), exp. 840, 131–150.
- [14] M. FRIED, M. JARDEN – *Field Arithmetic*, Ergebnisse der Mathematik und Ihrer Grenzgebiete (3), vol. 11, Springer-Verlag, Berlin, 1986.
- [15] A. GROTHENDIECK – *Revêtements étales et groupe fondamental*, Séminaire de Géométrie Algébrique du Bois-Marie 1960–1961 (SGA 1), Lecture Notes in Mathematics, vol. 224, Springer-Verlag, 1971.
- [16] A. GROTHENDIECK – *Cohomologie des faisceaux cohérents et théorèmes de Lefschetz locaux et globaux* (SGA2), North-Holland, 1968.
- [17] A. GROTHENDIECK – Brief an G. Faltings, reproduit dans *Geometric Galois Actions I* (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser., vol. 242, Cambridge Univ. Press, Cambridge, 1997, p. 49–58, trad. anglaise p. 285–293.
- [18] D. HARARI – Le théorème de Tamagawa II, in *Courbes semi-stables et groupe fondamental en géométrie algébrique (Luminy, 1998)*, J.-B. Bost, F. Loeser, M. Raynaud, eds., Progress in Math., 187, Birkhäuser, Basel, 2000, p. 203–216.
- [19] A. J. de JONG – Smoothness, semi-stability and alterations, *Inst. Hautes Études Sci. Publ. Math.* 83 (1996), 51–93.
- [20] N. KATZ, S. LANG – Finiteness theorems in geometric classfield theory, *L'Enseign. Math.* (2) 27 (1981), 285–319 (1982).
- [21] J. KOLLÁR – *Rational Curves on Algebraic Varieties*, Ergebnisse der Mathematik und Ihrer Grenzgebiete (3), vol. 32, Springer-Verlag, Berlin, 1996.
- [22] S. MOCHIZUKI – The local pro- p anabelian geometry of curves. *Invent. Math.* 138 (1999), 319–423.
- [23] H. NAKAMURA, A. TAMAGAWA, S. MOCHIZUKI, The Grothendieck conjecture on the fundamental groups of algebraic curves, *Sūgakū Expositions* 14 (1) (2001), 31–53.
- [24] J. NEUKIRCH – Kennzeichnung der p -adischen und der endlichen algebraischen Zahlkörper, *Invent. Math.* 6 (1969), 296–314.
- [25] J. NEUKIRCH – Kennzeichnung der endlich-algebraischen Zahlkörper durch die Galoisgruppe der maximalen auflösbaren Erweiterungen, *J. reine angew. Math.* 238 (1969), 135–147.
- [26] J. NEUKIRCH – Über die absoluten Galoisgruppen algebraischer Zahlkörper, in *Journées Arithmétiques de Caen*, Astérisque 41-42 (1977), 67–79.
- [27] J. NEUKIRCH, A. SCHMIDT, K. WINGBERG – *Cohomology of number fields*, Grundlehren der Mathematischen Wissenschaften vol. 323, Springer-Verlag, Berlin, 2000.

- [28] F. POP – On Grothendieck’s conjecture of birational anabelian geometry. *Ann. of Math.* (2) 139 (1994), 145–182.
- [29] F. POP – Glimpses of Grothendieck’s anabelian geometry, in *Geometric Galois Actions I* (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser., vol. 242, Cambridge Univ. Press, Cambridge, 1997, p. 145–182.
- [30] F. POP – On Grothendieck’s conjecture of birational anabelian geometry II, prépublication, 1996.
- [31] F. POP – Alterations and birational anabelian geometry, in *Resolution of singularities (Obergrugl, 1997)*, H. Hauser et al., eds., Progress in Math., vol. 181, Birkhäuser, Basel, 2000, p. 519–532.
- [32] F. POP – The birational anabelian conjecture revisited, prépublication, 2002.
- [33] F. POP – Pro- ℓ birational anabelian geometry over algebraically closed fields I, prépublication math.AG/0307076, 2003 ; deuxième partie en préparation.
- [34] M. ROVINSKY – On certain isomorphisms between absolute Galois groups, *Compositio Math.* 136 (2003), 61–67.
- [35] F. K. SCHMIDT – Mehrfach perfekte Körper, *Math. Ann.* 108 (1933), 1–25.
- [36] J-P. SERRE – *Cohomologie galoisienne*, cinquième édition, révisée et complétée, Lecture Notes in Mathematics, vol. 5, Springer-Verlag, Berlin, 1994.
- [37] M. SPIESS – An arithmetic proof of Pop’s theorem concerning Galois groups of function fields over number fields, *J. reine angew. Math.* 478 (1996), 107–126.
- [38] T. SZAMUELY – Le théorème de Tamagawa I, in *Courbes semi-stables et groupe fondamental en géométrie algébrique (Luminy, 1998)*, J.-B. Bost, F. Loeser, M. Raynaud, eds., Progress in Math., 187, Birkhäuser, Basel, 2000, p. 185–201.
- [39] K. UCHIDA – Isomorphisms of Galois groups, *J. Math. Soc. Japan* 28 (1976), 617–620.
- [40] K. UCHIDA – Isomorphisms of Galois groups of algebraic function fields. *Ann. of Math.* (2) 106 (1977), 589–598.
- [41] O. ZARISKI – The reduction of singularities of an algebraic surface, *Ann. of Math.* (2) 40 (1939), 639–689.

Tamás SZAMUELY

Alfréd Rényi Institute of Mathematics

Hungarian Academy of Sciences

PO Box 127

H-1364 Budapest, Hongrie

E-mail : szamuely@renyi.hu

LA CONJECTURE DE GREEN GÉNÉRIQUE

[d'après C. Voisin]

par Arnaud BEAUVILLE

1. Énoncé de la conjecture

La conjecture de Green est une vaste généralisation de deux résultats classiques de la théorie des courbes algébriques. Soit C une courbe complexe¹ projective et lisse (connexe), de genre $g \geq 2$. Soit K_C le fibré canonique (= fibré cotangent) de C . On associe à C son *anneau canonique*

$$R := \bigoplus_{n \geq 0} H^0(C, K_C^n).$$

Notons S l'algèbre symétrique $S \bullet H^0(C, K_C)$; c'est un anneau de polynômes en g indéterminées.

THÉORÈME 1 (M. Noether).— *L'homomorphisme naturel $S \rightarrow R$ est surjectif, sauf si C est hyperelliptique.*

Supposons désormais que C n'est pas hyperelliptique. À l'homomorphisme $S \rightarrow R$ correspond un plongement de C dans l'espace projectif $\mathbb{P}^{g-1} := \mathbb{P}(H^0(C, K_C)^*)$, dit *plongement canonique*, qui joue un rôle fondamental dans l'étude de la géométrie de C . L'étape suivante est d'essayer de comprendre les équations de C dans \mathbb{P}^{g-1} , c'est-à-dire les éléments de S qui s'annulent sur l'image de C ; ils forment un idéal gradué I_C de S , qui est le noyau de l'homomorphisme $S \rightarrow R$.

THÉORÈME 2 (Petri).— *L'idéal gradué I_C est engendré par ses éléments de degré 2, sauf si C est trigonale² ou isomorphe à une courbe plane de degré 5.*

Chacun de ces deux théorèmes décrit la structure du S -module R en termes de l'existence de certains systèmes linéaires sur la courbe C . Par exemple, le théorème de Petri se traduit (sauf pour les exceptions mentionnées dans l'énoncé) par une suite exacte

$$S(-2)^{b_1} \longrightarrow S \longrightarrow R \longrightarrow 0,$$

¹ Les théorèmes 1 et 2 ci-dessous sont vrais en toute caractéristique [S-D]. Ce n'est pas le cas de la conjecture de Green d'après [S1].

² La courbe C est dite trigonale si elle admet un morphisme $C \rightarrow \mathbb{P}^1$ de degré 3.

où l'on note comme d'habitude $S(-p)$ le S -module S muni de la graduation décalée de p crans vers la droite : $S(-p)_i = S_{i-p}$.

Cette présentation est un (petit) bout de la *résolution minimale* P_\bullet du S -module R , dont on sait depuis Hilbert qu'elle est de la forme

$$0 \rightarrow P_{g-2} \rightarrow P_{g-3} \rightarrow \cdots \rightarrow P_0 \rightarrow R \rightarrow 0,$$

où chaque P_i est une somme directe de modules $S(-p)$, et où les différentielles sont données par des matrices à coefficients homogènes de degré ≥ 1 . La résolution minimale est unique à isomorphisme (non unique) près³.

La dualité de Serre entraîne que le complexe $\text{Hom}_S(P_\bullet, S(-g-1))$, décalé de $(g-2)$ crans vers la gauche, définit encore une résolution minimale de R , donc est isomorphe à P_\bullet . Supposons C non hyperelliptique ; on a alors $P_0 = S$ (Th. 1), donc $P_{g-2} = S(-g-1)$, et on s'aperçoit qu'il reste très peu de degrés possibles pour les termes P_i intermédiaires. De manière précise, un argument élémentaire montre qu'il existe un entier $c \geq 1$ tel que P_\bullet soit de la forme :

$$\begin{aligned} 0 \rightarrow S(-g-1) \rightarrow S(-g+1)^{b_1} \rightarrow S(-g+2)^{b_2} \rightarrow \cdots \rightarrow S(-g+c-1)^{b_{c-1}} \rightarrow \\ S(-g+c+1)^{b'_c} \oplus S(-g+c)^{b''_c} \rightarrow \cdots \rightarrow S(-c-2)^{b'_c} \oplus S(-c-1)^{b''_c} \rightarrow \\ S(-c)^{b_{c-1}} \rightarrow \cdots \rightarrow S(-3)^{b_2} \rightarrow S(-2)^{b_1} \rightarrow S, \end{aligned}$$

où les entiers b_i, b'_j, b''_k sont strictement positifs.

La structure de la résolution minimale est donc essentiellement⁴ déterminée par l'entier c .

L'autre volet des théorèmes 1 et 2 porte sur la présence de systèmes linéaires spéciaux sur C . Si L est un fibré en droites sur C , de degré d , on note $h^i(L)$ la dimension de $H^i(C, L)$ ($i = 0, 1$), et l'on pose $\text{Cliff}(L) := g + 1 - (h^0(L) + h^1(L)) = d - 2h^0(L) + 2$; cet invariant vérifie la relation agréable $\text{Cliff}(L) = \text{Cliff}(K_C \otimes L^{-1})$. On définit alors l'*indice de Clifford* $\text{Cliff}(C)$ de C comme le minimum des entiers $\text{Cliff}(L)$ pour tous les fibrés en droites L sur C avec $h^0(L) \geq 2$ et $0 \leq d \leq g-1$. Un théorème classique de Clifford affirme que cet indice est toujours positif, et qu'il est nul si et seulement si C est hyperelliptique ; de plus les courbes d'indice 1 sont exactement celles qui apparaissent dans le théorème 2. Les théorèmes 1 et 2 admettent donc la reformulation suivante :

$$\text{Cliff}(C) \geq 1 \iff c \geq 1, \quad \text{Cliff}(C) \geq 2 \iff c \geq 2,$$

³ Dans le langage des faisceaux, il revient au même de considérer une résolution \mathcal{P}_\bullet du $\mathcal{O}_{\mathbb{P}^{g-1}}$ -module \mathcal{O}_C , où chaque \mathcal{P}_i est une somme directe de faisceaux $\mathcal{O}_{\mathbb{P}^{g-1}}(-p)$, et où les différentielles sont données par des matrices à coefficients homogènes de degré ≥ 1 .

⁴ Les entiers b_i ($i \leq c-1$) ainsi que b''_c sont déterminés par c et g , mais pas b'_i ni b''_i pour $i > c$: le premier cas où l'on trouve deux valeurs distinctes est $g = 7, c = 3$ [S1].

ce qui conduit naturellement à la

CONJECTURE DE GREEN [G].— $c = \text{Cliff}(C)$.

2. Résultats

Dans l'appendice de [G], Green et Lazarsfeld prouvent l'inégalité $c \leq \text{Cliff}(C)$, à l'aide des propriétés de la cohomologie de Koszul établies par Green dans le même article (voir § 3). Il s'agit donc de démontrer l'inégalité opposée, c'est-à-dire, vu ce qui précède, que la composante $(P_p)_{p+1}$ de degré $p+1$ de P_p , avec $p = g - 1 - \text{Cliff}(C)$, est nulle.

Cette conjecture remarquable a vite attiré l'attention des géomètres algébristes. Dans [S1] Schreyer la vérifie pour $g \leq 8$; il observe aussi qu'elle est fautive en caractéristique 2, déjà pour les courbes générales de genre 7. Le « cas suivant » de la conjecture, $\text{Cliff}(C) \geq 3 \iff c \geq 3$, a été démontré (indépendamment) par Voisin [V1] (pour $g \geq 11$), puis Schreyer [S3] en général. Le cas des courbes planes est traité dans [Lo]. Divers cas particuliers ou reformulations de la conjecture apparaissent dans [E], [P-R], [S2], [T]...

Claire Voisin vient de résoudre le cas particulièrement intéressant des courbes *générales* de genre g . Elles vérifient $\text{Cliff}(C) = \lfloor \frac{g-1}{2} \rfloor$ (voir l'Appendice), de sorte que l'énoncé prend une forme particulièrement simple (*conjecture de Green générique*) :

THÉORÈME 3 ([V2], [V3]).— *Posons $c = \lfloor \frac{g-1}{2} \rfloor$. Pour une courbe de genre g générale, la résolution minimale de R est de la forme*

$$0 \rightarrow S(-g-1) \rightarrow S(-g+1)^{b_1} \rightarrow \dots \rightarrow S(-c-2)^{b_c} \rightarrow S(-c)^{b_{c-1}} \rightarrow \dots \rightarrow S(-2)^{b_1} \rightarrow S$$

si g est impair, et

$$0 \rightarrow S(-g-1) \rightarrow S(-g+1)^{b_1} \rightarrow \dots \rightarrow S(-c-3)^{b_{c-1}} \rightarrow S(-c-2)^{b_{c/2}} \oplus S(-c-1)^{b_{c/2}} \rightarrow S(-c)^{b_{c-1}} \rightarrow \dots \rightarrow S(-2)^{b_1} \rightarrow S$$

si g est pair.

En fait la méthode de démonstration donne un résultat plus fort. Pour des entiers g et p fixés, considérons l'ensemble des courbes de genre g p -gonales, c'est-à-dire admettant un morphisme de degré p sur \mathbb{P}^1 . Elles sont paramétrées par un schéma irréductible, le schéma de Hurwitz. Nous verrons au § 6 qu'une variante de la démonstration du théorème 3 dans le cas g pair entraîne la conjecture de Green pour les courbes p -gonales assez générales, pour $p \geq \frac{g}{3} + 1$. Or il se trouve que M. Teixidor a obtenu (par une méthode très différente) le résultat correspondant pour $p \leq \frac{g}{3} + 2$ [T]. Ainsi :

THÉORÈME 4 ([V2], [T]).— *Une courbe p -gonale générale vérifie la conjecture de Green.*

Plus précisément, on a $c = \text{Cliff}(C) = p - 2$ pour $p \leq [\frac{g+3}{2}]$. L'intérêt de cet énoncé vient de ce que pour presque toutes⁵ les courbes C , l'indice de Clifford est égal à $\gamma - 2$, où γ (la « gonalité ») est le plus petit entier tel que C soit γ -gonale.

Signalons que le Th. 3 pour g impair a la conséquence suivante, qui avait été observée par Hirschowitz et Ramanan avant la démonstration de [V3], et qui apporte un peu plus d'eau au moulin de la conjecture de Green :

COROLLAIRE [H-R]. — *Supposons $g = 2k - 1$. Dans l'espace des modules des courbes de genre g , le lieu des courbes qui n'ont pas la résolution minimale générique coïncide avec celui des courbes k -gonales.*

3. Cohomologie de Koszul

Considérons plus généralement une variété projective X , munie d'un faisceau ample L . Notons

$$V = H^0(X, L) \quad S = \mathbf{S} \cdot V \quad R = \bigoplus_n H^0(X, L^n) ;$$

on s'intéresse à la résolution graduée libre minimale P_\bullet du S -module gradué R . Considérons \mathbb{C} comme un S -module via l'homomorphisme d'augmentation $S \rightarrow \mathbb{C}$. Le S -module gradué $\text{Tor}_i^S(\mathbb{C}, R)$ se calcule en substituant à R la résolution P_\bullet ; comme celle-ci est minimale, le complexe $\mathbb{C} \otimes_S P_\bullet$ est à différentielle nulle, et l'on trouve donc des isomorphismes de S -modules gradués $\text{Tor}_i^S(\mathbb{C}, R) \xrightarrow{\sim} \mathbb{C} \otimes_S P_i$. Mais on peut aussi calculer ce module en utilisant une résolution libre graduée de \mathbb{C} . Il en existe une bien connue, le complexe de Koszul

$$0 \rightarrow \Lambda^n V \otimes_{\mathbb{C}} S(-n) \rightarrow \dots \rightarrow \Lambda^2 V \otimes_{\mathbb{C}} S(-2) \rightarrow V \otimes_{\mathbb{C}} S(-1) \rightarrow S$$

(avec $n = \dim V$). La différentielle $d_p : \Lambda^p V \otimes_{\mathbb{C}} S(-p) \rightarrow \Lambda^{p-1} V \otimes_{\mathbb{C}} S(-p+1)$ applique $(v_1 \wedge \dots \wedge v_p) \otimes P$ sur $\sum_i (-1)^{i+1} (v_1 \wedge \dots \wedge v_{i-1} \wedge v_{i+1} \wedge \dots \wedge v_p) \otimes P.v_i$.

Ainsi la composante de degré $p+q$ du S -module gradué $\text{Tor}_p^S(\mathbb{C}, R)$ s'identifie à l'espace d'homologie $\mathcal{K}_{p,q}(X, L)$ du complexe

$$\Lambda^{p+1} V \otimes R_{q-1} \xrightarrow{d_{p+1}} \Lambda^p V \otimes R_q \xrightarrow{d_p} \Lambda^{p-1} V \otimes R_{q+1} .$$

Les espaces $\mathcal{K}_{p,q}(X, L)$ (« cohomologie de Koszul ») possèdent un grand nombre de propriétés intéressantes, étudiées notamment dans [G]. L'une d'elles sera fondamentale pour ce qui suit : supposons pour simplifier L très ample, de sorte que X est plongée dans un espace projectif de façon que $L = \mathcal{O}_X(1)$. Soit Y une section hyperplane⁶ de X , définie par une équation $\ell = 0$ (avec $\ell \in H^0(X, L)$). Considérons les anneaux

⁵ Au moins conjecturalement – voir l'Appendice pour une formulation précise.

⁶ Cela signifiera ici qu'aucune composante de X n'est contenue dans l'hyperplan $\ell = 0$.

$S_Y = \mathbf{S} \bullet H^0(Y, L|_Y)$ et $R_Y = \bigoplus_n H^0(Y, L|_Y^n)$. Faisons en outre l'hypothèse $H^1(X, L^i) = 0$ pour tout $i \geq 0$; elle garantit que S_Y s'identifie à $S/(\ell)$ et R_Y à $R/(\ell)$. Si P_\bullet est une résolution minimale du S -module R , alors $P_\bullet/\ell P_\bullet$ est une résolution minimale du S_Y -module R_Y . On en déduit un *isomorphisme canonique* $\mathcal{K}_{p,q}(X, L) \xrightarrow{\sim} \mathcal{K}_{p,q}(Y, L|_Y)$ (« théorème de Lefschetz »).

Revenons à notre courbe C . D'après le début du § 2, la conjecture de Green se traduit par l'annulation de $\mathcal{K}_{p,1}(C, K_C)$ pour $p = g - 1 - \text{Cliff}(C)$, ou encore par l'exactitude de la suite

$$\mathbf{A}^{p+1}H^0(C, K_C) \xrightarrow{d_{p+1}} \mathbf{A}^p H^0(C, K_C) \otimes H^0(C, K_C) \xrightarrow{d_p} \mathbf{A}^{p-1}H^0(C, K_C) \otimes H^0(C, K_C^2).$$

Pour une courbe C générale de genre g , l'indice de Clifford vaut $[\frac{g-1}{2}]$, et il s'agit donc de prouver l'annulation de $\mathcal{K}_{k,1}(C, K_C)$ avec $k = [\frac{g}{2}]$. Il suffit de l'obtenir pour *une* courbe de genre g ; C. Voisin utilise des courbes situées sur des surfaces très particulières, les surfaces K3.

Rappelons que les surfaces K3 sont, par définition, les surfaces (lisses, compactes) simplement connexes à fibré canonique trivial. Celles qui nous intéressent ici sont les surfaces K3 X polarisées de genre g , c'est-à-dire munies d'un fibré en droites très ample L de carré $2g - 2$; on supposera de plus que la classe de L dans le groupe de Picard $\text{Pic}(X)$ n'est divisible par aucun entier ≥ 2 . Les sections globales de L définissent un plongement de X dans \mathbb{P}^g , dans lequel les sections hyperplanes lisses de X sont des courbes de genre g , plongées dans \mathbb{P}^{g-1} par le plongement canonique. Pour chaque entier $g \geq 3$, les surfaces K3 polarisées de genre g forment une famille irréductible ; une surface assez générale⁷ dans cette famille vérifie $\text{Pic}(X) = \mathbb{Z}[L]$.

Les sections hyperplanes d'une telle surface X ne sont pas génériques pour $g \geq 12$, mais elles tendent à se comporter comme la courbe générique, en particulier du point de vue de la théorie de Brill-Noether [L] : par exemple leur indice de Clifford est l'indice générique $[\frac{g-1}{2}]$. Il est donc tout à fait naturel d'essayer de prouver l'annulation de $\mathcal{K}_{k,1}(C, K_C)$, avec $k = [\frac{g}{2}]$, pour ces courbes. D'après le « théorème de Lefschetz » pour la cohomologie de Koszul, elle est équivalente à l'annulation de $\mathcal{K}_{k,1}(X, L)$. La courbe C va désormais disparaître au profit de la surface K3 X .

4. Le cas de genre pair : stratégie de la preuve

La première idée force de la démonstration est l'interprétation de $\mathcal{K}_{p,1}(X, L)$ en termes du *schéma de Hilbert* de X . Si X est une variété projective et d un entier, le schéma de Hilbert X_d (noté plutôt d'habitude $X^{[d]}$ ou $\text{Hilb}^d(X)$) paramètre les sous-schémas finis de longueur d de X . Rappelons qu'un tel sous-schéma Z consiste

⁷ C'est-à-dire située en dehors d'une réunion dénombrable d'hypersurfaces dans l'espace des paramètres.

en la donnée de points x_1, \dots, x_m de X et en chacun de ces points d'un idéal \mathcal{I}_{x_i} de l'anneau local \mathcal{O}_{x_i} , de façon que $\sum_i \dim_{\mathbb{C}}(\mathcal{O}_{x_i}/\mathcal{I}_{x_i}) = d$. En associant à Z l'ensemble $\{x_1, \dots, x_m\}$, chaque x_i étant compté avec sa multiplicité $\dim(\mathcal{O}_{x_i}/\mathcal{I}_{x_i})$, on obtient un morphisme birationnel ε de X_d sur la puissance symétrique d -ième $\mathbf{S}^d X$. Lorsque X est une *surface*, X_d est lisse et irréductible, de sorte que ε fournit une résolution des singularités de $\mathbf{S}^d X$. Nous nous bornerons à ce cas dans la suite⁸.

Soit I_d la sous-variété de $X \times X_d$ formée des couples (x, Z) tels que x soit un point de Z . C'est une variété normale, munie de projections :

$$\begin{array}{ccc} & I_d & \\ p \swarrow & & \searrow q \\ X & & X_d ; \end{array}$$

la fibre de q en un point Z de X_d s'identifie au sous-schéma Z de X .

On associe à tout fibré en droites L sur X le fibré vectoriel $\mathcal{E}_L := q_*(p^*L)$ sur X_d , de rang d ; sa fibre en un point Z de X_d s'identifie à $H^0(Z, L|_Z)$. On pose $L_d := \det \mathcal{E}_L$. Une analyse précise du fibré en droites q^*L_d conduit alors au résultat suivant :

PROPOSITION 1. — *L'espace $\mathcal{K}_{d-1,1}(X, L)$ s'identifie au conoyau de l'homomorphisme $q^* : H^0(X_d, L_d) \rightarrow H^0(I_d, q^*L_d)$.*

La démonstration sera esquissée au § 5. Comme expliqué au § 3, le théorème 3 pour g pair résultera de :

PROPOSITION 2. — *Soit X une surface K3 dont le groupe de Picard est engendré par un fibré ample L , avec $L^2 = 2g - 2$ et $g = 2d - 2$. L'homomorphisme $q^* : H^0(X_d, L_d) \rightarrow H^0(I_d, q^*L_d)$ est surjectif.*

Comme le morphisme q est fini et plat, on dispose d'un homomorphisme dans l'autre sens $q_* : H^0(I_d, q^*L_d) \rightarrow H^0(X_d, L_d)$, qui vérifie $q_* \circ q^* = d$. La surjectivité de q^* est donc équivalente à l'injectivité de q_* .

Le cœur de la démonstration consiste alors à construire une variété Z , munie d'un morphisme $j : Z \rightarrow X_d$, telle que le carré cartésien

$$\begin{array}{ccc} \tilde{Z} & \xrightarrow{\tilde{j}} & I_d \\ q_Z \downarrow & & \downarrow q \\ Z & \xrightarrow{j} & X_d \end{array}$$

⁸ La Prop. 1 ci-dessous s'étend en toute dimension à condition de se limiter aux sous-schémas finis *curvilignes*, c'est-à-dire contenus dans une courbe lisse.

possède les deux propriétés suivantes :

- (i) l'homomorphisme $\tilde{j}^* : H^0(I_d, q^*L_d) \rightarrow H^0(\tilde{Z}, \tilde{j}^*q^*L_d)$ est injectif ;
- (ii) l'homomorphisme $(q_Z)_* : H^0(\tilde{Z}, q_Z^*(j^*L_d)) \rightarrow H^0(Z, j^*L_d)$ est injectif.

Au vu du diagramme commutatif

$$\begin{array}{ccc}
 H^0(\tilde{Z}, \tilde{j}^*q^*L_d) & \xleftarrow{\tilde{j}^*} & H^0(I_d, q^*L_d) \\
 (q_Z)_* \downarrow & & \downarrow q_* \\
 H^0(Z, j^*L_d) & \xleftarrow{j^*} & H^0(X_d, L_d)
 \end{array}$$

on en déduit aussitôt l'injectivité de q_* , et donc la Proposition 2.

La construction de Z repose sur l'existence d'un fibré vectoriel E de rang 2 remarquable sur X , introduit par Lazarsfeld dans [L]. C'est l'unique fibré de rang 2 stable sur X de déterminant L et seconde classe de Chern $c_2(E) = d$; il vérifie $\dim H^0(X, E) = d + 1$. En associant à une section s de E son schéma des zéros $Z(s)$, on définit un morphisme $\mathbb{P}(H^0(X, E)) \rightarrow X_d$ (qui est d'ailleurs un plongement). Notons W l'image réciproque de $\mathbb{P}(H^0(X, E))$ dans I_d . Elle est formée des couples (s, x) dans $\mathbb{P}(H^0(X, E)) \times X$ tels que $s(x) = 0$. Pour (s, x) dans un ouvert convenable W° de W , le schéma résiduel $Z(s) - x$ est bien défini. Considérons l'application rationnelle $j_0 : X \times W^\circ \dashrightarrow X_d$ qui associe à $(y, (s, x))$ le schéma $(Z(s) - x) \cup y$. En éclatant dans $X \times W$ le lieu des $(y, (s, x))$ tels que $y \in Z(s) - x$ et en restreignant à un gros ouvert, on obtient le morphisme $j : Z \rightarrow X_d$ cherché.

Le cœur de la démonstration consiste alors à vérifier les propriétés (i) et (ii) ci-dessus. Cette vérification prend 30 pages très denses de l'article [V2], qu'il n'est pas question de reproduire ici. J'essaierai d'en indiquer quelques étapes au paragraphe suivant.

5. Le cas de genre pair : quelques détails

a) Démonstration de la Proposition 1

Suivant [V2], nous dirons qu'un ouvert V° d'une variété normale V est *gros* si le fermé complémentaire est de codimension ≥ 2 . Si L est un fibré sur V , l'application de restriction $H^0(V, L) \rightarrow H^0(V^\circ, L)$ est alors un isomorphisme.

La première étape est le calcul de $H^0(X_d, L_d)$. Les homomorphismes de restriction $H^0(X, L) \rightarrow H^0(Z, L|_Z)$, pour $Z \in X_d$, définissent une flèche $H^0(X, L) \otimes_{\mathbb{C}} \mathcal{O}_{X_d} \rightarrow \mathcal{E}_L$, d'où en passant aux Λ^d un homomorphisme $\Lambda^d H^0(X, L) \rightarrow H^0(X_d, L_d)$, qui est en fait un *isomorphisme* : on le voit en remplaçant X_d par le gros ouvert des sous-schémas ayant au plus un point double, et en écrivant ce dernier comme quotient d'un gros ouvert de X^d éclaté le long des diagonales $x_i = x_j$.

On va désormais remplacer X_d par le gros ouvert des sous-schémas *curvilignes*, c'est-à-dire contenus dans une courbe lisse – et I_d par l'ouvert des couples (x, Z) où Z est curviligne. Pour un tel couple le schéma résiduel $Z - x$ est bien défini ; on dispose donc d'un morphisme

$$\tau : I_d \rightarrow X \times X_{d-1} \quad \text{défini par} \quad \tau(x, Z) = (x, Z - x) .$$

C'est un isomorphisme sur l'ouvert U de I_d formé des couples (x, Z) pour lesquels x est un point simple de Z ; il contracte le diviseur $D := I_d - U$ sur la variété d'incidence $I_{d-1} \subset X \times X_{d-1}$.

On déduit facilement de la définition de L_d un isomorphisme

$$(*) \quad q^*L_d \cong \tau^*(L \boxtimes L_{d-1})(-D) ,$$

d'où une suite exacte :

$$0 \rightarrow H^0(I_d, q^*L_d) \rightarrow H^0(X \times X_{d-1}, L \boxtimes L_{d-1}) \rightarrow H^0(I_{d-1}, (L \boxtimes L_{d-1})|_{I_{d-1}}) .$$

Notons $\tau' : I_{d-1} \rightarrow X \times X_{d-2}$ l'application correspondant à τ , et D' le diviseur de I_{d-1} contracté par τ' . Appliquant de nouveau $(*)$ on trouve un isomorphisme $(L \boxtimes L_{d-1})|_{I_{d-1}} \cong \tau'^*(L^2 \boxtimes L_{d-2})(-D')$, d'où une injection de $H^0(I_{d-1}, (L \boxtimes L_{d-1})|_{I_{d-1}})$ dans $H^0(X, L^2) \otimes H^0(X_{d-2}, L_{d-2})$. On a finalement une suite exacte :

$$0 \rightarrow H^0(I_d, q^*L_d) \rightarrow H^0(X, L) \otimes H^0(X_{d-1}, L_{d-1}) \xrightarrow{\beta} H^0(X, L^2) \otimes H^0(X_{d-2}, L_{d-2}) ,$$

de sorte que le conoyau de $q^* : H^0(X_d, L_d) \rightarrow H^0(I_d, q^*L_d)$ s'identifie à l'homologie d'un complexe

$$H^0(X_d, L_d) \xrightarrow{\alpha} H^0(X, L) \otimes H^0(X_{d-1}, L_{d-1}) \xrightarrow{\beta} H^0(X, L^2) \otimes H^0(X_{d-2}, L_{d-2}) ;$$

on vérifie que ce complexe s'identifie via les isomorphismes $\Lambda^p H^0(X, L) \xrightarrow{\sim} H^0(X_p, L_p)$ au complexe de Koszul

$$\Lambda^d H^0(X, L) \xrightarrow{d_d} H^0(X, L) \otimes \Lambda^{d-1} H^0(X, L) \xrightarrow{d_{d-1}} H^0(X, L^2) \otimes \Lambda^{d-2} H^0(X, L) ,$$

d'où la Proposition 1.

b) *Démonstration de la Proposition 2 : propriété (i)*

Dans la suite de ce paragraphe il est commode de poser $k = d - 1$ (de sorte qu'on a $g = 2k$). Notons \mathbb{P} le gros ouvert de $\mathbb{P}(H^0(X, E))$ formé des sections dont le schéma des zéros est curviligne. Reprenons le carré cartésien

$$\begin{array}{ccc} W & \longrightarrow & I_{k+1} \\ \pi \downarrow & & \downarrow \\ \mathbb{P} & \longrightarrow & X_{k+1} ; \end{array}$$

soit $\psi : W \rightarrow X_k$ le morphisme $(\sigma, x) \mapsto Z(\sigma) - x$. En explicitant la définition de Z on se ramène facilement à prouver l'injectivité de l'application

$$\psi^* : H^0(X_k, L_k) \rightarrow H^0(W, \psi^*L_k) .$$

Le point clé pour cela est la construction, à partir d'une étude fine du fibré de Lazarsfeld E , d'un isomorphisme canonique $\Lambda^k H^0(X, L) \xrightarrow{\sim} \mathbf{S}^k H^0(X, E)^*$. D'autre part on montre que le fibré ψ^*L_k est isomorphe à $\pi^* \mathcal{O}_{\mathbb{P}}(k)$, d'où un homomorphisme injectif $\mathbf{S}^k H^0(X, E)^* \hookrightarrow H^0(W, \psi^*L_k)$. On conclut en vérifiant que le diagramme

$$\begin{array}{ccc} \Lambda^k H^0(X, L) & \longrightarrow & \mathbf{S}^k H^0(X, E)^* \\ \downarrow & & \downarrow \\ H^0(X_k, L_k) & \xrightarrow{\psi^*} & H^0(W, \psi^*L_k) \end{array}$$

est commutatif à un scalaire près.

c) *Démonstration de la propriété (ii)*

Notons \widetilde{W} le produit fibré $W \times_{X_{k+1}} I_{k+1}$, de sorte qu'on a un carré cartésien

$$\begin{array}{ccc} \widetilde{W} & \longrightarrow & I_k \\ q_W \downarrow & & \downarrow q \\ W & \xrightarrow{\psi} & X_k \end{array} .$$

Après quelques péripéties, on se ramène à prouver la surjectivité de l'homomorphisme $q_W^* : H^0(W, \psi^*L_k) \rightarrow H^0(\widetilde{W}, q_W^* \psi^*L_k)$. Rappelons qu'on a $\psi^*L_k \cong \pi^* \mathcal{O}_{\mathbb{P}}(k)$. Notons r l'application composée $\widetilde{W} \rightarrow W \rightarrow \mathbb{P}$. En fait Voisin prouve un résultat plus fort, à savoir :

- *L'homomorphisme $r^* : H^0(\mathbb{P}, \mathcal{O}_{\mathbb{P}}(k)) \rightarrow H^0(\widetilde{W}, r^* \mathcal{O}_{\mathbb{P}}(k))$ est surjectif.*

La démonstration de ce résultat occupe 16 pages de [V2] et je ne peux faire mieux qu'y renvoyer le lecteur. Disons simplement qu'on réalise \widetilde{W} comme un sous-schéma de $B_{\Delta}(S \times S) \times \mathbb{P}$, où $B_{\Delta}(S \times S)$ est obtenu en éclatant $S \times S$ le long de la diagonale. La surjectivité cherchée est équivalente à l'annulation d'un H^1 convenable sur $B_{\Delta}(S \times S) \times \mathbb{P}$. Des calculs de cohomologie délicats sur cette variété ramènent cette annulation à des énoncés sur les sections globales du fibré de Lazarsfeld.

6. La conjecture de Green pour les courbes p -gonales générales

Soit toujours X notre surface K3, munie d'un fibré en droites L vérifiant $L^2 = 2g - 2$, $g = 2d - 2$ et $\text{Pic}(X) = \mathbb{Z}[L]$. Comme promis, nous allons voir que

l'annulation de $\mathcal{K}_{d-1,1}(X, L)$ entraîne le Théorème 4. Choisissons des points généraux x_1, \dots, x_δ de X , avec $\delta \leq (d-1)/2$. Comme $\dim H^0(X, E) = d+1$, il existe deux sections linéairement indépendantes s, t de E s'annulant en ces points. Pour un choix générique des x_i et de s, t , la courbe C où s'annule la section $s \wedge t$ de $\Lambda^2 E = L$ est lisse sauf en x_1, \dots, x_δ , où elle a des points doubles ordinaires. Soit $n : N \rightarrow C$ sa normalisation.

PROPOSITION 3.— *La courbe N est $(d-\delta)$ -gonale, et vérifie $c = \text{Cliff}(N) = d - \delta - 2$.*

La courbe N est de genre $\gamma = 2d - 2 - \delta$; les inégalités $0 \leq \delta \leq (d-1)/2$ se traduisent par $\frac{\gamma}{3} + 1 \leq d - \delta \leq \frac{\gamma}{2} + 1$. Cela donne la conjecture de Green pour les courbes p -gonales générales de genre γ , avec $\frac{\gamma}{3} + 1 \leq p \leq \frac{\gamma}{2} + 1$, et donc, compte tenu de [T], le théorème 4.

Démonstration de la Proposition 3 : Les sections s, t engendrent un sous-faisceau de rang 1 de $n^*(E|_C)$; la partie mobile du système linéaire correspondant est un pinceau de degré $d - \delta$ (le nombre de zéros de s ou t en dehors des x_i). La courbe N est donc $(d - \delta)$ -gonale, et il suffit de prouver qu'on a $\mathcal{K}_{p,1}(N, K_N) = 0$ pour $p = \gamma - 1 - (d - \delta - 2) = d - 1$. Or l'annulation de $\mathcal{K}_{d-1,1}(X, L)$ (Prop. 1 et 2) garantit celle de $\mathcal{K}_{d-1,1}(C, K_C) = 0$; il s'agit de comparer $\mathcal{K}_{d-1,1}(C, K_C)$ et $\mathcal{K}_{d-1,1}(N, K_N)$. L'application trace $n_*K_N \rightarrow K_C$ fournit des injections naturelles $H^0(N, K_N) \hookrightarrow H^0(C, K_C)$ et $H^0(N, K_N^2) \hookrightarrow H^0(C, K_C^2)$, d'où un diagramme commutatif

$$\begin{array}{ccccc}
 & & \xrightarrow{r_N} & & \\
 \Lambda^d H^0(N, K_N) & \xrightarrow{d_N} & \Lambda^{d-1} H^0(N, K_N) \otimes H^0(N, K_N) & \longrightarrow & \Lambda^{d-2} H^0(N, K_N) \otimes H^0(N, K_N^2) \\
 \downarrow j' & & \downarrow j & & \downarrow j'' \\
 \Lambda^d H^0(C, K_C) & \xrightarrow{d_C} & \Lambda^{d-1} H^0(C, K_C) \otimes H^0(C, K_C) & \longrightarrow & \Lambda^{d-2} H^0(C, K_C) \otimes H^0(C, K_C^2) \\
 & & \xleftarrow{r_C} & &
 \end{array}$$

Les différentielles d_N et d_C admettent des rétractions canoniques r_N et r_C , définies par $r_\bullet(\tau \otimes \omega) = \frac{1}{d} \omega \wedge \tau$, qui commutent aux flèches verticales; cela entraîne que l'homomorphisme $\mathcal{K}_{d-1,1}(N, K_N) \rightarrow \mathcal{K}_{d-1,1}(C, K_C)$ induit par j est injectif. En effet, si un élément v de $\Lambda^{d-1} H^0(N, K_N) \otimes H^0(N, K_N)$ est tel que $j(v)$ est un bord, on a

$$j(v) = d_C r_C j(v) = d_C j' r_N(v) = j d_N r_N(v),$$

d'où, puisque j est injectif, $v = d_N r_N(v)$. Ainsi $\mathcal{K}_{d-1,1}(C, K_C)$ est nul, d'où la proposition 3.

7. Le cas de genre impair

Ce qui précède repose de manière essentielle sur les propriétés du fibré de Lazarsfeld, qui n'existe qu'en genre pair. Pour traiter le cas g impair, C. Voisin considère

une surface K3 X dont le groupe de Picard est engendré par un fibré en droites très ample L , de carré $2g - 2$, et la classe d'une courbe rationnelle lisse Δ telle que $\deg(L|_{\Delta}) = 2$. Posons $L' = L(\Delta)$. On a $L'^2 = 2g$, $\deg(L'|_{\Delta}) = 0$; le morphisme $X \rightarrow \mathbb{P}^{g+1}$ associé à L' est un plongement en dehors de Δ et contracte Δ sur un point.

Posons $g = 2k + 1$. La première étape de la démonstration est de vérifier que la Proposition 2 s'étend à (X, L') , donnant $\mathcal{K}_{k+1,1}(X, L') = 0$. La démonstration de la propriété (i) s'adapte immédiatement, celle de (ii) demande nettement plus de travail.

Il s'agit maintenant d'en déduire l'annulation de $\mathcal{K}_{k,1}(X, L)$. Il est commode pour cela d'utiliser la dualité de Serre, qui fournit une dualité canonique entre $\mathcal{K}_{p,1}(X, L)$ et $\mathcal{K}_{g-2-p,2}(X, L)$. Ainsi $\mathcal{K}_{k-1,2}(X, L')$ est nul, et on veut en déduire l'annulation de l'espace $\mathcal{K}_{k-1,2}(X, L)$. Rappelons que celui-ci est l'homologie du complexe⁹

$$\Lambda^k H^0(L) \otimes H^0(L) \xrightarrow{d_L} \Lambda^{k-1} H^0(L) \otimes H^0(L^2) \xrightarrow{d_L} \Lambda^{k-2} H^0(L) \otimes H^0(L^3).$$

Au couple (X, L') est associé comme plus haut le fibré de Lazarsfeld E , de déterminant L' ; la preuve repose sur la construction d'un homomorphisme

$$\varphi : \mathbf{S}^k H^0(E) \rightarrow \Lambda^{k-1} H^0(L) \otimes H^0(L(-\Delta))$$

qui s'inspire d'une construction analogue utilisée par Green et Lazarsfeld pour prouver l'inégalité $c \leq \text{Cliff}(C)$ ([G], Appendice). Étant donné deux sections globales v, w de E , on notera $v \wedge w$ leur produit extérieur dans $H^0(\Lambda^2 E) = H^0(L')$. On choisit une base (w_1, \dots, w_{k+1}) de $H^0(E(-\Delta))$, et on pose¹⁰, pour $v \in H^0(E)$,

$$\varphi(v^k) = \sum_{i < j} (-1)^{i+j} (v \wedge w_1) \wedge \dots \wedge \widehat{(v \wedge w_i)} \wedge \dots \wedge \widehat{(v \wedge w_j)} \wedge \dots \wedge (v \wedge w_{k+1}) \otimes (w_i \wedge w_j);$$

la condition $w_i \in H^0(E(-\Delta))$ entraîne bien $v \wedge w_i \in H^0(L)$ et $w_i \wedge w_j \in H^0(L(-\Delta))$.

Choisissons d'autre part une section σ de $H^0(L')$ dont la restriction à Δ n'est pas nulle ; elle fournit un scindage de la suite exacte

$$0 \rightarrow H^0(L) \rightarrow H^0(L') \rightarrow H^0(L'|_{\Delta}) \cong \mathbb{C} \rightarrow 0,$$

d'où une décomposition $H^0(L') = H^0(L) \oplus \mathbb{C} \sigma$. Considérons le diagramme commutatif

$$\begin{array}{ccc} \Lambda^{k-1} H^0(L) \otimes H^0(L(-\Delta)) & \xrightarrow{\delta} & \Lambda^{k-2} H^0(L) \otimes H^0(L^2(-\Delta)) \\ \downarrow 1 \otimes \sigma & & \downarrow 1 \otimes \sigma \\ \Lambda^{k-1} H^0(L) \otimes H^0(L^2) & \xrightarrow{d_L} & \Lambda^{k-2} H^0(L) \otimes H^0(L^3). \end{array}$$

⁹ Dans ce paragraphe, pour tout faisceau F sur X on note simplement $H^0(F)$ l'espace $H^0(X, F)$.

¹⁰ Le chapeau sur un terme signifie comme d'habitude qu'on l'omet.

L'annulation de $\mathcal{K}_{k-1,2}(X, L)$ va résulter des quatre points suivants :

- (i) L'homomorphisme composé $\delta \circ \varphi$ est nul.
- (ii) L'homomorphisme induit $\varphi : \mathbf{S}^k H^0(E) \rightarrow \text{Ker } \delta$ est surjectif.
- (iii) $\mathcal{K}_{k-1,2}(X, L)$ est engendré par les classes d'éléments $(1 \otimes \sigma) \cdot \alpha$ pour $\alpha \in \text{Ker } \delta$.
- (iv) Pour $t \in \mathbf{S}^k H^0(E)$, la classe de $(1 \otimes \sigma) \cdot \varphi(t)$ dans $\mathcal{K}_{k-1,2}(X, L)$ est nulle.

Les assertions (i) et (iv) résultent d'un calcul sans mystères, basé sur l'identité

$$(v_1 \wedge v_2) \cdot (v_3 \wedge v_4) - (v_1 \wedge v_3) \cdot (v_2 \wedge v_4) + (v_1 \wedge v_4) \cdot (v_2 \wedge v_3) = 0 \quad \text{dans } H^0(L'^2)$$

quels que soient v_1, \dots, v_4 dans $H^0(E)$.

Prouvons (iii). Soit $\beta \in \text{Ker } d_L$. Puisque $\mathcal{K}_{k-1,2}(X, L') = 0$, il existe un élément γ de $\mathbf{\Lambda}^k H^0(L') \otimes H^0(L')$ tel que $\beta = d_{L'} \gamma$. La décomposition $H^0(L') = H^0(L) \oplus \mathbb{C} \sigma$ permet d'écrire

$$\gamma = \gamma_1 + \sigma \wedge \gamma_2 + \gamma_3 \otimes \sigma + (\sigma \wedge \gamma_4) \otimes \sigma, \quad \text{avec}$$

$$\gamma_1 \in \mathbf{\Lambda}^k H^0(L) \otimes H^0(L), \quad \gamma_2 \in \mathbf{\Lambda}^{k-1} H^0(L) \otimes H^0(L), \quad \gamma_3 \in \mathbf{\Lambda}^k H^0(L), \quad \gamma_4 \in \mathbf{\Lambda}^{k-1} H^0(L).$$

L'élément γ_4 s'identifie à l'image de $d_{L'} \gamma$ dans $\mathbf{\Lambda}^{k-1} H^0(L') \otimes H^0(L'_{|\Delta})$; comme $d_{L'} \gamma = \beta$ appartient à $\mathbf{\Lambda}^{k-1} H^0(L) \otimes H^0(L^2)$, on en déduit $\gamma_4 = 0$. Comme on peut modifier β par un bord on peut supposer $\gamma_1 = 0$. Enfin on a

$$\gamma_3 \otimes \sigma = d_{L'}(\sigma \wedge \gamma_3) + \sigma \wedge d_{L'} \gamma_3,$$

de sorte qu'en modifiant γ par un bord on peut supposer $\gamma_3 = 0$.

On a alors $\gamma = \sigma \wedge \gamma_2$, et par suite $\beta = d_{L'} \gamma = \gamma_2 \cdot (1 \otimes \sigma) - \sigma \wedge d_L \gamma_2$. En utilisant de nouveau la décomposition $H^0(L') = H^0(L) \oplus \mathbb{C} \sigma$, le fait que β appartient à $\mathbf{\Lambda}^{k-1} H^0(L) \otimes H^0(L^2)$ implique d'une part que $d_L \gamma_2$ est nul, d'autre part que γ_2 appartient au sous-espace $\mathbf{\Lambda}^{k-1} H^0(L) \otimes H^0(L(-\Delta))$. Par suite γ_2 appartient à $\text{Ker } \delta$, et $\mathcal{K}_{k-1,2}(X, L)$ est engendré par les classes des éléments $\gamma_2 \cdot (1 \otimes \sigma)$ avec $\gamma_2 \in \text{Ker } \delta$.

Le gros du travail est la démonstration de (ii), pour laquelle je ne peux que renvoyer à [V3], p. 12-26. Disons simplement qu'on se ramène à un énoncé sur la cohomologie d'un éclatement convenable de $\mathbb{P}(H^0(E)) \times X$, énoncé dont la démonstration demande une ingéniosité technique considérable.

Appendice : l'indice de Clifford

Nous utiliserons dans cet appendice une abréviation très classique : un système linéaire¹¹ $|D|$ sur C de degré d et de dimension projective r est appelé un g_d^r . L'indice de Clifford $\text{Cliff}(C)$ est alors le minimum des entiers $d - 2r$ sur l'ensemble des g_d^r avec $d \leq g - 1$ et $r \geq 1$. D'après le théorème de Clifford, on a $\text{Cliff}(C) \geq 0$, et $\text{Cliff}(C) = 0$

¹¹ Étant donné un diviseur D , le système linéaire $|D|$ est l'ensemble des diviseurs effectifs linéairement équivalents à D ; il s'identifie à l'espace projectif $\mathbb{P}(H^0(C, \mathcal{O}_C(D)))$.

si et seulement si C est hyperelliptique. Cet invariant a été introduit par Martens dans [M], où il montre entre autres que les courbes d'indice 1 sont exactement celles qui apparaissent dans le théorème 2.

Considérons les courbes de genre g fixé. Les courbes d -gonales, c'est-à-dire admettant un g_d^1 , ont un indice de Clifford $\leq d - 2$; lorsqu'elles sont assez générales, leur indice de Clifford est exactement $d - 2$ [Ba]. Il s'ensuit que l'indice de Clifford est $\lfloor \frac{g-1}{2} \rfloor$ pour une courbe générale, et prend toutes les valeurs entre 0 et $\lfloor \frac{g-1}{2} \rfloor$.

Les courbes dont l'indice de Clifford est fourni par un g_d^r avec $r > 1$ (et pas par un système linéaire de dimension plus petite) sont beaucoup plus rares. Pour $r = 2$, ce sont les courbes planes lisses de degré d , qui sont de genre $\frac{1}{2}(d-1)(d-2)$. Pour $3 \leq r \leq 9$, les auteurs de [ELMS] prouvent que cela impose $g = 4r - 2$, avec un indice de Clifford $2r - 3$ donné par un fibré en droites L tel que $L^2 \cong K_C$; ils conjecturent bien naturellement le même énoncé pour tout r (et construisent, pour tout r , une courbe ayant les propriétés indiquées). Si cette conjecture est correcte, et si C est une courbe de genre g et d'indice de Clifford c , alors :

- a) C est $(c+2)$ -gonale, ou
- b) $g = \frac{1}{2}(c+2)(c+3)$, C est une courbe plane lisse de degré $c+4$, ou
- c) c est impair ≥ 3 , $g = 2c+4$, et C admet un fibré en droites L tel que $L^2 \cong K_C$ et $\text{Cliff}(L) = c$.

Pour g fixé, il y a donc (moyennant la conjecture) au plus deux valeurs de c pour lesquelles il existe des courbes d'indice c qui ne soient pas $(c+2)$ -gonales.

Je remercie Olivier Debarre et Claire Voisin pour leurs commentaires pertinents sur une première version de ce texte.

BIBLIOGRAPHIE

- [E] L. EIN : *A remark on the syzygies of the generic canonical curves*. J. Differential Geom. **26** (1987), 361–365.
- [ELMS] D. EISENBUD, H. LANGE, G. MARTENS, F.-O. SCHREYER : *The Clifford dimension of a projective curve*. Compositio Math. **72** (1989), 173–204.
- [G] M. GREEN : *Koszul cohomology and the geometry of projective varieties*. J. Differential Geom. **19** (1984), 125–171.
- [H-R] A. HIRSCHOWITZ, S. RAMANAN : *New evidence for Green's conjecture on syzygies of canonical curves*. Ann. Sci. École Norm. Sup. (4) **31** (1998), 145–152.
- [L] R. LAZARSFELD : *Brill-Noether-Petri without degenerations*. J. Differential Geom. **23** (1986), 299–307.

- [Lo] F. LOOSE : *On the graded Betti numbers of plane algebraic curves*. Manuscripta Math. **64** (1989), 503–514.
- [M] H. MARTENS : *Varieties of special divisors on a curve II*. J. Reine Angew. Math. **233** (1968), 89–100.
- [N] M. NOETHER : *Über die invariante Darstellung algebraischer Funktionen*. Math. Ann. **17** (1880), 263–284.
- [P] K. PETRI : *Über die invariante Darstellung algebraischer Funktionen einer Veränderlichen*. Math. Ann. **88** (1923), 242–289.
- [P-R] K. PARANJAPÉ, S. RAMANAN : *On the canonical ring of a curve*. Algebraic geometry and commutative algebra, Vol. II, 503–516, Kinokuniya, Tokyo (1988).
- [S-D] B. SAINT-DONAT : *On Petri's analysis of the linear system of quadrics through a canonical curve*. Math. Ann. **206** (1973), 157–175.
- [S1] F.-O. SCHREYER : *Syzygies of canonical curves and special linear series*. Math. Ann. **275** (1986), 105–137.
- [S2] F.-O. SCHREYER : *Green's conjecture for general p -gonal curves of large genus*. Algebraic curves and projective geometry (Trento, 1988), 254–260, Lecture Notes in Math. **1389**, Springer, Berlin, 1989.
- [S3] F.-O. SCHREYER : *A standard basis approach to syzygies of canonical curves*. J. Reine Angew. Math. **421** (1991), 83–123.
- [T] M. TEIXIDOR I BIGAS : *Green's conjecture for the generic r -gonal curve of genus $g \geq 3r - 7$* . Duke Math. J. **111** (2002), 195–222.
- [V1] C. VOISIN : *Courbes tétraogonales et cohomologie de Koszul*. J. Reine Angew. Math. **387** (1988), 111–121.
- [V2] C. VOISIN : *Green's generic syzygy conjecture for curves of even genus lying on a K3 surface*. J. Eur. Math. Soc. **4** (2002), 363–404.
- [V3] C. VOISIN : *Green's canonical syzygy conjecture for generic curves of odd genus*. Preprint math.AG/0301359 (2003).

Arnaud BEAUVILLE
 Laboratoire J.-A. Dieudonné
 UMR 6621 du CNRS
 UNIVERSITÉ DE NICE
 Parc Valrose
 F-06108 NICE Cedex 2
 beauville@math.unice.fr

SLE ET INVARIANCE CONFORME
[d'après Lawler, Schramm et Werner]

par **Jean BERTOIN**

1. INTRODUCTION

La physique statistique s'intéresse à des systèmes macroscopiques définis à l'échelle microscopique sur un réseau par un grand nombre de données aléatoires, éventuellement corrélées. Le modèle microscopique est souvent facile à définir ; on s'attend à ce que le changement d'échelle du microscopique au macroscopique induise une limite en loi vers un modèle continu qui garde la trace des propriétés spécifiques du modèle discret.

Un exemple fondamental est la marche aléatoire simple symétrique sur \mathbb{Z}^d : on considère une trajectoire aléatoire $S : \mathbb{N} \rightarrow \mathbb{Z}^d$, disons issue de $S_0 = 0$, telle que les accroissements $\xi_n = S_n - S_{n-1}$ soient des variables indépendantes et identiquement distribuées (i.i.d.), avec $\mathbb{P}(\xi_n = x) = \frac{1}{2d}$, où x désigne l'un des $2d$ plus proches voisins de 0 sur \mathbb{Z}^d . Si l'on effectue le changement d'échelle

$$W_t^{(N)} = \sqrt{d/N} S_{[Nt]}, \quad t \geq 0$$

alors le Théorème Central Limite implique que pour tous $0 \leq t_1 \leq \dots \leq t_k$, le k -uplet $(W_{t_1}^{(N)}, \dots, W_{t_k}^{(N)})$ converge en loi vers $(W_{t_1}^{(\infty)}, \dots, W_{t_k}^{(\infty)})$, où $(W_t^{(\infty)}, t \geq 0)$ est un processus dont les accroissements sont indépendants et tels que pour $0 \leq s \leq t$, $W_t^{(\infty)} - W_s^{(\infty)}$ suit la loi de Gauss de moyenne nulle et de matrice de covariance $(t-s)\text{Id}$. Par ailleurs, le fait que la marche aléatoire simple ne saute que sur ses plus proches voisins suggère que sa limite macroscopique devrait avoir des trajectoires continues, et il est naturel de chercher à construire directement une mesure de probabilité \mathcal{W} sur $\mathcal{C}([0, \infty[, \mathbb{R}^d)$, dont les marginales fini-dimensionnelles sont celles de $W^{(\infty)}$. C'est ce qu'a accompli Norbert Wiener, la mesure \mathcal{W} s'appelle la mesure de Wiener, et le processus correspondant est le mouvement brownien.

Pour de nombreux modèles discrets de la physique statistique en dimension 2, et pour la valeur critique du paramètre à laquelle se produit la transition de phase, les physiciens prédisent non seulement l'existence d'une limite continue, mais également que cette limite

vérifie une propriété d'invariance conforme. Pour reprendre l'exemple du mouvement brownien dans $\mathbb{C} \sim \mathbb{R}^2$, donnons-nous deux domaines simplement connexes $D, D' \neq \mathbb{C}$ contenant l'origine et $f : D \rightarrow D'$ une application conforme qui met D et D' en bijection. Notons \mathcal{W}_D (respectivement $\mathcal{W}_{D'}$) la loi sur l'espace des courbes continues à valeurs dans D (respectivement D') induite par la trajectoire brownienne arrêtée la première fois où elle quitte D (respectivement D'). Alors l'image de \mathcal{W}_D par l'application qui transforme une courbe dans D en une courbe dans D' par l'application conforme f , est $\mathcal{W}_{D'}$. Cette propriété importante du mouvement brownien complexe a été observée par Paul Lévy ; il est intéressant de noter qu'elle est vérifiée pour la limite continue, mais pas pour le modèle discret, c'est-à-dire la marche aléatoire symétrique simple.

Penchons-nous maintenant sur un modèle voisin, celui de la marche aléatoire auto-évitante. Plus précisément, on se donne toujours un domaine simplement connexe $D \neq \mathbb{C}$, on considère la marche aléatoire simple symétrique sur le réseau $N^{-1}\mathbb{Z}^2$ que l'on arrête lorsqu'elle quitte D . La probabilité que cette marche ne visite jamais deux fois ou plus le même site est strictement positive. On peut donc la conditionner par cet événement et on obtient ainsi une marche auto-évitante dans D . D'après ce qui précède, on s'attend d'une part à l'existence d'une limite continue donnée par une mesure de probabilité sur les trajectoires auto-évitantes à valeurs dans D , et d'autre part à ce que cette limite continue vérifie la propriété d'invariance conforme. Pour l'instant, ceci est toujours un problème ouvert. On peut également chercher à construire directement un modèle continu qui satisfasse aux propriétés fondamentales qu'on attend. Une approche naïve consisterait à tenter de conditionner une trajectoire brownienne dans D à ne jamais se recouper, mais ce conditionnement ne peut être réalisé rigoureusement¹ car il est bien connu que presque sûrement, le mouvement brownien plan visite à nouveau des points par lesquels il est déjà passé, et ce pour des intervalles de temps arbitrairement petits.

Depuis quelques années, des avancées spectaculaires ont été réalisées dans ce domaine par Schramm, Lawler, Werner et Smirnov. Elles reposent de façon essentielle sur une famille à un paramètre de mesures de probabilités sur les courbes du plan complexe vérifiant une propriété d'invariance conforme, qui ont été construites par Oded Schramm. Il a été établi rigoureusement que certains modèles discrets en physique statistique admettent effectivement une limite continue qui peut être décrite par ces mesures (pour des valeurs adéquates du paramètre), et des conjectures très précises prédisent que c'est encore le cas pour d'autres modèles importants. En outre, ces mesures permettent de calculer rigoureusement des valeurs des exposants critiques pour le mouvement brownien plan ou la percolation (cf. [8, 9, 17]), et de vérifier ainsi les prédictions des physiciens (voir notamment [5] et [6]). Schramm a construit ces objets en introduisant l'aléa dans une méthode due à Loewner, qui, de façon informelle, permet de coder certaines familles

¹Symanzik [18] a cherché à rendre rigoureux ce conditionnement, mais son programme conduit à une loi absolument continue par rapport à la mesure de Wiener, et donc pour laquelle les trajectoires ont toujours des points multiples.

croissantes de compacts du demi-plan à l'aide de courbes réelles. Il les a nommés *Stochastic Loewner Evolution* (SLE) ; il est désormais plus juste d'utiliser ces mêmes initiales pour *Schramm-Loewner Evolution*.

L'objet de cet exposé est de proposer une présentation succincte et non technique de ce thème, l'un des plus importants actuellement en théorie des probabilités. Nous renvoyons au papier de Lawler [7] pour une introduction plus détaillée à SLE, et aux excellentes notes de cours de Werner [19] et aux travaux qui y sont cités pour une présentation complète.

2. L'ÉQUATION DE LOEWNER ET SA VERSION STOCHASTIQUE

2.1. L'équation de Loewner

Notons $\mathbb{H} := \{z \in \mathbb{C} : \text{Im}z > 0\}$ le demi-plan supérieur. Soit $K \subseteq \overline{\mathbb{H}}$ un compact tel que $H := \mathbb{H} \setminus K$ soit simplement connexe ; on dira qu'un tel K est une *cosse* (hull en anglais). Le théorème de Riemann assure l'existence d'applications conformes $g_K : H \rightarrow \mathbb{H}$ avec $g_K(\infty) = \infty$. Le principe de réflexion de Schwarz permet d'étendre analytiquement g_K à $\{z \in \mathbb{C} : z \notin K \text{ and } \bar{z} \notin K\}$. Le développement analytique de g_K au voisinage de ∞ (ce qui revient à développer $z \rightarrow 1/g_K(1/z)$ au voisinage de $z = 0$) s'écrit alors sous la forme

$$g_K(z) = bz + a_0 + a_1z^{-1} + a_2z^{-2} + \dots$$

avec $b > 0$ et $a_j \in \mathbb{R}$. Il y a une famille à 2 paramètres de telles applications conformes g_K (observer que pour tous $c > 0$ et $d \in \mathbb{R}$, l'application $z \rightarrow g_K(cz) + d$ vérifie les mêmes propriétés), et on convient de choisir la normalisation, parfois dite hydrodynamique, pour laquelle $b = 1$ et $a_0 = 0$. Autrement dit, on a

$$g_K(z) = z + a_1/z + O(z^{-2}), \quad z \rightarrow \infty.$$

Le coefficient $a_1 := a(K)$ s'appelle la capacité de K , il a une interprétation probabiliste simple : considérons un mouvement brownien complexe W issu de iy , et notons τ le premier temps d'atteinte de $K \cup \mathbb{R}$. Alors il est facile de voir que

$$a(K) = \lim_{y \rightarrow \infty} y \mathbb{E}^{iy}(\text{Im}(W_\tau)).$$

L'observation élémentaire suivante va jouer un rôle important dans la suite : si $J \subseteq K$ sont deux cosses, correspondant aux applications conformes normalisées g_J et g_K , et si on note $L = \overline{g_J(K \setminus J)}$, alors il est facile de vérifier la formule dite de subordination

$$(1) \quad g_K = g_L \circ g_J.$$

En particulier, on en déduit l'identité entre capacités

$$a(K) = a(J) + a(L).$$

De façon informelle, ceci montre que la croissance des compacts de \mathbb{H} est très naturellement liée à la composition des applications conformes.

Loewner s'est intéressé à la situation suivante. On se donne une courbe continue simple $\gamma : [0, \infty[\rightarrow \overline{\mathbb{H}}$ issue de $\gamma(0) = 0$, avec $\gamma(t) \in \mathbb{H}$ pour $t > 0$, et on prend $K_t := \gamma[0, t]$, de sorte que $(K_t, t \geq 0)$ est une famille de compacts simples de $\overline{\mathbb{H}}$ qui croît continûment. On suppose que $\lim_{t \rightarrow \infty} a(\gamma[0, t]) = \infty$, et pour simplifier, on se ramène au cas où la paramétrisation de γ est choisie de sorte que $a(K_t) = 2t$. Pour chaque $t \geq 0$, on posera $g_t := g_{K_t}$. Les applications conformes g_t permettent de transformer la courbe γ à valeurs dans \mathbb{H} en une courbe $\omega : [0, \infty[\rightarrow \mathbb{R}$ définie par

$$\omega(t) := g_t(\gamma(t)), \quad t \geq 0.$$

Loewner a établi une relation très simple entre les applications conformes g_t et la courbe réelle ω (voir par exemple [1]).

PROPOSITION 2.1. — *Pour tout $z \in \overline{\mathbb{H}}$, $g_t(z)$ vérifie l'équation différentielle*

$$(2) \quad \partial_t g_t(z) = \frac{2}{g_t(z) - \omega(t)}, \quad g_0(z) = z.$$

Cette équation est satisfaite jusqu'au premier temps T_z en lequel $g_t(z) = 0$.

Il est également naturel de prendre un autre point de vue. Au lieu de partir d'une courbe simple γ dans $\overline{\mathbb{H}}$ et de construire la courbe réelle ω , on se donne une fonction $\omega : [0, \infty[\rightarrow \mathbb{R}$ avec $\omega(0) = 0$, puis on construit pour chaque $z \in \overline{\mathbb{H}}$ une fonction $t \rightarrow g_t(z)$ en résolvant l'équation (2). Cette dernière est bien définie jusqu'à ce que $g_t(z) - \omega(t)$ atteigne 0, temps que l'on note T_z . Introduisons encore

$$(3) \quad K_t = \{z \in \overline{\mathbb{H}} : T_z \leq t\}.$$

On appelle $(K_t, t \geq 0)$ la *chaîne de Loewner* associée à la fonction ω . On a alors le résultat suivant.

PROPOSITION 2.2. — *Avec les notations et hypothèses précédentes, pour tout $t > 0$, g_t est la transformation conforme normalisée de $\mathbb{H} \setminus K_t$ vers \mathbb{H} .*

REMARQUE 2.3. — Lorsque la fonction ω est suffisamment régulière (plus précisément, Hölderienne d'ordre $1/2$ avec un coefficient de Hölder assez petit), la fonction réciproque g_t^{-1} se prolonge continûment à $\omega(t) \in \partial\mathbb{H}$ (voir [13]). On peut alors récupérer une courbe γ par l'identité $\gamma(t) = g_t^{-1}(\omega(t))$, de sorte que les compacts K_t sont donnés par des courbes simples $\gamma : [0, \infty[\rightarrow \overline{H}$. Mais en général, K_t n'est pas nécessairement donné par une courbe simple.

2.2. SLE cordal

Schramm [15] a introduit la notion d'évolution de Loewner stochastique en prenant pour la fonction ω dans les propositions 2.1–2.2, un mouvement brownien réel de variance $\kappa > 0$, i.e.

$$\omega(t) = W_t = \sqrt{\kappa} B_t,$$

où $(B_t, t \geq 0)$ est un mouvement brownien standard. Pour un probabiliste, ce choix est bien sûr très naturel, car les mouvements browniens sont les seuls processus aléatoires à trajectoires continues, qui vérifient une propriété d'invariance par changement d'échelle, et dont les accroissements sont indépendants.

DÉFINITION 2.4 (Schramm). — *Pour $\kappa > 0$, on appelle SLE_κ cordal la chaîne de Loewner associée à un mouvement brownien $(W_t, t \geq 0)$ de variance κ .*

Le qualificatif *cordal* fait référence au fait que la famille de compacts $(K_t, t \geq 0)$ croît entre deux points de la frontière $\partial\mathbb{H}$ du domaine, 0 et ∞ . Schramm a construit également des processus SLE radiaux, i.e. pour lesquels les compacts croissent d'un point de la frontière jusqu'à atteindre un point intérieur donné. Dans cet exposé, nous nous concentrerons sur le cas cordal, à l'exception de la section 4.4.

Sans entrer dans les détails, ayant défini $(K_t, t \geq 0)$, SLE_κ cordal dans \mathbb{H} croissant de 0 vers ∞ , pour tout domaine simplement connexe $D \neq \mathbb{C}$, et A, B deux points distincts de la frontière de D , on construit SLE_κ dans D croissant de A vers B en considérant $(g(K_t), t \geq 0)$, où $g : \mathbb{H} \rightarrow D$ désigne une application conforme avec $g(0) = A$ et $g(\infty) = B$.

Les propriétés de base du mouvement brownien se traduisent aisément pour SLE :

- Le brownien est symétrique, i.e. W et $-W$ ont la même loi ; il en découle que si \tilde{K} désigne le symétrique de K par rapport à l'axe imaginaire, alors $(K_t, t \geq 0)$ et $(\tilde{K}_t, t \geq 0)$ ont la même loi.
- Invariance par changement d'échelle, i.e. pour tout $c > 0$, $(W_t, t \geq 0)$ et $(c^{-1/2}W_{ct}, t \geq 0)$ ont la même loi. Il s'ensuit que $(K_{ct}, t \geq 0)$ et $(cK_t, t \geq 0)$ ont la même loi.
- Propriété de Markov : si T est un temps d'arrêt dans $(\mathcal{F}_t)_{t \geq 0}$, la filtration naturelle de W , alors $(W_{T+t} - W_T, t \geq 0)$ est un mouvement brownien indépendant de \mathcal{F}_T . Il s'ensuit que le processus $(g_{T+t}(K_{T+t} \setminus K_T) - W_T, t \geq 0)$ est également indépendant de \mathcal{F}_T et est encore un SLE_κ .

La propriété suivante, due à Rohde et Schramm [14], fait appel à des propriétés plus fines du brownien.

PROPOSITION 2.5. — (i) *Pour $\kappa \in]0, 4]$, SLE_κ est une courbe simple p.s.*

(ii) *Pour $\kappa > 4$, il existe une courbe aléatoire continue γ qui n'est pas simple, mais qui ne se croise pas elle-même, telle que pour tout $t \geq 0$, $\mathbb{H} \setminus K_t$ coïncide avec la composante connexe infinie de $\mathbb{H} \setminus \gamma[0, t]$. Autrement dit, K_t est la réunion de $\gamma[0, t]$ et de l'intérieur de ses boucles.*

REMARQUE 2.6. — Il est intéressant de rappeler que la trajectoire brownienne n'est pas Hölderienne d'ordre $1/2$ (mais presque : localement, elle est Hölderienne d'ordre $1/2 - \varepsilon$ pour tout $\varepsilon > 0$), et de comparer le résultat (i) avec la remarque qui suit la proposition 2.2.

On voit clairement dans cet énoncé à quel point le comportement de SLE dépend de la valeur du paramètre κ , ce qui peut paraître a priori assez surprenant. Dans cette direction, mentionnons un résultat récent de Beffara [2] qui a montré que la dimension de Hausdorff de la courbe γ permettant de représenter SLE_κ vaut $\min(2, 1 + \kappa/8)$, presque sûrement (sauf peut-être pour $\kappa = 4$).

On va maintenant présenter des propriétés cruciales de SLE_κ pour des valeurs particulières de κ , propriétés qui auront un rôle essentiel dans l'analyse de la limite continue de certains modèles de physique statistique.

3. PROPRIÉTÉS DE LOCALITÉ ET DE RESTRICTION

3.1. Localité pour $\kappa = 6$

On s'intéresse tout d'abord à l'image de SLE_κ par des applications conformes. Considérons une cosse A qui ne contient pas l'origine. Rappelons que g_A désigne l'application conforme normalisée de $\mathbb{H} \setminus A$ sur \mathbb{H} , et posons

$$\tilde{K}_t = g_A(K_t) \quad \text{pour } t < T := \inf \{s : K_s \cap A \neq \emptyset\} .$$

On a défini ainsi une nouvelle famille croissante de compacts $(\tilde{K}_t, t < T)$; il est important d'observer que la paramétrisation est donnée en fonction de la capacité de $K_t = g_A^{-1}(\tilde{K}_t)$ et non pas directement de \tilde{K}_t .

On peut montrer que la famille $(\tilde{K}_t, t < T)$ est elle aussi, à un changement de temps près, une chaîne de Loewner, qu'on peut donc représenter sous la forme des propositions 2.1–2.2. Plus précisément, si on note \tilde{g}_t l'application conforme normalisée de $g_A(\mathbb{H} \setminus (A \cup K_t))$ dans \mathbb{H} , et si on note $\alpha(t) = a(A \cup K_t) = a(A) + a(\tilde{K}_t)$, alors on obtient à partir de (2) l'équation différentielle

$$\partial \tilde{g}_t(z) = \frac{\partial \alpha(t)}{\tilde{g}_t(z) - \tilde{W}_t} ,$$

où $\tilde{W}_t = h_t(W_t)$ et h_t est l'application conforme normalisée de $g_t(\mathbb{H} \setminus (A \cup K_t))$ dans \mathbb{H} .

Bien sûr, \tilde{W} n'est plus un mouvement brownien, mais à l'aide de la formule d'Itô (i.e. la formule de changement de variables du calcul différentiel pour le mouvement brownien), permet de vérifier que \tilde{W} satisfait l'équation différentielle stochastique

$$d\tilde{W}_t = h'_t(W_t)dW_t + \left(\frac{\kappa}{2} - 3\right) h''_t(W_t)dt .$$

On voit sur cette expression que quelque chose de remarquable se passe lorsque $\kappa = 6$, car alors le terme de dérive est identiquement nul, de sorte que

$$\tilde{W}_t = \int_0^t h'_s(W_s)dW_s ,$$

où le terme de droite est une intégrale stochastique d'Itô. On sait qu'une telle intégrale peut être représentée sous la forme d'un mouvement brownien changé de temps. Plus

précisément, on pose $\langle \tilde{W} \rangle_t = \int_0^t h'_s(W_s)^2 ds$, et on définit implicitement \hat{W} par $\hat{W}_{\langle \tilde{W} \rangle_t} = \tilde{W}_t$. Alors \hat{W} est un mouvement brownien de variance κ ($= 6$). Par ailleurs, on peut vérifier également que $\langle \tilde{W} \rangle_t = \frac{1}{2}(\alpha(t) - \alpha(0))$, de sorte que, finalement, si on définit implicitement \hat{g}_t par $\tilde{g}_t = \hat{g}_{\alpha(t)/2}$, alors

$$\partial_t \hat{g}_t(z) = \frac{2}{\hat{g}_t(z) - \hat{W}_t}.$$

On a donc le théorème suivant, dû à Lawler, Schramm et Werner [8, 12] :

THÉORÈME 3.1 (Propriété de localité). — *Désignons par \tilde{T} le premier temps en lequel K_t atteint $g_t(\partial A)$. Pour $\kappa = 6$, à un changement de temps près, les processus $(\tilde{K}_t - g_A(0), t < T)$ et $(K_t, t < \tilde{T})$ ont la même loi.*

3.2. Restriction pour $\kappa = 8/3$

Si on applique les mêmes techniques pour étudier $h'_t(W_t)$ au lieu de $h_t(W_t)$ comme précédemment, on obtient par application de la formule d'Itô que

$$dh'_t(W_t) = h''_t(W_t)dW_t + \left(\frac{h''_t(W_t)}{2h'_t(W_t)} + (\kappa/2 - 4/3)h'''_t(W_t) \right) dt.$$

Cette fois, c'est la valeur $\kappa = 8/3$ qui joue un rôle remarquable ; en effet, on voit qu'alors

$$d[h'_t(W_t)^{5/8}] = \frac{5h''_t(W_t)}{8h'_t(W_t)^{3/8}}dW_t,$$

de sorte que $h'_t(W_t)^{5/8}$ est une martingale locale. Ceci permet de calculer les probabilités d'atteinte pour $SLE_{8/3}$:

PROPOSITION 3.2. — *Soit $(K_t, t \geq 0)$ un $SLE_{8/3}$ cordal. Alors pour toute cosse A qui ne contient pas l'origine,*

$$\mathbb{P}(\forall t \geq 0, K_t \cap A = \emptyset) = g'_A(0)^{5/8},$$

où $g_A : \mathbb{H} \setminus A \rightarrow A$ est l'application conforme normalisée.

Donnons juste l'esquisse de la preuve. Notons T le premier temps en lequel K_t atteint A ($T = \infty$ lorsque K_t n'atteint jamais A). On vérifie d'abord que $M_t := h'_t(W_t)^{5/8}$ reste compris entre 0 et 1 pour tout $t \in [0, T]$, et converge vers 1 sur l'événement $T = \infty$ (rappelons que la normalisation des applications conformes impose $h'_t(\infty) = 1$) ; par ailleurs, on montre que $\lim_{t \rightarrow T^-} h'_t(W_t) = 0$ lorsque $T < \infty$. Le théorème d'arrêt pour les martingales donne alors

$$\mathbb{P}(T = \infty) = \mathbb{E}(M_T) = \mathbb{E}(M_0) = h'_0(0)^{5/8} = g'_A(O)^{5/8}.$$

Une conséquence importante de ce calcul explicite est la propriété dite de restriction, vérifiée par $SLE_{8/3}$ (cf. [12]).

THÉORÈME 3.3 (Propriété de restriction). — Soit $A \subset \mathbb{H}$ une cosse qui ne contient pas 0, et $(K_t, t \geq 0)$ un $SLE_{8/3}$ cordal. Alors la loi conditionnelle de K_∞ sachant $K_\infty \cap A = \emptyset$ est la même que celle de $f_A^{-1}(K_\infty)$, où f_A^{-1} désigne la réciproque de l'application conforme $f_A : \mathbb{H} \setminus A \rightarrow \mathbb{H}$ donnée par $f_A(z) = g_A(z) - g_A(0)$.

Donnons encore seulement l'esquisse de la preuve. Soit $B \subset \mathbb{H}$ une cosse qui ne contient pas l'origine, et disjoint de $f_A(A)$. Il s'agit de calculer la probabilité conditionnelle

$$\mathbb{P}(f_A(K_\infty) \cap B = \emptyset \mid K_\infty \cap A = \emptyset) = \frac{\mathbb{P}(K_\infty \cap (f_A^{-1}(B) \cup A) = \emptyset)}{\mathbb{P}(K_\infty \cap A = \emptyset)}.$$

Si on note $C = A \cup f(B)$, alors on sait d'après (1) que $f_C = f_B \circ f_A$, puisque $B = f_A(f_A^{-1}(B))$. On déduit de la proposition 3.2 que le quotient ci-dessus vaut

$$\left(\frac{f'_C(0)}{f'_A(0)} \right)^{5/8} = \left(\frac{f'_A(0)f'_B(f_A(0))}{f'_A(0)} \right)^{5/8} = f'_B(0)^{5/8}.$$

En appliquant à nouveau la proposition 3.2, on peut identifier cette dernière quantité comme $\mathbb{P}(K_\infty \cap B = \emptyset)$. Autrement dit, les probabilités d'atteinte pour $f_A(K_\infty)$ sachant que $K_\infty \cap A = \emptyset$ sont les mêmes que pour K_∞ , ce qui entraîne aisément l'assertion.

On peut montrer par ailleurs que $SLE_{8/3}$ cordal est la seule mesure de probabilités sur les courbes continues simples de 0 à ∞ dans \mathbb{H} pour laquelle cette propriété de restriction est vraie. Ceci est à mettre en rapport avec le fait que la propriété de restriction est vérifiée également par la courbe brownienne. Plus précisément, on considère un mouvement brownien β dans \mathbb{H} de 0 à ∞ (ce qui est facile à définir rigoureusement), et une cosse $A \subset \mathbb{H}$ qui ne contient pas l'origine. On note $f_A : \mathbb{H} \setminus A \rightarrow \mathbb{H}$ l'application conforme avec $f_A(0) = 0$ et $f_A(z) \sim z$ lorsque $z \rightarrow \infty$, et f_A^{-1} son inverse. Alors $f_A^{-1}(\beta)$ a, à un changement de temps près, la même loi que β conditionné à éviter A . Bien sûr, la courbe brownienne n'est pas simple (elle a même presque sûrement des points de multiplicité infinie) ; on ne peut donc pas identifier la courbe de β et $SLE_{8/3}$. Cependant, de façon tout à fait informelle, si on conditionne β à ne pas avoir de points multiples, ce conditionnement devrait préserver la propriété de restriction, et donc on peut s'attendre à ce que la courbe brownienne dans \mathbb{H} de 0 à ∞ et conditionnée à être simple, soit décrite par $SLE_{8/3}$. Toujours de façon informelle, la courbe brownienne conditionnée à être simple est sensée apparaître à la limite continue de la marche aléatoire auto-évitante (déjà mentionnée dans l'introduction). Autrement dit, il est naturel de faire la conjecture que la limite continue de la marche aléatoire auto-évitante existe et est décrite par $SLE_{8/3}$. Cette conjecture est toujours ouverte pour l'instant ; elle a été confirmée par des simulations numériques.

Dans la dernière partie de cet exposé, nous allons présenter rapidement des situations dans lesquelles il a été établi de façon rigoureuse que SLE permet de décrire des courbes aléatoires remarquables du plan.

4. SLE ET QUELQUES COURBES ALÉATOIRES REMARQUABLES

4.1. $SLE_{8/3}$ et excursions browniennes

Appelons excursion brownienne dans \mathbb{H} un mouvement brownien complexe $Z = X + iY$ à valeurs dans \mathbb{H} , issu de $Z_0 = 0$, et conditionné à tendre vers ∞ . Un tel conditionnement est seulement formel (car la probabilité qu'un mouvement brownien complexe reste dans \mathbb{H} est nulle), mais il est aisé à rendre rigoureux. Plus précisément, les coordonnées X et Y sont deux processus indépendants, X est un mouvement brownien réel standard, et Y un processus de Bessel de dimension 3, c'est-à-dire la norme euclidienne d'un mouvement brownien dans \mathbb{R}^3 .

Soit $A \subset \overline{\mathbb{H}}$ une cosse qui ne contient pas l'origine, et $g_A : \mathbb{H} \setminus A \rightarrow \mathbb{H}$ l'application conforme normalisée. On peut montrer que la probabilité que la trajectoire de Z n'atteigne jamais A est donnée par

$$(4) \quad \mathbb{P}(Z_t \notin A, \forall t \geq 0) = g'_A(0).$$

Esquissons la preuve de cette identité : Travaillons d'abord avec un mouvement brownien dans \mathbb{H} issu de $z \in \mathbb{H}$, et conditionné à tendre vers ∞ . La fonction $\Im g_A$ est harmonique dans $\mathbb{H} \setminus A$, et il en découle que le processus

$$\frac{\Im g_A(Z_t)}{Y_t}, \quad t < T_A := \inf \{t \geq 0, Z_t \in A\}$$

est une martingale. Cette martingale converge vers 1 sur l'événement $T_A = \infty$ (puisque $g_A(z) \sim z$ quand $z \rightarrow \infty$), et vers 0 quand $T_A < \infty$. Par application du théorème d'arrêt, on a donc

$$\mathbb{P}(T_A = \infty \mid Z_0 = z) = \frac{\Im g_A(z)}{\Im z},$$

et il ne reste alors qu'à faire tendre z vers 0.

La formule (4) est bien sûr à rapprocher de la proposition 3.2. On en tire immédiatement le résultat suivant (voir [12]).

THÉORÈME 4.1. — *Soit \mathcal{H}_8 la frontière extérieure de l'union de 8 $SLE_{8/3}$ cordaux indépendants. Soit \mathcal{B}_5 la frontière extérieure de l'union de 5 excursions browniennes indépendantes. Alors \mathcal{H}_8 et \mathcal{B}_5 ont la même loi.*

En effet, les probabilités que ces différentes courbes ne rencontrent pas une cosse A sont les mêmes.

Ce théorème a plusieurs conséquences intéressantes. En particulier, si on le relie à la conjecture de la section précédente, qui énonce que $SLE_{8/3}$ est la limite continue de la marche aléatoire auto-évitante, on peut alors observer que localement, la frontière extérieure de la courbe brownienne dans le plan peut être vue comme la limite continue d'une marche aléatoire auto-évitante (propriété qui avait été énoncée de façon informelle par Mandelbrot).

4.2. SLE_6 et percolation critique

L'un des résultats mathématiques les plus remarquables obtenus récemment en ce qui concerne le comportement asymptotique de modèles discrets de la physique statistique est dû à Smirnov [16]. Il porte sur la *percolation sur le réseau triangulaire*.

Commençons par considérer un réseau régulier dans le plan (pas nécessairement le réseau triangulaire). On fixe un paramètre $p \in]0, 1[$, et on décide d'ouvrir chaque site de ce réseau avec probabilité p , indépendamment les uns des autres. On s'intéresse aux composantes connexes (amas) formées par les sites ouverts. Il existe un paramètre critique p_c tel que, presque sûrement, pour $p \leq p_c$, il n'y a aucun amas de taille infinie, alors que pour $p > p_c$, il y a exactement un amas infini.

Le paramètre critique p_c dépend du réseau ; sa valeur exacte n'est pas connue en général. Il y a cependant une exception remarquable : dans le cas du réseau triangulaire, Kesten et Wierman ont montré que $p_c = 1/2$. De façon très informelle et rapide, disons que des arguments de renormalisation suggèrent que, quand on se place dans une grande échelle et pour la valeur critique du paramètre, ce qu'on observe de la percolation ne dépend pas du réseau. Par ailleurs, si on admet l'existence d'une limite continue de la percolation, cette limite doit nécessairement être invariante par changement d'échelle et par rotation (puisque la limite ne dépend pas du réseau). Ceci conduit naturellement à la conjecture plus forte que la percolation continue est invariante conforme, i.e. les connexions respectives dans des domaines simplement connexes devraient avoir la même loi à des transformations conformes près.

En particulier, la probabilité qu'il existe un amas permettant de traverser de gauche à droite un rectangle de longueur Na et de largeur Nb doit converger quand N tend vers l'infini, vers une limite $F(b/a)$ qui ne dépend que du rapport b/a , et où la fonction F est « universelle » (i.e. ne dépend pas du réseau). Cardy [3], à l'aide de considérations provenant de la théorie conforme des champs, a prédit très précisément ce que devait être cette fonction F . Par la suite Carleson a observé qu'on pouvait reformuler de façon beaucoup plus simple la conjecture de Cardy en considérant la percolation non pas dans un rectangle, mais dans un triangle équilatéral ABC . Plus précisément, si on considère un point $X \in [CA]$, la probabilité qu'il existe un amas pour la percolation critique dans le triangle ABC qui relie $[AB]$ à $[CX]$ vaut CX/AB .

Smirnov [16] a démontré ces conjectures (existence d'une limite continue, invariance conforme et formule de Cardy) pour la percolation critique sur le réseau triangulaire. De plus, SLE_6 permet de décrire l'amas de percolation de la façon suivante. Pour faciliter la présentation, il est commode de voir le réseau triangulaire comme le pavage du plan en « nid d'abeille », i.e. en identifiant chaque site du réseau avec un hexagone dont il est le centre. On se donne un domaine simplement connexe $D \neq \mathbb{C}$, et deux points distincts a et b sur la frontière. On fixe un petit paramètre $\delta > 0$, on considère D_δ une approximation simplement connexe de D obtenue par réunion de petits hexagones de taille δ , et on note

respectivement a_δ et b_δ les points du réseau hexagonal sur la frontière de D_δ les plus proches de a et de b . Les cellules de la frontière de D_δ entre a_δ et b_δ sont coloriées en blanc, celles entre b_δ et a_δ en noir (avec l'orientation trigonométrique). On colorie ensuite chaque hexagone de l'intérieur de D_δ en blanc avec probabilité $1/2$, indépendamment les uns des autres, et on décide que les hexagones non coloriés sont noirs.

L'interface partant de a_δ et finissant en b_δ , qui sépare les cellules blanches des cellules noires, définit une courbe aléatoire γ_δ qui parcourt la frontière extérieure de l'amas de percolation contenant les points a et b . On l'appelle le processus d'exploration. Le théorème de Smirnov entraîne que quand $\delta \rightarrow 0+$, γ_δ converge en loi vers une courbe aléatoire dans \overline{D} qui débute en a et finit en b . La propriété d'invariance conforme pour la percolation critique se transpose à la courbe limite ; plus précisément on montre que cette dernière vérifie la propriété de localité (cf. Théorème 3.1). Enfin, la formule de Cardy-Smirnov permet de calculer les probabilités d'atteinte de cosses, et de vérifier ainsi qu'elles coïncident avec celles pour SLE_6 cordal dans D , allant de a à b . Il est intéressant d'observer que la courbe limite γ n'est pas simple (cf. Proposition 2.5), et donc ne décrit pas seulement la frontière extérieure de l'amas de percolation, mais également une partie de la frontière intérieure qui affleure la frontière extérieure.

Pour conclure cette section, il convient de mentionner que SLE_6 permet également de décrire les points déconnectés de l'infini par la trajectoire d'un mouvement brownien réfléchi, disons dans \mathbb{H} , où l'angle de réflexion vaut $\pi/3$ sur la frontière $]0, \infty[$ et $2\pi/3$ sur $] - \infty, 0[$ (en particulier, la réflexion tend à éloigner de l'origine). Ceci permet de relier la frontière extérieure d'un amas de percolation à la frontière extérieure d'une trajectoire brownienne, et donc à $SLE_{8/3}$ d'après la section précédente.

4.3. SLE_8 et arbre couvrant aléatoire

Donnons-nous comme précédemment un domaine simplement connexe $D \neq \mathbb{C}$. Fixons un petit paramètre $\delta > 0$ et notons cette fois D_δ une approximation (connexe) de D par une portion du réseau $\delta\mathbb{Z}^2$. On appelle *arbre aléatoire couvrant* dans D_δ un sous-graphe (donné par des sommets et des arêtes) connexe qui passe par tous les sommets de D_δ et qui ne contient pas de boucle. Donnons-nous deux points distincts a et b sur la frontière de D . Soient respectivement a_δ et b_δ les sommets de D_δ les plus proches de a et b . On s'intéresse aux arbres couvrants qui contiennent toutes les arêtes sur la frontière de D_δ entre a_δ et b_δ (dans le sens trigonométrique). On peut coder de façon unique un tel arbre couvrant par une courbe de Peano (i.e. qui passe par tous les sommets du graphe) : On note η^d la courbe qui parcourt l'arbre en partant de a_δ et en suivant les branches dans l'ordre anti-trigonométrique jusqu'à atteindre b_δ (si on suit ensuite la frontière de D_δ de b_δ à a_δ toujours dans le sens anti-trigonométrique, on a ainsi entièrement parcouru l'arbre couvrant).

Lorsqu'on choisit l'arbre couvrant au hasard suivant la probabilité uniforme, la courbe de Peano η^δ est prise elle aussi au hasard suivant l'équiprobabilité sur l'ensemble des courbes de Peano dans D_δ partant de a_δ et finissant en b_δ . À l'aide d'arguments délicats, dont certains reposent sur l'algorithme de Wilson (qui permet de construire des arbres aléatoires couvrants à partir de trajectoires de marches aléatoires), Lawler, Schramm et Werner [10] ont établi le résultat suivant.

THÉORÈME 4.2. — *Quand $\delta \rightarrow 0$, la courbe de Peano discrète η^δ converge en loi vers une courbe de Peano aléatoire continue η , telle que, à un changement de temps près, $(\eta[0, t], t \geq 0)$ a la même loi que SLE_8 cordal dans D de a à b .*

4.4. SLE_2 radial et marche aléatoire à boucles effacées

Le dernier exemple que nous présenterons est très lié au précédent ; il fait intervenir un autre type de chaîne de Loewner, dite *radiale* (voir par exemple [4]), qui correspond à une famille de compacts croissant de la frontière vers un point intérieur du domaine.

Notons d'abord \mathbb{U} le disque unité, et considérons une courbe continue simple $\gamma : \mathbb{R}_+ \rightarrow \overline{\mathbb{U}}$ avec $\gamma_0 = 1$, $\gamma_t \in \mathbb{U}$ pour tout $t > 0$ et $\lim_{t \rightarrow \infty} \gamma_t = 0$. Pour chaque $t \geq 0$, on note f_t l'application conforme de $U_t := \mathbb{U} \setminus \gamma[0, t]$ dans \mathbb{U} , normalisée par $f_t(0) = 0$ et $f_t(\gamma_t) = 1$. Pour simplifier, on supposera que la paramétrisation de γ est telle que $|f_t'(0)| = e^t$, ce qui n'induit pas de perte de généralité. Définissons ensuite

$$\zeta_t = \left(\frac{f_t'(0)}{|f_t'(0)|} \right)^{-1},$$

de sorte que ζ décrit une courbe sur le cercle $\partial\mathbb{U}$ telle que

$$g_t(z) := \zeta_t f_t(z)$$

soit l'application conforme de U_t dans \mathbb{U} telle que $g_t(0) = 0$ et $g_t'(\gamma_t) = e^t$. Noter également que $g_t(\gamma_t) = \zeta_t$.

Loewner a établi qu'on pouvait récupérer g_t à partir de ζ_t en résolvant l'équation différentielle

$$(5) \quad \partial g_t(z) = -g_t(z) \frac{g_t(z) + \zeta_t}{g_t(z) - \zeta_t}$$

tant que $z \in U_t$, puis γ_t par la formule $\gamma_t = g_t^{-1}(\zeta_t)$, au moins dès qu'elle a un sens.

Comme dans la section 2, on prend un point de vue différent en se donnant a priori une courbe ζ sur le cercle unité, et en construisant des applications conformes $z \rightarrow g_t(z)$ en résolvant (5) jusqu'au temps $T(z)$ en lequel $g_t(z)$ atteint ζ_t . On définit ensuite la chaîne de Loewner radiale

$$K_t := \{z \in \overline{\mathbb{U}} : T(z) \leq t\}.$$

Pour tout $\kappa > 0$, on appelle SLE_κ radial (croissant de 1 à 0) la chaîne de Loewner aléatoire obtenue lorsque ζ est un mouvement brownien sur le cercle de variance κ .

On va maintenant conclure cet exposé en présentant un dernier résultat de Lawler, Schramm et Werner [10] (voir également Schramm [15]) qui relie SLE_2 radial et la limite continue de la *marche aléatoire simple à boucles effacées*.

Pour cela, considérons un domaine simplement connexe $D \neq \mathbb{C}$ contenant 0, et fixons un petit paramètre δ . On introduit une marche aléatoire simple symétrique sur $\delta\mathbb{Z}^2$ jusqu'à ce qu'elle sorte de D , et on note η^δ la courbe dans D obtenue en effaçant les boucles de la marche dans l'ordre inverse de leur apparition (i.e. on retourne la marche quand elle sort de D , et on efface au fur et à mesure les boucles de marche retournée).

Soit encore $\Phi : \mathbb{U} \rightarrow D$ une application conforme avec $\Phi(0) = 0$ et η la valeur terminale d'un SLE_2 radial dans \mathbb{U} , partant d'un point uniformément distribué sur le cercle unité. Alors η^δ converge en loi quand δ tend vers 0 vers $\Phi(\eta)$.

BIBLIOGRAPHIE

- [1] L.V. Ahlfors. *Conformal Invariants, Topics in Geometric Function Theory*. McGraw-Hill, 1973.
- [2] V. Beffara. The dimension of the SLE curves. Prépublication.
- [3] J. Cardy. Critical percolation in finite geometries. *J. Phys.* **A** (1992), L201-L206.
- [4] J.B. Conway. *Functions of one complex variable. II*. Springer-Verlag, New York, 1995.
- [5] B. Duplantier et K.-H. Kwon. Conformal invariance and intersections of random walks. *Phys. Rev. Lett.* **61** (1988), 2514-2517.
- [6] C. Itzykson et J.-M. Drouffe. *Statistical field theory. Vol. 2*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1989.
- [7] G.F. Lawler. An introduction to the Stochastic Loewner Evolution. Prépublication. <http://www.math.duke.edu/~jose/esi.html>
- [8] G.F. Lawler, O. Schramm et W. Werner. Values of Brownian intersection exponents I: Half-plane exponents et II: Plane exponents. *Acta Mathematica* **187** (2001), 237-273 et 275-308.
- [9] G.F. Lawler, O. Schramm et W. Werner. Values of Brownian intersection exponents III: Two-sided exponents *Ann. Inst. Henri Poincaré* **38** (2002), 109-123.
- [10] G.F. Lawler, O. Schramm et W. Werner. Conformal invariance of planar loop-erased random walks and uniform spanning trees. À paraître dans *Ann. Probab.*
- [11] G.F. Lawler, O. Schramm et W. Werner. On the scaling limit of self-avoiding walks. À paraître dans *Fractal geometry and application, A jubilee of Benoît Mandelbrot*, AMS Proc. Symp. Pure Math.
- [12] G.F. Lawler, O. Schramm et W. Werner. Conformal restriction properties. The chordal case. *J. Amer. Math. Soc.* **16** (2003), 917-955.

- [13] D. Marshall et S. Rohde. The Loewner differential equation and slit mappings. Prépublication.
- [14] S. Rohde et O. Schramm. Basic properties of SLE. À paraître dans *Ann. Math.*
- [15] O. Schramm. Scaling limits of loop-erased random walks and uniform spanning trees. *Israel J. Math.* **118** (2001), 221-288.
- [16] S. Smirnov. Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *C. R. Acad.Sc. Paris série 1 Math.* **333** (2001), 239-244.
- [17] S. Smirnov et W. Werner. Critical exponents for two-dimensional percolation. *Math. Res. Lett.* **8** (2001), 729-744.
- [18] K. Symanzik. Euclidean Quantum Field Theory. In *Local Quantum Theory*, Proc. International School of Physics "Enrico Fermi" XLV, Ed. L. Jost (Academic Press, New York 1969), p. 152.
- [19] W. Werner. Lectures on random planar curves and Schramm-Loewner evolutions. À paraître dans *École d'été de probabilités de Saint-Flour 2002*, Springer.

Jean BERTOIN

Laboratoire de Probabilités
et Modèles Aléatoires

et

Institut universitaire de France
Université Pierre et Marie Curie
175, rue du Chevaleret
F-75013 PARIS

E-mail : `jbe@ccr.jussieu.fr`

RÉSULTATS RÉCENTS SUR LA LIMITE INCOMPRESSIBLE

par Isabelle GALLAGHER

1. INTRODUCTION

Les équations d'Euler compressibles modélisent le mouvement d'un fluide parfait (sans forces de viscosité interne) : elles s'écrivent sous la forme d'un système d'équations aux dérivées partielles d'évolution non linéaires, portant sur le champ de vitesse et la densité du fluide. Ces équations s'obtiennent par minimisation d'une fonctionnelle d'énergie, dans l'espace des difféomorphismes (voir par exemple [2], [9]) : une particule au point x à l'instant t_0 est au point $\psi_1(x)$ au temps t_1 , où ψ_1 est un difféomorphisme. La traduction de cette écriture lagrangienne en variables eulériennes conduit aux équations d'Euler compressibles. Les équations d'Euler incompressibles s'obtiennent quant à elles en se restreignant aux difféomorphismes préservant la mesure : le fluide est incompressible, et ainsi tout ouvert de l'espace est transporté en un ouvert de même volume au cours du mouvement. Le système d'équations aux dérivées partielles obtenu ne porte plus que sur le champ de vitesse (si l'on suppose de plus la densité constante), maintenant de divergence nulle. Le lecteur intéressé par l'obtention de ces équations à partir des lois fondamentales pourra par exemple consulter [40]. Les systèmes des fluides parfaits compressibles et incompressibles seront écrits précisément dans la section 1.1 suivante (systèmes (1) et (2) respectivement).

La question du passage du premier système (Euler compressible) au second (Euler incompressible), dans la limite où un «paramètre de compressibilité» (le nombre de Mach) tend vers zéro, a fait l'objet d'un grand nombre de travaux mathématiques depuis une vingtaine d'années. Les motivations sont nombreuses : mathématiquement tout d'abord, les équations incompressibles sont mieux comprises que les équations compressibles, et il est intéressant d'essayer de transposer à l'un des systèmes des résultats connus pour l'autre, par un argument de comparaison ou de passage à la limite. Physiquement ensuite, on a souvent tendance à considérer qu'un fluide «peu compressible» possède des propriétés semblables à celles d'un fluide incompressible — encore faut-il le justifier (voir

par exemple [63] et [64]). Enfin dans des calculs numériques il est souvent préférable de considérer qu'un fluide est incompressible même lorsqu'il ne l'est pas, pour simplifier l'implémentation (voir à ce propos les travaux [37], [49], [53] et [54]).

Comme nous le verrons ici, ce problème du passage du compressible à l'incompressible fait intervenir des théories mathématiques variées, suivant l'espace physique dans lequel on plonge le fluide : ainsi des idées de l'optique géométrique sont utilisées dans le cas de conditions aux limites périodiques. Pour un fluide dans l'espace entier, ces aspects sont remplacés par des phénomènes dispersifs, et l'on fait appel à des estimations de type «Strichartz», ou encore à des techniques de mesures de défaut et des estimations de décroissance de l'énergie locale.

L'objectif de cet exposé est de donner un aperçu de «l'état de l'art» sur la question de la limite incompressible et de l'implémentation des techniques évoquées ci-dessus : nous verrons les différentes approches suivies depuis les travaux de S. Klainerman et A. Majda au début des années quatre-vingt dans le cas de «données bien préparées» (notion naturelle portant sur la donnée initiale, que nous présentons dans la section 1.1 ci-dessous) jusqu'aux travaux récents de G. Métivier et S. Schochet portant sur les équations non isentropiques (où l'on couple au système d'Euler compressible l'équation de transport de l'entropie).

Le plan de l'exposé est le suivant.

Dans la section 1.1 suivante, nous présentons les équations relatives aux fluides compressibles et incompressibles respectivement, et nous rappelons brièvement les résultats principaux concernant le problème de Cauchy pour ces deux systèmes. Suit dans la section 1.2 la mise en évidence du paramètre de compressibilité dans l'équation compressible ; on étudie alors la limite formelle de ce système (retrouvant ainsi le système incompressible), et les notions de données «bien» et «mal préparées» sont dégagées.

Le paragraphe 2 traite du cas où les données initiales sont bien préparées, au sens défini dans la section 1.2 ; les premiers travaux concernant la limite incompressible se sont attachés à comprendre ce cas, plus simple.

Dans le paragraphe 3, on s'intéresse au cas général des données mal préparées. On commence par l'étude du cas où les équations sont posées dans l'espace entier (sections 3.1.1 et 3.1.2) : le mot clef de ces études est la dispersion. On s'aperçoit immédiatement que le cas de domaines bornés ne peut être résolu par de tels arguments. Le cas des conditions aux limites périodiques est ainsi traité en détail dans la section 3.2.1, alors que le cas de domaines à bords est effleuré dans la section 3.2.2.

Le dernier paragraphe est consacré à des travaux récents de G. Métivier et S. Schochet concernant la limite incompressible dans le cas non isentropique. Les difficultés mises en évidence dans les paragraphes précédents pour le cas isentropique sont amplifiées ici par le transport couplé de l'entropie : le cas de l'espace entier présente déjà de nouvelles difficultés (surmontées) alors que le cas périodique reste encore largement à comprendre.

REMARQUE 1.1. — Le souci de concision nous a amenée à opérer de nombreuses « impasses » dans cette présentation. Ainsi nous avons choisi de ne pas parler de la dimension 1 d'espace (sauf à l'extrême fin de cet exposé pour étudier les équations des fluides non isentropiques), qui relève plutôt de techniques de lois de conservation ; nous renvoyons le lecteur intéressé par les problèmes spécifiques au cas monodimensionnel au livre de P.-L. Lions [41], Chapitre 8.7, et ses références. Notons que pour une étude d'oscillations dans le cas monodimensionnel on pourra aussi consulter l'article [17].

Nous ne parlerons pas non plus d'équations proches des équations d'Euler compressibles comme les équations de l'élasticité ou de la viscoélasticité, ni d'équations de la géophysique ou de la magnéto-hydrodynamique, qui soulèvent des problèmes mathématiques proches de ceux traités ici.

Enfin nous omettrons tout aspect lié à une éventuelle viscosité du fluide : les techniques pour traiter la limite incompressible dans le cas visqueux (passage de Navier–Stokes compressible à Navier–Stokes incompressible) sont bien sûr spécifiques dès que l'on cherche à utiliser l'effet régularisant du laplacien dans l'équation de Navier–Stokes (voir par exemple [10]-[12], [14], [42], [43] ou encore [48]), mais les effets purement « limite incompressible » sont indépendants de la présence ou non de viscosité — du moins tant qu'on ne traite pas de problèmes aux bords, nous y reviendrons en section 3.2.2 (pour d'autres travaux à ce sujet nous renvoyons le lecteur intéressé à [6], [15], [26], [38]).

REMERCIEMENTS. Je souhaite remercier P. Gérard pour de nombreux conseils sur la rédaction de ce texte. J'adresse aussi mes remerciements à R. Danchin et ainsi qu'à G. Métivier pour des discussions sur les questions évoquées ici.

1.1. Présentation des équations et rappels sur le problème de Cauchy

Considérons un fluide compressible parfait évoluant dans un domaine Ω (que nous préciserons) de l'espace d -dimensionnel \mathbb{R}^d , dont la vitesse et la densité en un point $x \in \Omega$ et à un instant $t \in \mathbb{R}^+$ sont notées respectivement $u(t, x)$ (vecteur à d composantes) et $\rho(t, x)$ (scalaire). Ces deux quantités sont reliées par les équations d'Euler compressibles (isentropiques) suivantes :

$$(1) \quad \begin{aligned} \partial_t \rho + \operatorname{div}(\rho u) &= 0 \\ \rho &\geq 0 \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) + \nabla \rho^\gamma &= 0 \end{aligned}$$

avec $\gamma > 1$, et les conditions initiales

$$\rho|_{t=0} = \rho_0 \geq 0 \quad \text{et} \quad u|_{t=0} = u_0.$$

Les équations incompressibles quant à elles s'écrivent ainsi :

$$(2) \quad \partial_t u + u \cdot \nabla u + \nabla p = 0, \quad \operatorname{div} u = 0$$

avec toujours la condition initiale $u|_{t=0} = u_0$. Dans l'équation d'Euler incompressible (2), la fonction scalaire p représente la pression du fluide (c'est une inconnue du système, due à la contrainte d'incompressibilité). Remarquons que la densité a été choisie constante dans (2), égale à un pour simplifier. Dans le cas d'un domaine borné (régulier) Ω , il convient de prescrire dans les deux équations une condition au bord : $u \cdot n = 0$ sur $\partial\Omega$.

Rappelons quelques résultats sur le problème de Cauchy pour ces deux systèmes d'équations. Ce sont des EDP du premier ordre d'apparence très simple mais nos connaissances en sont relativement faibles : sur les équations compressibles par exemple, la théorie classique permet d'associer à une donnée initiale assez régulière une solution unique sur un temps maximal $[0, T_c]$, qui devient discontinue en T_c . Les solutions discontinues (simplement bornées par exemple) vérifiant l'équation au sens des distributions, ne sont ni uniques, ni stables : pour assurer leur stabilité (et parfois leur unicité), il faut imposer des restrictions, dites «conditions d'entropie de Lax» (voir la théorie de Lax [39]). Nous n'allons pas détailler ici cette théorie, mais simplement énoncer l'une de ses conséquences : dès que la dimension est plus grande que deux, il y a très peu d'entropies et donc très peu d'informations sur les solutions. Nous ne parlerons pas du problème de l'apparition de chocs ici, et notre étude partira du théorème fondamental suivant, qui se démontre en symétrisant le système (1) : en définissant la vitesse du son $c \stackrel{\text{déf}}{=} \frac{2}{\gamma-1} \left(\frac{\partial(\rho^\gamma)}{\partial\rho} \right)^{\frac{1}{2}} = \frac{2\sqrt{\gamma}}{\gamma-1} \rho^{\frac{\gamma-1}{2}}$ le système d'Euler compressible (1) devient

$$(3) \quad \begin{aligned} \partial_t c + u \cdot \nabla c + \frac{\gamma-1}{2} c \operatorname{div} u &= 0 \\ \partial_t u + u \cdot \nabla u + \frac{\gamma-1}{2} c \nabla c &= 0 \end{aligned}$$

et la théorie des systèmes symétriques hyperboliques conduit au théorème suivant (voir par exemple [44], et [45] ou [55] pour le cas de domaines à bords).

THÉORÈME 1.2. — *Soit Ω un domaine régulier de l'espace \mathbb{R}^d , et soit s un entier strictement plus grand que $\frac{d}{2} + 1$. Il existe une constante C telle que si (u_0, c_0) sont des éléments de $H^s(\Omega)$ (u_0 s'annulant au voisinage du bord par exemple), alors il existe un temps T et une unique solution (u, c) dans l'espace $C^0([0, T], H^s(\Omega))$ au système (3). En outre on a la minoration suivante sur le temps d'existence : $T \geq \frac{C}{\|(u_0, c_0)\|_{H^s(\Omega)}}$.*

Dans ce théorème (et partout dans ce texte), l'espace $H^s(\Omega)$ désigne l'espace des fonctions de carré intégrable sur Ω , dont toutes les dérivées jusqu'à l'ordre s sont de carré intégrable. Dorénavant nous noterons s_0 le plus petit entier strictement plus grand que $\frac{d}{2} + 1$. Notons que l'hypothèse d'annulation de la donnée initiale au voisinage du bord donnée dans l'énoncé du théorème 1.2 est l'une des conditions de compatibilité possibles (voir par exemple [55]). Notons en outre qu'il existe des solutions qui effectivement deviennent singulières en temps fini. Nous renvoyons le lecteur intéressé au travail de T. Sideris [59].

Concernant le système incompressible, la situation est un peu meilleure, du moins en dimension deux d'espace. En effet, en dimension deux, le tourbillon $\omega \stackrel{\text{d\u00e9f}}{=} \partial_1 u_2 - \partial_2 u_1$ est conserv\u00e9 le long des caract\u00e9ristiques : il v\u00e9rifie l'\u00e9quation de transport $\partial_t \omega + u \cdot \nabla \omega = 0$ gr\u00e2ce \u00e0 la condition de divergence nulle sur le champ de vitesse. Cette remarque permet de montrer le th\u00e9or\u00e8me suivant (voir [9], [40], [46], [62]).

TH\u00c9OR\u00c8ME 1.3. — *Soit Ω un domaine r\u00e9gulier de l'espace \mathbb{R}^2 , et soit $u_0 \in L^2(\Omega)$ dont le tourbillon ω_0 est un \u00e9l\u00e9ment de $L^r(\Omega)$ pour un r\u00e9el r dans l'intervalle $]1, +\infty[$. Alors il existe une solution $u \in C^0(\mathbb{R}^+, L^2(\Omega))$ telle que $\nabla u \in C^0(\mathbb{R}^+, L^r(\Omega))$. Si $r = +\infty$, alors cette solution $u \in C^0(\mathbb{R}^+, L^2(\Omega))$ est unique, et le tourbillon est alors born\u00e9 en temps et en espace.*

En dimension sup\u00e9rieure la situation est moins comprise, et le r\u00e9sultat est le suivant.

TH\u00c9OR\u00c8ME 1.4. — *Soit Ω un domaine r\u00e9gulier de l'espace \mathbb{R}^d , et soit $s \geq s_0$. Si (u_0, c_0) est un \u00e9l\u00e9ment de $H^s(\Omega)$, alors il existe une unique solution maximale $u \in C^0([0, T], H^s(\Omega))$. Si T est fini alors la norme de u dans $H^s(\Omega)$ tend vers $+\infty$ en T .*

Puisque nous souhaitons ici mettre en valeur les ph\u00e9nom\u00e8nes intervenant dans la limite incompressible plus que les probl\u00e8mes li\u00e9s \u00e0 l'\u00e9tude du probl\u00e8me de Cauchy, nous n'avons pas \u00e9nonc\u00e9 tous les r\u00e9sultats dans leur plus grande g\u00e9n\u00e9ralit\u00e9 ni leur plus grande subtilit\u00e9. Nous renvoyons le lecteur int\u00e9ress\u00e9 par exemple aux livres de J.-Y. Chemin [9] ou de P.-L. Lions [40] et [41] pour des r\u00e9sultats beaucoup plus pr\u00e9cis.

1.2. La limite incompressible

Il y a de multiples fa\u00e7ons de mettre en \u00e9vidence le petit param\u00e8tre dans les \u00e9quations des fluides compressibles (3) (voir par exemple [25] ou [35]). Nous pr\u00e9sentons ici la m\u00e9thode d\u00e9taill\u00e9e dans [51]. La limite incompressible \u00e9tant comprise comme la limite o\u00f9 la vitesse du fluide devient n\u00e9gligeable devant la vitesse du son, on commence par remettre la vitesse \u00e0 l'\u00e9chelle, en rempla\u00e7ant u par εu . Le param\u00e8tre ε est le nombre de Mach. La vitesse \u00e9tant la d\u00e9riv\u00e9e de la position d'une particule de fluide par rapport au temps, une particule va donc se d\u00e9placer d'une distance $O(\varepsilon)$ pendant un temps de l'ordre de 1 (ou de m\u00eame d'une distance de l'ordre de 1 en un temps $O(1/\varepsilon)$). Cela sugg\u00e8re de r\u00e9-\u00e9chelonner les variables d'espace x en les rempla\u00e7ant par x/ε (ou encore la variable de temps t en la changeant en εt). En associant le changement d'\u00e9chelle en vitesse \u00e0 l'un ou l'autre de ces changements d'\u00e9chelle spatiale ou temporelle, le syst\u00e8me (3) devient le suivant :

$$\begin{aligned} \partial_t c + u \cdot \nabla c + \frac{\gamma - 1}{2} c \operatorname{div} u &= 0 \\ \partial_t u + u \cdot \nabla u + \frac{\gamma - 1}{2\varepsilon^2} c \nabla c &= 0. \end{aligned}$$

La derni\u00e8re hypoth\u00e8se maintenant consiste \u00e0 supposer que $c(t, x) = c_1 + \varepsilon \tilde{c}(t, x)$, o\u00f9 c_1 est une constante donn\u00e9e. En renommant \tilde{c} en c , il vient finalement le syst\u00e8me des fluides

faiblement compressibles suivant :

$$(4) \quad \begin{aligned} \partial_t c + u \cdot \nabla c + \frac{\gamma - 1}{2} c \operatorname{div} u + \frac{\gamma - 1}{2\varepsilon} c_1 \operatorname{div} u &= 0 \\ \partial_t u + u \cdot \nabla u + \frac{\gamma - 1}{2} c \nabla c + \frac{\gamma - 1}{2\varepsilon} c_1 \nabla c &= 0. \end{aligned}$$

Les données initiales sont $u|_{t=0} = u_0$ et $c|_{t=0} = c_0$. Nous verrons ci-dessous qu'il peut être pertinent de remplacer cette donnée initiale fixe par une famille de données dépendant du paramètre ε . Nous allons dorénavant nous intéresser à ce système, et chercher à en comprendre le comportement quand ε tend vers zéro. Formellement, prendre la limite $\varepsilon \rightarrow 0$ dans (4) revient à annuler les termes ayant le plus grand facteur en ε dans l'équation. On obtient ainsi formellement à la limite

$$(5) \quad \operatorname{div} u = 0, \quad \nabla c = 0,$$

ce qui revient à considérer un fluide incompressible. On s'attend donc à retrouver dans la limite $\varepsilon \rightarrow 0$ les équations des fluides incompressibles. On se retrouve alors immédiatement devant une condition de compatibilité sur la donnée initiale : suivant qu'elle vérifie elle-même la condition (5) ou non, on peut avoir plus ou moins de difficultés à passer à la limite dans le système (4). On est ainsi amené à introduire la notion de donnée « bien » ou « mal préparée », portant sur une famille de données initiales indexée par le paramètre ε .

DÉFINITION 1.5. — Soient Ω un domaine de \mathbb{R}^d et s un réel. Soit $(u_{\varepsilon,0})_{\varepsilon>0}$ une famille de champs de vecteurs et soit $(c_{\varepsilon,0})_{\varepsilon>0}$ une famille de fonctions, toutes deux bornées dans $H^s(\Omega)$. On dit que $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ est « bien préparée » si

$$\lim_{\varepsilon \rightarrow 0} \operatorname{div} u_{\varepsilon,0} = 0 \quad \text{dans } H^s(\Omega) \quad \text{et} \quad \lim_{\varepsilon \rightarrow 0} \nabla c_{\varepsilon,0} = 0 \quad \text{dans } H^{s-1}(\Omega).$$

Dans le cas contraire, la famille $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ est dite « mal préparée ».

REMARQUE 1.6. — Historiquement l'étude du passage des équations des fluides compressibles aux équations des fluides incompressibles a commencé dans le cas de données « bien préparées » au sens de la définition 1.5. Des notions plus fortes de préparation de données peuvent être trouvées dans la littérature (voir par exemple [7]) mais celle proposée ci-dessus est la plus utilisée. Dans le paragraphe suivant nous présentons les méthodes employées pour démontrer des théorèmes de convergence des solutions compressibles vers les solutions incompressibles dans ce cadre « bien préparé ». Les autres paragraphes de cet exposé sont consacrés à l'étude (plus difficile) du cas de données quelconques.

2. LE CAS BIEN PRÉPARÉ

Le premier travail mathématique concernant la limite incompressible est dû à S. Klainerman et A. Majda, dans [35] et [36]. Il s'agit d'étudier le temps d'existence des solutions du système (4) et sa dépendance en ε , ainsi que le passage à la limite dans l'équation.

Dans [35] les auteurs s'appuient sur l'étude du système quasilinear abstrait suivant :

$$(6) \quad \partial_t u_\varepsilon + \sum_{j=1}^d B^j(u_\varepsilon, \varepsilon) \partial_j u_\varepsilon = 0 \quad \text{dans } \mathbb{T}^d$$

avec donnée initiale $u_\varepsilon|_{t=0} = u_{\varepsilon,0}$. Ils supposent que ce système est symétrisable, au sens où il existe une matrice symétrique définie positive $A^0(u, \varepsilon)$ telle que $A^j \stackrel{\text{déf}}{=} A^0 B^j(u, \varepsilon)$ est symétrique pour tout j . Sous des conditions structurelles sur les matrices B^j et A^j que nous ne détaillons pas, et sous une condition fondamentale d'initialisation sur la famille de données initiales – qui revient à supposer, dans le cas particulier où le système est précisément (4) que la famille de données initiales est bien préparée au sens de la définition 1.5, le résultat de [35] est le suivant. Rappelons que partout dans ce texte s_0 désigne le plus petit entier strictement plus grand que $\frac{d}{2} + 1$.

THÉORÈME 2.1 ([35]). — *Soit $\Omega = \mathbb{T}^d$ un domaine périodique, et soit $s \geq s_0 + 1$. Sous les hypothèses structurelles et d'initialisation évoquées ci-dessus, il existe un temps $T > 0$ indépendant du paramètre ε tel qu'il existe une unique solution u_ε au système (6), bornée dans l'espace $C_w^0([0, T], H^s) \cap C_w^1([0, T], H^{s-1})$.*

L'espace $C_w^0([0, T], H^s)$ est défini en munissant H^s de la topologie faible.

Ce théorème est appliqué dans [35] à de nombreux systèmes de la mécanique des fluides, comme les fluides compressibles (visqueux ou non) ou la magnéto-hydrodynamique. Dans le cas de la limite incompressible (ce résultat n'est pas démontré pour le système abstrait (6) dans [35]), la borne *a priori* sur $\partial_t u_\varepsilon$ donnée par le théorème 2.1 permet en outre de démontrer le résultat de convergence suivant, par application d'un théorème de compacité de type Ascoli ou Lions-Aubin.

THÉORÈME 2.2 ([35]). — *Soit $\Omega = \mathbb{T}^d$ un domaine périodique, et soit $s \geq s_0 + 1$. Si $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ est une famille de données bien préparées dans $H^s(\mathbb{T}^d)$, avec $u_{\varepsilon,0}$ convergeant vers un champ de divergence nulle u_0 dans $H^s(\mathbb{T}^d)$, alors il existe un temps $T > 0$ indépendant du paramètre ε tel qu'il existe une unique solution $(u_\varepsilon, c_\varepsilon)$ au système (4), bornée dans l'espace $C_w^0([0, T], H^s) \cap C_w^1([0, T], H^{s-1})$. En outre, la famille u_ε tend vers \bar{u} dans $C_w^0([0, T], H^s)$ où \bar{u} est solution du système incompressible (2) avec donnée initiale u_0 , et ∇c_ε tend vers zéro dans $C_w^0([0, T], H^{s-1})$.*

Ce théorème indique que, dans le cas de données bien préparées avec des conditions aux limites périodiques, il y a une solution unique au système (4), bornée uniformément en ε , et cette solution converge vers la solution du système incompressible. Ce théorème est analysé dans [35] comme un nouveau théorème, constructif, d'existence locale aux équations d'Euler incompressibles (2). En revanche, les auteurs ne comparent pas le temps d'existence T obtenu par ce biais de la limite incompressible avec celui que l'on obtiendrait par d'autres méthodes plus directes. En particulier dans le cas bidimensionnel

on sait construire des solutions globales au système des fluides incompressibles (voir le théorème 1.3) mais le théorème 2.2 n'indique pas que $T = +\infty$ en dimension deux. Cette question est résolue par les mêmes auteurs dans [36] : dans ce travail l'étude est restreinte aux équations d'Euler compressibles (sans viscosité), aussi bien dans le cas périodique que dans tout l'espace. Le premier résultat de [36] est un théorème d'existence sur un temps uniforme en ε , sans exiger que la donnée soit bien préparée, ce qui est un net progrès par rapport à [35]. Bien entendu dans le cas mal préparé on ne peut espérer obtenir de borne *a priori* sur $\partial_t u_\varepsilon$, mais seulement sur $\varepsilon \partial_t u_\varepsilon$ (et $\varepsilon \partial_t c_\varepsilon$). Ainsi l'on ne peut appliquer de résultat de compacité pour passer à la limite dans l'équation, et dans [36] la restriction aux données bien préparées est nécessaire pour passer à la limite.

THÉORÈME 2.3 ([36]). — *Soit $\Omega = \mathbb{T}^d$ ou \mathbb{R}^d , et soit $s \geq s_0 + 1$. Si $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ est une famille bornée de $H^s(\Omega)$, alors il existe un temps $\bar{T} > 0$ indépendant du paramètre ε tel qu'il existe une unique solution $(u_\varepsilon, c_\varepsilon)$ au système (4), bornée dans $C_w^0([0, \bar{T}], H^s)$. En outre, supposons que $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ tend vers (u_0, c_0) dans $H^s(\Omega)$ avec $\operatorname{div} u_0 = 0$ et $\nabla c_0 = 0$. On suppose ici que $s \geq s_0 + 2$. Soit alors \bar{u} la solution du système incompressible (2), élément de $C^0([0, T_0[, H^s(\Omega)) \cap C^1([0, T_0[, H^{s-1}(\Omega))$ pour un certain $T_0 > 0$ (qui est infini si $d = 2$). Alors pour tout $T < T_0$, il existe ε_0 tel que pour tout $\varepsilon \leq \varepsilon_0$, l'équation (4) admet une unique solution sur le temps $[0, T]$, telle que*

$$(7) \quad u_\varepsilon \rightarrow \bar{u} \quad \text{dans} \quad C^0([0, T], H^{s-1}(\Omega)), \quad \text{et} \quad \nabla c_\varepsilon \rightarrow 0 \quad \text{dans} \quad C^0([0, T], H^{s-2}(\Omega)).$$

Nous ne détaillerons pas la démonstration de ce théorème ici. Pour obtenir une solution sur un temps uniforme (indépendant de ε et ce sans préparation de la donnée initiale), il faut utiliser l'antisymétrie du terme de « pénalisation » qui intervient dans l'équation : de ce fait ce terme non borné n'intervient pas dans les estimations d'énergie permettant de montrer l'existence de solutions. L'idée pour obtenir le résultat de convergence (pour une donnée bien préparée) est de former la différence entre la solution de (4) et la solution de (2) et de montrer que tant que cette dernière solution est définie, la différence reste bornée, et tend vers zéro avec ε par application de méthodes d'énergie (pour lesquelles encore une fois le terme pénalisé disparaît).

Notons que des extensions de ce résultat ont permis d'améliorer la classe d'espaces fonctionnels dans lequel la convergence (7) est vraie ; nous renvoyons le lecteur par exemple aux travaux de H. Beirão da Veiga [4]-[5].

3. LE CAS MAL PRÉPARÉ

Dans ce paragraphe nous allons étudier la limite incompressible sans hypothèse de préparation de la donnée initiale. D'après le théorème 2.3, on sait construire une solution à ce système sur un temps uniforme en ε (en tous cas si Ω est l'espace entier ou le tore, mais nous admettrons que ce résultat est général pour tout domaine régulier) et l'on

cherche maintenant à déterminer le comportement asymptotique de cette solution. Nous allons tout d'abord nous concentrer sur le cas où l'équation est posée dans un domaine extérieur : dans la section 3.1.1 nous détaillerons la démonstration du passage à la limite incompressible dans le cas de l'espace entier, en introduisant des estimations dispersives (de type « Strichartz »). Le paragraphe 3.1.2 donnera une indication de l'application de cette méthode au cas d'un domaine extérieur. Le cas de domaines bornés est plus délicat, et nous commencerons par expliquer les phénomènes d'oscillations dans le cas périodique (section 3.2.1), en nous attardant très peu sur le cas de domaines à bords dans la section 3.2.2.

3.1. Le cas de domaines non bornés — dispersion

3.1.1. *Dans l'espace entier.* Considérons le système (4) posé dans l'espace entier \mathbb{R}^d . Dans [61], S. Ukai démontre le résultat suivant.

THÉORÈME 3.1 ([61]). — *Soit $\Omega = \mathbb{R}^d$, et soit $s \geq s_0$. Si $(u_{\varepsilon,0}, c_{\varepsilon,0})_{\varepsilon>0}$ est une famille bornée de $H^s(\Omega)$, convergeant vers (u_0, c_0) dans $H^s(\Omega)$, alors la solution $(u_\varepsilon, c_\varepsilon)$ construite dans le théorème 2.3 vérifie quand ε tend vers zéro, avec les notations de ce théorème,*

$$u_\varepsilon \rightarrow \bar{u} \quad \text{dans} \quad C_{loc}^0((0, T] \times \Omega), \quad \text{et} \quad \nabla c_\varepsilon \rightarrow 0 \quad \text{dans} \quad C_{loc}^0((0, T] \times \Omega), \quad \forall T < T_0.$$

Remarquons que le résultat de convergence n'est pas vrai en $t = 0$, ce qui est naturel puisque la donnée initiale n'est pas supposée de type incompressible ; il y a donc un phénomène de couche limite en temps. Notons d'autre part que le résultat démontré dans [61] n'est pas tout à fait celui énoncé ci-dessus : S. Ukai démontre la convergence sur le temps d'existence de (4) plutôt que celui de (2), mais nous verrons ci-dessous comment obtenir le résultat de convergence sur $(0, T)$ pour tout $T < T_0$.

La démonstration de ce théorème repose sur l'étude du système linéarisé : la partie de la solution $(u_\varepsilon, c_\varepsilon)$ qui n'est pas dans le noyau de l'opérateur de pénalisation vérifie une équation d'ondes, et l'on utilise alors des estimations de type « Strichartz » pour montrer que cette partie tend vers zéro. On décompose ainsi tout champ $U \stackrel{\text{déf}}{=} (c, u)$ de la manière suivante : $U = \bar{U} + U_{osc}$ avec $\bar{U} = (0, \bar{u}) = (0, Pu)$, où P est le projecteur de Leray, orthogonal dans L^2 sur les champs de vecteurs de divergence nulle. On pose alors $W_\varepsilon = U_\varepsilon - (0, \bar{u})$, et en écrivant $W_{\varepsilon,osc} = (c_\varepsilon, w_{\varepsilon,osc}) = (c_\varepsilon, u_{\varepsilon,osc})$, il vient :

$$\begin{aligned} \partial_t \bar{w}_\varepsilon + P(u_\varepsilon \cdot \nabla u_\varepsilon) - P(\bar{u} \cdot \nabla \bar{u}) &= 0 \\ \partial_t w_{\varepsilon,osc} + (Id - P)(u_\varepsilon \cdot \nabla u_\varepsilon) + \frac{\gamma - 1}{2} c_\varepsilon \nabla c_\varepsilon + \frac{\gamma - 1}{2\varepsilon} \nabla c_\varepsilon &= 0 \\ \partial_t c_\varepsilon + c_\varepsilon \operatorname{div} w_{\varepsilon,osc} + \frac{\gamma - 1}{2} u_\varepsilon \cdot \nabla c_\varepsilon + \frac{\gamma - 1}{2\varepsilon} \operatorname{div} w_{\varepsilon,osc} &= 0. \end{aligned}$$

Il s'agit maintenant de montrer que W_ε tend vers zéro, ce qui s'obtient en deux étapes : on commence par montrer que la partie « oscillante » $W_{\varepsilon,osc}$ tend vers zéro, puis on en déduit le résultat cherché sur la partie « non oscillante » \bar{w}_ε en appliquant des méthodes relevant

de la stabilité des équations d'Euler incompressibles (essentiellement par application d'un lemme de Gronwall pour absorber le terme $P(u_\varepsilon \cdot \nabla u_\varepsilon) - P(\bar{u} \cdot \nabla \bar{u})$).

Nous n'allons pas détailler ces étapes, mais donner un schéma de preuve. Commençons donc par introduire le linéarisé du système « oscillant » : posons ainsi $A(D)U \stackrel{\text{déf}}{=} {}^t(\operatorname{div} u, \nabla c)$, et étudions

$$(8) \quad \partial_t U + \frac{1}{\varepsilon} A(D)U = 0, \quad U|_{t=0} = U_0.$$

La matrice $A(D)$ s'écrit en variables de Fourier $A(\xi) = i \begin{pmatrix} 0 & \xi \\ {}^t\xi & 0 \end{pmatrix}$, et ses valeurs propres sont $\pm i|\xi|$ avec vecteurs propres associés $\pm {}^t(|\xi|, \xi)$. Nous avons affaire à une équation des ondes (il s'agit de l'acoustique), et une estimation de phase stationnaire conduit aux estimations dispersives suivantes : dès que U_0 est à support compact en fréquences, on a

$$(9) \quad \|U(t)\|_{L^\infty(\mathbb{R}^d)} \leq C \left(\frac{\varepsilon}{t}\right)^{\frac{d-1}{2}} \|U_0\|_{L^1(\mathbb{R}^d)}.$$

Il n'est pas question ici de rappeler la théorie des estimations dispersives et de Strichartz ; nous renvoyons le lecteur par exemple à [22] pour leur démonstration dans le cas des équations des ondes. Ces estimations ont été popularisées dans le domaine des équations aux dérivées partielles à la fois pour leurs propriétés dispersives (qui donnent des résultats asymptotiques pour la limite incompressible comme pour d'autres problèmes de limite singulière) que pour leurs propriétés régularisantes (qui permettent d'améliorer les résultats classiques sur le problème de Cauchy pour de nombreuses équations telles les équations des ondes ou de Schrödinger non linéaires — voir par exemple [3], [33], [34] pour les ondes). Dans notre cadre, l'inégalité (9) permet de montrer par un argument de densité que la norme L^∞ de la solution de (8) tend vers zéro quand ε tend vers zéro, pour tout temps strictement positif. Un argument de dualité permet d'obtenir le même résultat pour une équation non linéaire du type $\partial_t U + \frac{1}{\varepsilon} A(D)U + Q(U, DU) = 0$, où D est un multiplicateur de Fourier d'ordre un et Q une forme quadratique.

On retourne alors au système vérifié par W_ε : on peut faire disparaître les termes oscillants du système, quand ε tend vers zéro, en interpolant les estimations dispersives avec des bornes *a priori* sur W^ε dans $H^s(\mathbb{R}^d)$. En utilisant parallèlement le fait que la solution d'un système symétrique hyperbolique existe tant que sa norme Lipschitz en espace reste bornée, on peut en déduire que c'est le temps d'existence de \bar{u} qui va déterminer le temps d'existence de W_ε , ainsi que le fait que W_ε tend vers zéro. Nous ne détaillerons pas plus la démonstration ici, et renvoyons le lecteur intéressé à l'article original [61] ; pour le cas de Navier–Stokes on pourra consulter [13].

REMARQUE 3.2. — Toujours par utilisation des estimations de Strichartz mais en quantifiant précisément le taux de décroissance en ε donné par la dispersion, A. Dutrifoy et T. H'midi ont montré récemment dans [16] un résultat beaucoup plus précis que celui

énoncé dans le théorème 3.1. En particulier ils autorisent la partie «compressible» de la donnée initiale à être non bornée en ε , ce qui leur permet d'obtenir en particulier un théorème d'existence en temps grand pour le problème des poches de tourbillon bidimensionnelles (le rotationnel du champ de vitesses initial est la fonction caractéristique d'un domaine borné de l'espace).

3.1.2. *Un domaine extérieur.* Dans le cas d'un domaine extérieur, on s'attend à ce que les mêmes arguments de type dispersifs soient valables ; c'est en effet le cas. Dans le cas d'un demi-espace par exemple, T. Iguchi montre dans [27] la convergence des solutions du système des fluides compressibles (4) vers un champ de vecteur incompressible vérifiant (2), et ce sans hypothèse de préparation de la donnée initiale. Dans [28]-[30], H. Isozaki quant à lui démontre le même résultat pour le cas d'un domaine extérieur. Là encore la démonstration repose sur une estimation dispersive sur le linéarisé, obtenue par application de méthodes de théorie spectrale. La forme du bord du domaine n'intervient pas.

Nous ne nous attarderons pas plus sur ces cas ici, qui relèvent finalement des mêmes idées que dans le cas de l'espace entier, avec des techniques de théorie spectrale additionnelles.

3.2. Le cas de domaines bornés — oscillations

Nous allons maintenant discuter du cas où les équations sont posées dans un domaine borné régulier de l'espace. Il n'est plus question alors d'utiliser d'estimations dispersives, qui ne peuvent être valables globalement en temps que dans un domaine non borné ; notons qu'il existe des estimations dispersives dans des domaines bornés ou sur le tore pour des équations d'ondes, mais ces estimations ne sont vraies que localement en temps — remises à l'échelle dans notre cadre, elles ne sont donc pas utilisables car elles deviennent vraies sur un temps de l'ordre de ε seulement.

3.2.1. *Conditions aux limites périodiques.* Nous allons présenter ici les techniques de «filtrage» dues à S. Schochet [56] (voir aussi [23]). Considérons une équation abstraite du type suivant

$$(10) \quad \partial_t U_\varepsilon + Q(U_\varepsilon, U_\varepsilon) + \frac{1}{\varepsilon} A U_\varepsilon = 0 \quad \text{dans } \mathbb{R}^+ \times \mathbb{T}^d, \quad U_{\varepsilon|t=0} = U_0,$$

où A est une matrice antisymétrique de multiplicateurs de Fourier, et où Q est une forme quadratique du type $Q(a, b) = \frac{1}{2}(a \cdot \nabla b + b \cdot \nabla a)$. On pourrait bien sûr généraliser le cadre d'étude en choisissant une forme quadratique Q moins particulière, mais nous gardons celle-ci ici afin d'alléger les notations.

On suppose ici que le champ de vecteurs U_ε a d' composantes, avec $d' \geq d$ (ce qui est le cas pour la limite incompressible en particulier). La question de l'existence d'une solution sur un temps uniforme est résolue assez simplement, en remarquant que l'antisymétrie

de A conduit à des estimations d'énergie sur (10) indépendantes de ε . On cherche à déterminer le comportement asymptotique de cette solution, ainsi que la dépendance en ε de son temps d'existence. Posons $\tilde{U}_\varepsilon \stackrel{\text{déf}}{=} e^{-\frac{t}{\varepsilon}A}U_\varepsilon \stackrel{\text{déf}}{=} \mathcal{A}\left(-\frac{t}{\varepsilon}\right)U_\varepsilon$. En remarquant que \mathcal{A} commute avec les dérivations, le champ \tilde{U}_ε vérifie le système suivant :

$$(11) \quad \partial_t \tilde{U}_\varepsilon + \mathcal{A}\left(-\frac{t}{\varepsilon}\right)Q \left(\mathcal{A}\left(\frac{t}{\varepsilon}\right)\tilde{U}_\varepsilon, \mathcal{A}\left(\frac{t}{\varepsilon}\right)\tilde{U}_\varepsilon \right) = 0.$$

Comme $\mathcal{A}\left(\pm\frac{t}{\varepsilon}\right)$ est unitaire dans tous les espaces de Sobolev H^s (par l'antisymétrie de A), on dispose d'estimations *a priori* sur $\partial_t \tilde{U}_\varepsilon$ et l'on en déduit que le champ \tilde{U}_ε converge fortement (dans un espace de type Sobolev) vers un champ limite U . Toute la question revient maintenant à déterminer l'équation vérifiée par ce champ limite ; nous allons montrer qu'aussi longtemps que U existe, il existe aussi une solution au système de départ (10) et que celle-ci converge (en un certain sens) vers U . Commençons par déterminer l'équation vérifiée par U , en passant à la limite (au sens des distributions) dans l'équation sur \tilde{U}_ε . S'il est facile de passer à la limite au sens des distributions dans les termes linéaires, c'est *a priori* moins aisé dans les termes non linéaires de (11) ; en effet on ne dispose pas d'information de convergence forte de $\mathcal{A}\left(\frac{t}{\varepsilon}\right)\tilde{U}_\varepsilon$. Écrivons la matrice A en variables de Fourier ; son caractère antisymétrique permet d'introduire une base orthogonale de vecteurs propres $(e_j(n))_{1 \leq j \leq d'}$ avec les valeurs propres associées $i\lambda_j(n)$, pour tout $n \in \mathbb{Z}^d$, avec $\lambda_j(n) \in \mathbb{R}$. Notons ici que les variables de Fourier n sont discrètes puisque l'on est dans un cadre périodique. Ainsi l'on a pour tout champ de vecteurs a et en notant \mathcal{F} la transformée de Fourier et (\cdot, \cdot) le produit scalaire dans $\mathbb{C}^{d'}$,

$$\forall n \in \mathbb{Z}^d, \quad \mathcal{F}\left(e^{-\frac{t}{\varepsilon}A}a\right)(n) = \sum_{j=1}^{d'} e^{-i\frac{t}{\varepsilon}\lambda_j(n)} (\mathcal{F}a(n), e_j(n)) e_j(n).$$

La forme quadratique apparaissant dans l'équation satisfaite par \tilde{U}_ε s'écrit ainsi

$$(12) \quad \mathcal{F}Q_\varepsilon(a, a)(n) \stackrel{\text{déf}}{=} \mathcal{F}\left\{ \mathcal{A}\left(-\frac{t}{\varepsilon}\right)Q \left(\mathcal{A}\left(\frac{t}{\varepsilon}\right)a, \mathcal{A}\left(\frac{t}{\varepsilon}\right)a \right) \right\}(n) = \sum_{j=1}^{d'} q_{\varepsilon,j}(n) e_j(n), \text{ avec}$$

$$q_{\varepsilon,j}(n) \stackrel{\text{déf}}{=} \sum_{j',j''=1}^{d'} \sum_{k+m=n} e^{i\frac{t}{\varepsilon}\lambda_{j,j',j''}^{k,m,n}} F_{j,j',j''}^{k,m,n}(\hat{a}_k; \hat{a}_m)$$

et où $\lambda_{j,j',j''}^{k,m,n} \stackrel{\text{déf}}{=} \lambda_{j'}(k) + \lambda_{j''}(m) - \lambda_j(n)$. On a noté \hat{a}_k la transformée de Fourier de a au point k . La forme exacte du terme bilinéaire $F_{j,j',j''}^{k,m,n}(\hat{a}_k; \hat{a}_m)$ ne nous sera pas nécessaire ici, nous la présentons ci-dessous pour être complet : on a précisément

$$F_{j,j',j''}^{k,m,n}(\hat{a}_k; \hat{a}_m) \stackrel{\text{déf}}{=} (\hat{a}(k), e_{j'}(k))(e_{j'}(k), m)(\hat{a}(m), e_{j''}(m))(e_{j''}(m), e_j(n)).$$

Pour obtenir la limite au sens des distributions de cette quantité (12), il suffit d'appliquer un théorème de phase non stationnaire et l'on obtient immédiatement la forme quadratique limite : il s'agit de restreindre les fréquences (k, m, n) , dans la sommation intervenant

dans (12), à celles annulant la phase d'oscillation $\lambda_{j,j',j''}^{k,m,n}$. On obtient

$$\mathcal{FQ}(a, a)(n) \stackrel{\text{d\u00e9f}}{=} \sum_{j=1}^{d'} q_j(n) e_j(n), \quad \text{avec}$$

$$(13) \quad q_j(n) \stackrel{\text{d\u00e9f}}{=} \sum_{j',j''=1}^{d'} \sum_{(k,m) \in R_{j,j',j''}(n)} F_{j,j',j''}^{k,m,n}(\widehat{a}_k; \widehat{a}_m).$$

On a not\u00e9 $R_{j,j',j''}(n)$ l'ensemble des r\u00e9sonances

$$R_{j,j',j''}(n) \stackrel{\text{d\u00e9f}}{=} \{(k, m) \in \mathbb{Z}^{2d}, \quad k + m = n \quad \text{et} \quad \lambda_{j'}(k) + \lambda_{j''}(m) - \lambda_j(n) = 0\}.$$

Cet ensemble a une importance fondamentale dans la compr\u00e9hension du syst\u00e8me limite, nous y reviendrons plus bas. \u00c9non\u00e7ons pour l'instant le th\u00e9or\u00e8me de convergence de $\widetilde{U}_\varepsilon$ vers U , aussi longtemps qu'existe U . Ce r\u00e9sultat est d\u00fb \u00e0 S. Schochet [56] (voir aussi [19], o\u00f9 en particulier le terme suivant de l'asymptotique est calcul\u00e9).

TH\u00c9OR\u00c8ME 3.3 ([56]). — Soit $U_0 \in H^s(\mathbb{T}^d)$, avec $s \geq s_0 + 2$, et soit $U \in C^0([0, T], H^s(\mathbb{T}^d))$ la solution du syst\u00e8me

$$(14) \quad \partial_t U + \mathcal{Q}(U, U) = 0,$$

o\u00f9 la forme quadratique \mathcal{Q} a \u00e9t\u00e9 d\u00e9finie en (13). Alors il existe ε_T tel que pour tout $\varepsilon \leq \varepsilon_T$ la solution U_ε de (10) est un \u00e9l\u00e9ment de l'espace $C^0([0, T], H^s(\mathbb{T}^d))$ et $U_\varepsilon - \mathcal{A}\left(\frac{t}{\varepsilon}\right)U$ tend vers z\u00e9ro dans $C^0([0, T], H^{s-1}(\mathbb{T}^d))$.

La d\u00e9monstration de ce th\u00e9or\u00e8me consiste \u00e0 \u00e9tudier $W_\varepsilon \stackrel{\text{d\u00e9f}}{=} \mathcal{A}\left(-\frac{t}{\varepsilon}\right)U_\varepsilon - U$, qui v\u00e9rifie

$$(15) \quad \partial_t W_\varepsilon + \mathcal{Q}_\varepsilon(W_\varepsilon, W_\varepsilon + 2U) = (\mathcal{Q} - \mathcal{Q}_\varepsilon)(U, U).$$

Il y a ici une r\u00e9elle difficult\u00e9 due au fait que $(\mathcal{Q} - \mathcal{Q}_\varepsilon)(U, U)$ ne converge vers z\u00e9ro qu'au sens des distributions, par application du th\u00e9or\u00e8me de la phase non stationnaire. Ainsi l'on ne peut esp\u00e9rer conclure par une simple estimation d'\u00e9nergie. L'id\u00e9e de S. Schochet est de constater que $(\mathcal{Q} - \mathcal{Q}_\varepsilon)(U, U)$ est en fait fortement oscillante en temps et n'est ainsi qu'une petite perturbation de l'\u00e9quation $\partial_t W_\varepsilon + \mathcal{Q}_\varepsilon(W_\varepsilon, W_\varepsilon + 2U) = 0$. Pr\u00e9cisons cette id\u00e9e. On commence par se ramener au cas o\u00f9 les fr\u00e9quences dans la sommation de (13) (et de m\u00eame dans celle intervenant dans \mathcal{Q}_ε) sont uniform\u00e9ment born\u00e9es par un param\u00e8tre N . L'erreur commise en omettant les « hautes » fr\u00e9quences en k , m ou n est arbitrairement petite pourvu que N soit assez grand, et ce uniform\u00e9ment en ε : on n'utilise pour cela que la r\u00e9gularit\u00e9 de U , qui ne d\u00e9pend pas de ε . Une fois les fr\u00e9quences k, m et n restreintes \u00e0 la boule de \mathbb{Z}^{3d} de rayon N , on peut introduire la fonction auxiliaire suivante

$$\widetilde{W}_{\varepsilon, N} \stackrel{\text{d\u00e9f}}{=} W_\varepsilon + \varepsilon R_{\varepsilon, N}, \quad \text{avec}$$

$$\mathcal{F}R_{\varepsilon, N}(n) \stackrel{\text{d\u00e9f}}{=} \sum_{\substack{|k|, |m|, |n| \leq N \\ k+m=n}} \sum_{(k,m) \notin R_{j,j',j''}(n)} \frac{i}{\lambda_{j,j',j''}^{k,m,n}} e^{i\frac{t}{\varepsilon} \lambda_{j,j',j''}^{k,m,n}} F_{j,j',j''}^{k,m,n}(\widehat{U}_k; \widehat{U}_m) e_j(n).$$

Cette fonction est aussi régulière qu'on peut le souhaiter (elle est à fréquences bornées), et sa vertu est d'éliminer le terme oscillant de l'équation (15) sur W_ε . On a en effet, en notant $(\mathcal{Q} - \mathcal{Q}_\varepsilon)^N(U, U)$ la partie «hautes fréquences» du membre de droite de (15) (qui comme noté ci-dessus est arbitrairement petite, uniformément en ε),

$$\partial_t \widetilde{W}_{\varepsilon, N} + \mathcal{Q}_\varepsilon(\widetilde{W}_{\varepsilon, N} - \varepsilon R_{\varepsilon, N}, \widetilde{W}_{\varepsilon, N} - \varepsilon R_{\varepsilon, N} + 2U) = (\mathcal{Q} - \mathcal{Q}_\varepsilon)^N(U, U) + \varepsilon R_{\varepsilon, N}^t(U, U),$$

où $R_{\varepsilon, N}^t$ provient de la dérivée en temps de $R_{\varepsilon, N}$: cette dérivée est égale d'une part à l'opposé du terme «basses fréquences» du membre de droite de (15) (ce qui permet de l'éliminer) et d'autre part à des termes résiduels dus aux dérivées temporelles de U . Mais comme ce terme $R_{\varepsilon, N}^t$ est aussi à fréquences bornées, il est régulier et $\varepsilon R_{\varepsilon, N}^t$ est donc arbitrairement petit pour tout N . Finalement $\widetilde{W}_{\varepsilon, N}$ vérifie une équation de type quasilinéaire à données petites (de l'ordre de ε), avec un terme source arbitrairement petit. Des arguments classiques permettent de montrer que le temps d'existence de cette équation n'est limité que par celui de U , et que $\widetilde{W}_{\varepsilon, N}$ tend vers zéro quand N tend vers l'infini et ε vers zéro. Cela clôt la démonstration du théorème 3.3 : rappelons en effet que $\varepsilon R_{\varepsilon, N}$ est arbitrairement régulier pour tout N fixé, et arbitrairement petit (pour ε assez petit), donc la limite de $\widetilde{W}_{\varepsilon, N}$ est celle de W_ε , aussi longtemps que U est définie. Nous renvoyons le lecteur à [56] pour des détails.

Nous pouvons à présent appliquer ce théorème aux équations des fluides compressibles (4), et obtenir ainsi le comportement asymptotique de ses solutions ; le problème revient à analyser le système limite (14) dans le cas de la limite incompressible, c'est-à-dire plus précisément comprendre l'ensemble des résonances $R_{j, j', j''}(n)$ dans le cas où l'opérateur A est la matrice de l'acoustique, introduite dans l'équation (8) au paragraphe 3.1.1. Nous avons dans ce cas $d' = d + 1$ où d est la dimension d'espace, et les valeurs propres sont $\lambda_\pm(n) \stackrel{\text{déf}}{=} \pm|n|$ et $\lambda_0 \stackrel{\text{déf}}{=} 0$. Le mode 0 correspond aux champs incompressibles, dans le noyau de l'opérateur acoustique. Décrivons plus précisément l'ensemble résonant : si nous posons $\varepsilon_j \in \{-1, 0, 1\}$ pour tout $j \in \{1, \dots, d'\}$, alors $R_{j, j', j''}$ est du type

$$R_{j, j', j''}(n) = \{(k, m) \in \mathbb{Z}^{2d}, \quad k + m = n, \quad \text{et} \quad \varepsilon_{j'}|k| + \varepsilon_{j''}|m| - \varepsilon_j|n| = 0\}.$$

On s'aperçoit ainsi que dans le cas particulier où $\varepsilon_j = \varepsilon_{j'} = \varepsilon_{j''} = 0$, qui correspond à l'interaction des modes incompressibles, le terme bilinéaire dans l'équation limite (14) est précisément le terme quadratique habituel des équations incompressibles $P(\bar{u} \cdot \nabla \bar{u})$, où P est le projecteur de Leray sur les champs de divergence nulle (c'est un fait général, les termes non oscillants n'interagissent pas entre eux autrement que de manière non oscillante). Plus intéressante est certainement l'étude des interactions compressible-compressible. On peut montrer que cette interaction prise sur le mode incompressible ($\varepsilon_j = 0$ et $\varepsilon_{j'}\varepsilon_{j''} \neq 0$) produit un terme gradient, et est donc responsable du terme $-\nabla p$ dans l'équation limite incompressible (voir par exemple [23], [47]). Ce qui importe donc

est d'étudier l'interaction compressible–compressible sur un mode compressible, c'est-à-dire l'ensemble des résonances dans le cas où le produit $\varepsilon_j \varepsilon_{j'} \varepsilon_{j''}$ est non nul. Cette analyse a été menée précisément par N. Masmoudi dans [47]. Ainsi il a démontré que la condition de résonance

$$k + m = n \quad \text{et} \quad \pm |k| \pm |m| \pm |n| = 0$$

détermine un très faible nombre de triplets (k, m, n) , si faible que finalement le terme bilinéaire de type compressible–compressible se comporte sur les modes compressibles comme si l'on était en dimension un d'espace. Plus précisément on peut se convaincre (voir [47], ainsi que [11] pour des calculs plus quantitatifs) que deux éléments k et m de l'ensemble de résonance sont nécessairement colinéaires. N. Masmoudi restreint alors son étude aux éléments de

$$\mathcal{P} \stackrel{\text{déf}}{=} \{p \in \mathbb{Z}^d, \quad p_1 \wedge \dots \wedge p_d = 1 \quad \text{et} \quad p_1 > 0\}.$$

En réécrivant la forme quadratique en décomposant les variables de Fourier sur \mathcal{P} , il est alors possible de se ramener à des calculs purement monodimensionnels. Cela induit naturellement des estimations bien meilleures que les estimations de produit habituelles (la régularité Sobolev du produit de deux fonctions se dégrade essentiellement d'un facteur $1/2$ avec chaque dimension d'espace supplémentaire) ; ainsi l'on peut démontrer assez aisément que le facteur limitant pour le temps d'existence des solutions du système limite (14) est le temps d'existence des équations d'Euler incompressibles. En particulier en dimension deux d'espace ce système limite est globalement bien posé.

On peut finalement démontrer le théorème suivant.

THÉORÈME 3.4 ([56],[47]). — *Soit $U_0 \in H^s(\mathbb{T}^d)$, avec $s \geq s_0 + 2$. Soit T^* le temps d'existence des solutions de l'équation des fluides incompressibles (2) dans $H^s(\mathbb{T}^d)$. Alors la solution U du système limite (14) associé à l'équation faiblement compressible (4) appartient à $C^0([0, T^*), H^s(\mathbb{T}^d))$ et pour tout temps $T < T^*$, il existe ε_T tel que pour tout $\varepsilon \leq \varepsilon_T$ la solution U_ε de (14) est un élément de l'espace $C^0([0, T], H^s(\mathbb{T}^d))$ et $U_\varepsilon - \mathcal{A}(\frac{t}{\varepsilon})U$ tend vers zéro dans $C^0([0, T], H^{s-1}(\mathbb{T}^d))$.*

REMARQUE 3.5. — Nous n'avons pas précisé ici la forme exacte du système limite des équations des fluides compressibles périodiques. Il s'agit du système suivant : la limite U s'écrit $U = \bar{U} + U_{osc} = (0, \bar{u}) + U_{osc}$ avec

$$\begin{aligned} \partial_t \bar{u} + \bar{u} \cdot \nabla \bar{u} + \nabla p &= 0, \quad \text{div } \bar{u} = 0 \\ \partial_t U_{osc} + \mathcal{Q}_{osc}(U_{osc}, U_{osc} + 2\bar{U}) &= 0, \end{aligned}$$

où \mathcal{Q}_{osc} est définie comme dans (13), restreinte aux modes $\lambda_j \neq 0$ (le mode nul est représenté dans l'équation en \bar{u}). En particulier cette forme quadratique \mathcal{Q}_{osc} , agissant sur des modes compressibles (*i.e* associés à une valeur propre non nulle de A) vérifie des estimations de type monodimensionnelles.

3.2.2. *Un domaine à bord.* Pour clore cette présentation de la limite incompressible dans le cas isentropique, nous allons très brièvement considérer le cas où l'équation est posée dans un domaine de l'espace, avec condition $u \cdot n = 0$ au bord.

Dans le cas des équations non visqueuses il y a assez peu de travaux sur le sujet ; nous présentons ici les travaux de P. Secchi [57]-[58]. L'existence de solutions sur un temps uniforme est démontrée sans hypothèse de préparation de la donnée initiale, comme dans [36]. La méthode repose toujours sur des estimations d'énergie utilisant l'antisymétrie de l'opérateur acoustique. Pour démontrer des résultats de convergence en revanche, P. Secchi démontre la convergence faible des solutions vers les solutions du système des fluides incompressibles, et la convergence forte n'est obtenue que dans le cas de données préparées au sens de la définition 1.5.

La situation est donc pour l'instant assez peu comprise dans le cas d'un domaine borné général. Notons au passage que pour les équations de Navier–Stokes compressibles avec conditions aux limites de Dirichlet homogènes, la situation est plus favorable : dans [14], B. Desjardins, E. Grenier, P.-L. Lions et N. Masmoudi démontrent en effet que sous une hypothèse géométrique sur le domaine (toutes les solutions du problème surdéterminé $-\Delta\varphi = \lambda\varphi$ dans Ω avec φ constant sur le bord et $\partial_n\varphi = 0$ sur le bord, doivent être nulles) alors il y a convergence forte des solutions compressibles vers les solutions incompressibles. C'est un résultat surprenant au vu de ce que nous avons constaté dans le cas périodique (les ondes acoustiques dans le cas périodique ne disparaissent pas à la limite) ; cela est dû à l'absorption par le bord des ondes acoustiques, par un phénomène de couche limite propre au cas visqueux.

4. LES ÉQUATIONS NON ISENTROPIQUES

4.1. Présentation

Le dernier paragraphe de ce texte concerne une extension de l'étude de la limite incompressible au cas d'équations non isentropiques. Jusqu'ici en effet nous avons occulté le fait qu'un fluide compressible peut vérifier des équations plus générales que (1), dans lesquelles il faut prendre en compte le transport de l'entropie. Les équations sont ainsi les suivantes :

$$\begin{aligned}
 (16) \quad & \partial_t \rho + u \cdot \nabla \rho + \rho \operatorname{div} u = 0 \\
 & \rho(\partial_t u + u \cdot \nabla u) + \nabla p = 0 \\
 & \partial_t S + u \cdot \nabla S = 0 \\
 & (\rho, u, S)|_{t=0} = (\rho_0, u_0, S_0)
 \end{aligned}$$

où ρ est la densité, p est la pression, u est la vitesse et S l'entropie. Ces variables sont reliées par l'équation d'état $\rho = R(S, p)$, et la vitesse du son c est donnée par $\frac{1}{c^2} = \frac{\partial R}{\partial p} > 0$.

Dans la suite nous définirons aussi $A \stackrel{\text{déf}}{=} \frac{1}{R} \frac{\partial R}{\partial p}$.

Nous renvoyons à l'exposé [50] pour une présentation de ces équations, ainsi que pour la symétrisation et le changement d'échelle dans (16) conduisant finalement au système suivant :

$$(17) \quad \begin{aligned} a_\varepsilon(\partial_t q_\varepsilon + v_\varepsilon \cdot \nabla q_\varepsilon) + \frac{1}{\varepsilon} \operatorname{div} v_\varepsilon &= 0 \\ r_\varepsilon(\partial_t v_\varepsilon + v_\varepsilon \cdot \nabla v_\varepsilon) + \frac{1}{\varepsilon} \nabla q_\varepsilon &= 0 \\ \partial_t S_\varepsilon + v_\varepsilon \cdot \nabla S_\varepsilon &= 0 \\ (q_\varepsilon, v_\varepsilon, S_\varepsilon)|_{t=0} &= (q_{\varepsilon,0}, v_{\varepsilon,0}, S_{\varepsilon,0}) \end{aligned}$$

où $a_\varepsilon \stackrel{\text{déf}}{=} \mathcal{A}(S, \varepsilon q)$ et $r_\varepsilon \stackrel{\text{déf}}{=} \mathcal{R}(S, \varepsilon q)$. Les fonctions \mathcal{A} et \mathcal{R} correspondent aux fonctions A et R précédentes, écrites dans les nouvelles variables (S, q) plutôt que (S, p) .

Comme dans le cas isentropique étudié ci-dessus, deux questions se posent : y a-t-il des solutions au système (17) bornées uniformément en ε , sur un intervalle de temps indépendant de ε ? Si oui, quel est leur comportement quand ε tend vers zéro ? Pour deviner le système limite, comme dans le cas isentropique on annule les termes non bornés dans l'équation (17) : on trouve $\operatorname{div} v_\varepsilon = 0$ et $\nabla q_\varepsilon = 0$, et ainsi on s'attend à obtenir à la limite le système incompressible suivant :

$$(18) \quad \begin{aligned} r_0(\partial_t v + v \cdot \nabla v) + \nabla P &= 0 \\ \partial_t S + v \cdot \nabla S &= 0 \\ \operatorname{div} v &= 0 \end{aligned}$$

avec naturellement $r_0 \stackrel{\text{déf}}{=} \mathcal{R}(S, 0)$.

4.2. Le problème de Cauchy

Contrairement au cas isentropique étudié dans les paragraphes précédents, la question de l'existence de solutions bornées en ε sur un temps indépendant de ε est loin d'être immédiate : à première vue on peut chercher à faire fonctionner l'argument habituel d'antisymétrie du terme de pénalisation, mais cette méthode tombe immédiatement en défaut à cause du couplage avec l'entropie. On se convainc aisément de ce fait en étudiant le système modèle linéarisé suivant (nous reprenons ici un exemple de [51]) :

$$(19) \quad \begin{aligned} a(S)\partial_t u_\varepsilon + \frac{1}{\varepsilon} \partial_x u_\varepsilon &= 0 \\ \partial_t S &= 0 \\ (u_\varepsilon, S)|_{t=0} &= (u_0, S_0). \end{aligned}$$

Si S est un état constant $S = \underline{S}$ alors

$$u_\varepsilon(t, x) = u_0 \left(x - \frac{t}{\varepsilon a(\underline{S})} \right).$$

Ainsi une petite perturbation de l'état constant \underline{S} induit une grande perturbation dans u_ε , et le système (19) est donc instable. Dans [51] G. Métivier et S. Schochet démontrent néanmoins le théorème suivant sur le système complet (17).

THÉORÈME 4.1 ([50]-[51]). — *Soit $s \geq s_0$, soit $M_0 > 0$ et soit $\Omega = \mathbb{R}^d$ ou \mathbb{T}^d . Si les données initiales $(v_{\varepsilon,0}, q_{\varepsilon,0}, S_{\varepsilon,0})$ sont uniformément bornées dans $H^s(\Omega)$ par M_0 , alors il existe une unique solution $(v_\varepsilon, q_\varepsilon, S_\varepsilon)$ à (17), uniformément bornée dans $C^0([0, T], H^s(\Omega))$ où T ne dépend que de M_0 ; en outre $\partial_t S_\varepsilon$ et $\partial_t \text{rot}(r_\varepsilon v_\varepsilon)$ sont bornées respectivement dans $C^0([0, T], H^{s-1}(\Omega))$ et $C^0([0, T], H^{s-2}(\Omega))$.*

L'idée de la démonstration de ce théorème peut se lire sur le système modèle (19) : l'étude du système linéarisé ne donnant pas de résultat satisfaisant, il convient d'étudier directement le système non linéaire de départ. Afin de simplifier la présentation, nous allons plutôt expliquer la démarche de [51] sur le système modèle suivant :

$$(20) \quad \begin{aligned} \sigma_\varepsilon \partial_t u_\varepsilon + \frac{1}{\varepsilon} \partial_x u_\varepsilon &= 0 \\ \partial_t \sigma_\varepsilon + a(u_\varepsilon) \partial_x \sigma_\varepsilon &= 0. \end{aligned}$$

Dans ce système u_ε joue le rôle des variables $(q_\varepsilon, \text{div } v_\varepsilon)$ et σ_ε celui de $(S_\varepsilon, \text{rot}(r_\varepsilon v_\varepsilon))$.

Pour éviter de se ramener par une estimation d'énergie à un cadre linéarisé (instable), on commute la première équation de (20) non pas avec ∂_x^k , méthode usuelle mais qui conduirait ici à une équation instable, mais plutôt avec $\left(\frac{1}{\sigma_\varepsilon} \partial_x \right)^k$. Ainsi comme $\partial_t \sigma_\varepsilon$ et σ_ε sont estimées (dans H^{s-1} et H^s respectivement) par $\|u_\varepsilon\|_{H^s}$ d'après la seconde équation de (20), les commutateurs supplémentaires obtenus peuvent être absorbés par une inégalité de type Gronwall. Cette démarche peut être mise en œuvre de manière analogue pour le système de départ (17) ce qui démontre le théorème 4.1.

Notons que dans [51] ce théorème d'existence est démontré pour une classe plus générale d'équations, faisant apparaître l'importance de la structure spéciale non linéaire. Nous ne donnerons pas plus de détails sur la preuve du théorème 4.1 et renvoyons à [51] pour toute la démonstration.

La question de l'existence de solutions uniformément bornées étant désormais résolue, nous allons à présent nous attacher à comprendre le comportement asymptotique de ces solutions quand ε tend vers zéro. Dans le cas isentropique nous avons constaté que l'étude est plus simple dans le cas de \mathbb{R}^d que dans \mathbb{T}^d . C'est le cas aussi (et bien plus encore) dans le cas non isentropique ; nous allons donc commencer par nous pencher sur le cas d'équations posées dans l'espace entier dans la section 4.3, en présentant le résultat de convergence de [51]. Nous étudierons le cas périodique dans la dernière section 4.4, en

nous appuyant sur le travail [52], ainsi que sur [50] ; il ne sera pas question de domaines à bords dans cette étude du cas non isentropique (pour une étude du cas de domaines à bords nous renvoyons à l'article de S. Schochet [55] pour des données bien préparées, ainsi qu'au travail récent de T. Alazard [1] dans le cas général). Notons pour terminer que des études formelles dans le cas visqueux (périodique) ont été menées dans [6] ; ici comme partout ailleurs dans ce texte l'on se restreint aux équations sans viscosité.

4.3. Asymptotique dans \mathbb{R}^d

Commençons par énoncer le théorème de convergence démontré dans [51].

THÉORÈME 4.2 ([50]-[51]). — Soit $s \geq s_0$ et soit $\Omega = \mathbb{R}^d$. On suppose que les données initiales $(v_{\varepsilon,0}, q_{\varepsilon,0}, S_{\varepsilon,0})$ sont uniformément bornées dans $H^s(\Omega)$ et convergent vers (v_0, q_0, S_0) dans $H^s(\Omega)$ quand ε tend vers zéro, avec l'hypothèse de décroissance à l'infini suivante pour S_0 :

$$(21) \quad |S_0(x)| \leq \frac{C}{|x|^{1+\delta}} \quad \text{et} \quad |\nabla S_0(x)| \leq \frac{C}{|x|^{2+\delta}}, \quad \delta > 0.$$

Alors la famille de solutions $(v_\varepsilon, q_\varepsilon, S_\varepsilon)$ construite dans le théorème 4.1 converge dans l'espace $L^2([0, T], H_{loc}^{s'}(\Omega))$ pour tout $s' < s$, vers $(v, 0, S)$ où (v, S) est l'unique solution de (18) avec donnée (w_0, S_0) et w_0 est l'unique solution de

$$\operatorname{div} w_0 = 0, \quad \operatorname{rot}(\bar{r}_0 w_0) = \operatorname{rot}(\bar{r}_0 v_0), \quad \text{où} \quad \bar{r}_0 \stackrel{\text{déf}}{=} \mathcal{R}(S_0, 0).$$

Donnons une idée de la démonstration de ce résultat.

La première étape consiste à obtenir de la compacité en espace, en utilisant les estimations uniformes données par le théorème 4.1. Ainsi l'on peut montrer facilement que S_ε converge fortement (quitte à extraire une sous-suite, ce que nous omettrons de signaler dorénavant) vers une fonction S dans $C^0([0, T], H_{loc}^{s'}(\Omega))$ et que $(q_\varepsilon, v_\varepsilon)$ converge faiblement dans $L^\infty([0, T], H^s(\Omega))$ vers un certain (q, v) . Enfin $\operatorname{rot}(r_\varepsilon v_\varepsilon)$ converge fortement dans $L^\infty([0, T], H_{loc}^{s'-1}(\Omega))$ vers $\operatorname{rot}(r_0 v)$, où nous rappelons que $r_0 = \mathcal{R}(S, 0)$. On se convainc aussi (il suffit de multiplier les équations par ε et de passer à la limite) que

$$q = 0, \quad \operatorname{div} v = 0 \quad \text{et} \quad \partial_t S + v \cdot \nabla S = 0.$$

Reste à montrer le point le plus délicat du théorème, c'est-à-dire la convergence forte de q_ε vers zéro, et celle de v_ε vers v vérifiant le système limite (18).

Comme dans le cas isentropique, on commence par décomposer v_ε en une composante sur le noyau de l'opérateur de pénalisation (qui sera compacte en temps et en espace), et un reste : écrivons donc

$$v_\varepsilon = \bar{v}_\varepsilon + \tilde{v}_\varepsilon, \quad \text{avec} \quad P v_\varepsilon = \bar{v}_\varepsilon,$$

où nous rappelons que P est le projecteur de Leray sur les champs de vecteurs de divergence nulle. La suite \bar{v}_ε est compacte en temps, elle converge donc fortement vers v et l'on va maintenant s'attacher à montrer que $\tilde{V}_\varepsilon \stackrel{\text{déf}}{=} (q_\varepsilon, \tilde{v}_\varepsilon)$ converge fortement vers zéro. Pour

cela il faut contrôler les oscillations en temps de cette suite de fonctions, ce qui passe par l'étude du linéarisé

$$\partial_t(\mathcal{A}(S_\varepsilon, 0)\partial_t W_\varepsilon) - \frac{1}{\varepsilon^2} \operatorname{div} \left(\frac{1}{\mathcal{R}(S_\varepsilon, 0)} \nabla W_\varepsilon \right) = 0.$$

Ce système est proche de l'équation des ondes (8) au fait près qu'il est à coefficients variables. Ainsi le calcul direct en variables de Fourier que nous avons pratiqué en section 3.1.1 pour obtenir la dispersion des ondes acoustiques n'est plus opérant ici. Nous obtiendrons à la place un résultat de décroissance de l'énergie locale qui permettra de conclure. Un tel résultat est obtenu dans [51] par utilisation de la théorie des mesures de défaut. Nous renvoyons le lecteur à [21] et [60] pour l'introduction de cette théorie (ainsi qu'à [8] pour une présentation générale et des applications).

L'idée de la démonstration de [51] est assez simple à décrire, et nous allons en rester à cette description générale sans rentrer dans l'implémentation technique de la preuve (celle-ci requiert entre autres une écriture par paquets d'ondes (en temps) : un résumé de la méthode peut être trouvé dans [50]). Notons \mathcal{K} (respectivement \mathcal{L}) l'espace des opérateurs compacts (respectivement à trace) sur $L^2(\mathbb{R}^d)$. Les mesures de défaut microlocales de \tilde{V}_ε se définissent de la manière suivante : quitte à extraire une sous-suite, il existe une mesure de Radon positive μ sur \mathbb{R}^2 et il existe M μ -intégrable à valeurs dans \mathcal{L} telle que pour toute fonction a dans $\mathcal{S}(\mathbb{R}^2)$ à valeurs dans \mathcal{K} , l'opérateur pseudo-différentiel semi-classique associé $A \stackrel{\text{d\u00e9f}}{=} a(t, \varepsilon D_t)$ vérifie

$$\lim_{\varepsilon \rightarrow 0} \left(A \tilde{V}_\varepsilon, \tilde{V}_\varepsilon \right) = \int_{\mathbb{R}^2} \operatorname{Tr} (M(t, \tau) a(t, \tau)) \mu(dt, d\tau).$$

Le point important est que les mesures de défaut microlocales sont supportées sur la variété caractéristique de l'équation, ce qui signifie ici que

$$\mathcal{A}(S(t, \cdot), 0) \tau^2 M(t, \tau) + \operatorname{div} \left(\frac{1}{\mathcal{R}(S(t, \cdot), 0)} \nabla \right) M(t, \tau) = 0, \quad \mu \text{ p.p.}$$

On utilise alors le fait que la condition (21) est propagée par l'équation de transport $\partial_t S + v \cdot \nabla S = 0$ ainsi que le résultat suivant (dont une démonstration peut se trouver dans [51], paragraphe 5) : le noyau de $a\tau^2 + \operatorname{div}(b\nabla)$ est réduit à zéro, pour tout $\tau \in \mathbb{R}$, s'il existe une constante \bar{a} telle que

$$|a(x) - \bar{a}| \leq \frac{C}{(1 + |x|)^{1+\delta}}, \quad |\nabla a(x)| \leq \frac{C}{(1 + |x|)^{2+\delta}}$$

et de même pour b (avec une autre constante $\bar{b} \neq 0$). En combinant ces deux propriétés l'on obtient que $M = 0$. En corollaire on peut en déduire que la mesure de défaut de \tilde{V}_ε est nulle, et les estimations a priori sur \tilde{v}_ε et q_ε permettent finalement de démontrer que \tilde{V}_ε converge fortement vers zéro dans $L^2([0, T], H_{loc}^{s'}(\mathbb{R}^d))$, pour tout $s' < s$.

On conclut enfin la démonstration en passant à la limite dans l'équation en v_ε : tous les termes de l'équation ont une limite forte, et l'équation (18) est obtenue sans difficulté.

L'implémentation de ces principes est techniquement délicate et nous renvoyons donc le lecteur à [51] pour la démonstration détaillée (un bref exposé peut être trouvé dans [50]).

4.4. Asymptotique dans \mathbb{T}^d

Dans cette section finale nous allons nous intéresser aux équations non isentropiques (17) dans le cas où le domaine d'espace est périodique.

Le résultat que nous allons présenter est celui de G. Métivier et S. Schochet [52] (nous nous inspirerons aussi de l'exposé [50]). Considérons une suite de solutions $(q_\varepsilon, v_\varepsilon, S_\varepsilon)_{\varepsilon>0}$ donnée par le théorème 4.1, associée à une famille de données initiales $(q_{\varepsilon,0}, v_{\varepsilon,0}, S_{\varepsilon,0})_{\varepsilon>0}$. Cette suite de solutions vérifie les mêmes estimations *a priori* que dans le cas \mathbb{R}^d (rappelons que le théorème 4.1 est valable aussi bien dans tout l'espace que dans le cas périodique). L'étape de compacité en espace se déroule donc aussi bien que dans la section 4.3 précédente. On a donc convergence forte de S_ε et $\text{rot}(r_\varepsilon v_\varepsilon)$ dans $C^0([0, T], H^{s'}(\mathbb{T}^d))$ et $C^0([0, T], H^{s'-1}(\mathbb{T}^d))$ respectivement, et le couple $(q_\varepsilon, v_\varepsilon)$ converge faiblement dans l'espace $L^\infty([0, T], H^s(\mathbb{T}^d))$ vers (q, v) , avec

$$(22) \quad \nabla q = 0, \quad \text{div } v = 0, \quad \partial_t S + v \cdot \nabla S = 0.$$

On peut aussi montrer (voir [52] Lemme 8.2) que q ne dépend pas non plus du temps, et est donc une constante en t et en x . Cherchons maintenant l'équation vérifiée par v . C'est ici que se fait la différence avec le cas \mathbb{R}^d : de la même façon que dans le cas isentropique, on ne peut espérer *a priori* dans le cas périodique de convergence forte sur les composantes acoustiques, et l'on va introduire la décomposition suivante :

$$v_\varepsilon = \bar{v}_\varepsilon + \frac{1}{r_\varepsilon} \nabla h_\varepsilon \quad \text{où} \quad \text{div } \bar{v}_\varepsilon = 0.$$

Comme précédemment \bar{v}_ε converge fortement vers v dans $C^0([0, T], H^{s'}(\mathbb{T}^d))$, et l'on peut passer à la limite au sens des distributions dans tous les termes de l'équation vérifiée par v_ε , sauf dans les termes bilinéaires en h_ε (puisque h_ε ne converge que faiblement). En suivant les notations de [50], posons $\mathcal{B} \stackrel{\text{déf}}{=} \frac{\partial \mathcal{R}}{\partial \rho}$ et $b_0 \stackrel{\text{déf}}{=} \mathcal{B}(S, 0)$. En écrivant

$$\mathcal{E} \stackrel{\text{déf}}{=} -\frac{b_0}{2r_0} \nabla \kappa^{(1)} + \frac{1}{2r_0} \nabla \kappa^{(2)}$$

$$\kappa^{(1)} \stackrel{\text{déf}}{=} \lim_{\varepsilon \rightarrow 0} |\varepsilon \partial_t h_\varepsilon|^2 \quad \text{et} \quad \kappa^{(2)} \stackrel{\text{déf}}{=} \lim_{\varepsilon \rightarrow 0} |\nabla h_\varepsilon|^2,$$

il vient à la limite (rappelons que $r_0 = \mathcal{R}(S, 0)$)

$$(23) \quad r_0(\partial_t v + v \cdot \nabla v) + \mathcal{E} + \nabla P = 0, \quad \text{div } v = 0,$$

avec la même donnée initiale que dans le cas \mathbb{R}^d : $v|_{t=0} = w_0$ où w_0 est l'unique solution de $\text{div } w_0 = 0$, $\text{rot}(r_0 w_0) = \text{rot}(r_0 v_0)$.

Le problème maintenant est de clore le système (22, 23) en déterminant \mathcal{E} ; il n'y a pas de raison *a priori* pour que \mathcal{E} soit un gradient.

Étudions un instant (comme dans [50] et [52]) le système modèle suivant :

$$\varepsilon \partial_t u_\varepsilon = A(S_\varepsilon(t))u_\varepsilon,$$

où A est un opérateur antisymétrique. Dans le cas où $S_\varepsilon = S$ ne dépend pas du temps, la solution est donnée par $u_\varepsilon = e^{\frac{tA(S)}{\varepsilon}}u_0$. Dans le cas contraire on s'aperçoit immédiatement d'une difficulté considérable liée au croisement des valeurs propres de $A(S(t))$. En effet si le spectre de A est de multiplicité constante il est connu (voir [32]) que l'évolution de u_ε est asymptotiquement diagonalisée dans la résolution spectrale de $A(S(t))$. Mais lorsque deux valeurs propres se croisent en un temps t_0 , la limite de $|u_\varepsilon|^2$ peut dépendre non seulement de S , mais aussi de la façon dont S_ε approche S , et même encore de la sous-suite en ε choisie... Des exemples à ce sujet peuvent être trouvés dans [52] Section 7, s'inspirant de travaux de [18], [24], [31]. Dans notre cadre, cela semble indiquer que si les valeurs propres de $A(S(t))$ ne sont pas de multiplicité constante, la fonction \mathcal{E} ne sera pas uniquement déterminée par v et S .

L'on constate ainsi que le cas non isentropique périodique est redoutablement plus difficile que le cas non isentropique dans \mathbb{R}^d , mais aussi surtout que le cas isentropique périodique.

L'idée de G. Métivier et S. Schochet dans [52] est de tirer parti des estimations *a priori* disponibles sur les solutions, dues au théorème 4.1, pour commencer l'étude en se restreignant à un nombre fini de modes de Fourier. Cela signifie que l'on peut commencer par considérer des modèles en dimension finie. La première partie de [52] (et la section 5 de [50]) est ainsi dévolue à l'étude du système dynamique suivant, à deux échelles en temps :

$$(24) \quad \begin{aligned} \varepsilon \partial_t u_\varepsilon &= A(S)u_\varepsilon + \varepsilon Q(S)(u_\varepsilon, u_\varepsilon) \\ \partial_t S_\varepsilon &= F(S)u_\varepsilon \end{aligned}$$

sur le domaine $\mathbb{D} \stackrel{\text{déf}}{=} [0, L_1] \times \cdots \times [0, L_d]$ où les L_i sont des réels strictement positifs.

Des hypothèses de structure sur A et F sont dégagées, portant notamment sur les valeurs propres de A : l'ensemble des S tels que $A(S)$ a au moins une valeur propre non nulle multiple, est analytique et de codimension 2. Cette hypothèse est cruciale dans l'analyse car elle permet de montrer que génériquement les valeurs propres non nulles de $A(S(t))$ sont simples. Notons que dans [52], il est montré (Théorème 6.1) que sous certaines restrictions sur les tailles L_i (L_j^{-4} et $L_j^{-2}L_k^{-2}$ doivent être indépendantes dans \mathbb{Q}) ces hypothèses sont toujours satisfaites dans le cas où le système modèle est une troncature spectrale des équations d'Euler compressibles non isentropiques.

L'hypothèse importante de généricité est alors la suivante : en notant $(i\lambda_j(S(t)))_{j \neq 0}$ les valeurs propres non nulles de $A(S(t))$ (répétées suivant leur multiplicité) on définit

$$\Omega \stackrel{\text{déf}}{=} \{S \mid \lambda_j(S(t)) \neq \lambda_k(S(t)), \forall j \neq k, jk \neq 0\}.$$

DÉFINITION 4.3. — *La fonction S est générique si pour tout $t \in [0, T]$, $S(t) \in \Omega$ et*

$$\lambda_j(S) - \lambda_k(S) = \lambda_\ell(S) \Rightarrow \partial_t (\lambda_j(S) - \lambda_k(S) - \lambda_\ell(S)) \neq 0.$$

Sous cette hypothèse sur S on peut obtenir une équation sur la limite faible de u_ε , qui a une solution unique dont on montre que pour presque toutes données initiales elle est générique au sens de la définition 4.3 (ce qui justifie bien la terminologie employée dans cette définition). Ainsi l'on obtient un théorème de convergence des solutions $(u_\varepsilon, q_\varepsilon, S_\varepsilon)$ vers les solutions du système limite.

Nous n'allons pas détailler ce système limite dans le cadre abstrait, mais revenir plutôt aux équations d'Euler pour déduire de l'analyse abstraite des théorèmes de convergence des solutions des équations d'Euler non isentropiques (17). La matrice de pénalisation ici est la suivante

$$A(S) = \begin{pmatrix} 0 & \frac{1}{a_0} \operatorname{div} \\ \frac{1}{r_0} \nabla & 0 \end{pmatrix}.$$

On rappelle que $a_0 = \mathcal{A}(S, 0)$ et $r_0 = \mathcal{R}(S, 0)$.

Le spectre de $A(S)$ s'obtient en analysant l'opérateur $W_S \stackrel{\text{déf}}{=} -\frac{1}{a_0} \operatorname{div} \left(\frac{1}{r_0} \nabla \right)$, autoadjoint positif dans $L^2(\mathbb{D}, a \, dx)$, et à résolvante compacte. On note $0 = \mu_{0,S} < \mu_{1,S} \leq \dots$ ses valeurs propres, et $\lambda_{j,S} \stackrel{\text{déf}}{=} \sqrt{\mu_{j,S}}$. Les vecteurs propres (normalisés) associés sont notés $\Phi_{j,S}$, et le théorème de convergence va reposer sur l'hypothèse suivante, liée à la définition de généricité 4.3.

DÉFINITION 4.4. — *La fonction $S \in C^0([0, T], H^s(\mathbb{D}))$ vérifie la condition (G) si*

- (i) *Pour tout $t \in [0, T]$, pour tout $j > 0$, $\mu_{j,S(t)}$ est simple.*
- (ii) *Pour tout triplet (j, k, l) d'entiers non nuls, on a presque partout sur $[0, T]$,*

$$\lambda_{j,S(t)} - \lambda_{k,S(t)} \neq \lambda_{l,S(t)}.$$

Le théorème principal est le suivant.

THÉORÈME 4.5 ([50], [52]). — *Si S vérifie la condition (G) alors (v, S) vérifient les équations limites (22,23) avec*

$$\kappa^{(1)} = \frac{1}{2} \sum_{j=1}^{\infty} \sigma_j \lambda_{j,S(t)} (\Phi_{j,S(t)}(x))^2, \quad \kappa^{(2)} = \frac{1}{2} \sum_{j=1}^{\infty} \frac{\sigma_j}{\lambda_{j,S(t)}} |\nabla \Phi_{j,S(t)}(x)|^2$$

où σ_j est calculé explicitement en fonction des données initiales et des Φ_{j,S_0} .

La démonstration de ce résultat sort du cadre de cet exposé. Indiquons simplement que la démonstration consiste comme prévu à se ramener à la dimension finie en faisant agir un projecteur spectral sur h_ε , en tirant parti des estimations a priori fournies par le théorème 4.1. On peut alors décomposer h_ε sur les J premiers vecteurs propres de $W_{S(t)}$ (pour $\varepsilon \leq \varepsilon_J$) — en utilisant la compacité de S_ε en temps on peut montrer que J peut

être choisi indépendamment de $t \in [0, T]$. Les termes non linéaires dont on cherche la limite faible s'écrivent de la manière suivante :

$$\sum_{j \leq J} \beta_{j,\ell}^\varepsilon \exp \left(\pm \frac{i}{\varepsilon} \int_0^t \lambda_j^\varepsilon(s) ds \pm \frac{i}{\varepsilon} \int_0^t \lambda_\ell^\varepsilon(s) ds \right).$$

Notons qu'il n'y a que des interactions bilinéaires dans la phase, parce que l'on cherche la projection de ces termes non linéaires sur le noyau de l'opérateur de pénalisation, c'est-à-dire sur le mode $\Phi_{0,S(t)}$ associé à la valeur propre 0. On peut montrer que les $\beta_{j,\ell}^\varepsilon$ ont une limite forte quand ε tend vers zéro. Par le théorème de la phase stationnaire, ne demeurent à la limite dans la somme que les indices j, ℓ tels que

$$\pm \int_0^t \lambda_j^\varepsilon(s) ds \pm \int_0^t \lambda_\ell^\varepsilon(s) ds = 0,$$

c'est-à-dire que $\ell = j$ et les signes sont opposés.

Il reste alors à vérifier que les séries obtenues sont convergentes dans $C^0([0, T], H^s(\mathbb{D}))$, ce qui conclut la première partie de l'analyse : le terme \mathcal{E} est obtenu ainsi. Il reste maintenant à déterminer l'évolution temporelle des coefficients de cette série, afin d'obtenir l'expression des coefficients $\kappa^{(1)}$ et $\kappa^{(2)}$. Des hypothèses de non résonance interviennent ici, ainsi que des éléments de théorie spectrale. Nous ne rentrons pas dans les détails techniques mais renvoyons à [52], Section 9.

Nous commenterons ce résultat, en présentant les problèmes ouverts qui y sont rattachés, dans la section 4.5 suivante. Pour terminer cet exposé, plaçons-nous en dimension 1 d'espace : ce cas est bien plus favorable, d'abord parce que $W_{S(t)}$ y prend une forme plus simple (voir [52] Section 11), mais surtout parce que la contrainte incompressible sur la limite v se traduit en dimension 1 par $v(t, x) = \underline{v}(t)$, constant en x .

Le théorème dans ce cas est le suivant.

THÉORÈME 4.6 ([50], [52]). — *Supposons que $\mathcal{R}_0 \stackrel{\text{déf}}{=} \mathcal{R}(S, 0)$ et $\mathcal{A}_0 \stackrel{\text{déf}}{=} \mathcal{A}(S, 0)$ vérifient*

$$\forall \sigma \in \mathbb{R}, \quad \frac{\partial \mathcal{R}_0}{\partial S}(\sigma) \neq 0 \quad \text{et} \quad \mathcal{A}_0(\sigma) - \frac{\partial \mathcal{A}_0}{\partial S}(\sigma) \neq 0.$$

Alors pour tout $s \geq 2$, il existe un G_δ -dense $\mathcal{G} \subset H^s(\mathbb{R}/L\mathbb{Z})$ tel que pour tout S dans \mathcal{G} , le spectre de W_S est simple et la condition de non résonance est vérifiée : $\lambda_{j,S} \neq \lambda_{k,S} + \lambda_{\ell,S}$.

En particulier, $S(t)$ vérifie alors la condition (G) et les conclusions du théorème 4.5 sont vraies.

REMARQUE 4.7. — Notons comme dans [52] que dans le cas des équations compressibles on a en fait $\mathcal{A}(S, p) = \frac{1}{\mathcal{R}} \frac{\partial \mathcal{R}}{\partial p}$. Il n'y a alors pas d'hypothèse de généricité à faire dans le théorème 4.6, et l'on trouve pour la limite faible de v_ε la solution de l'équation $\frac{d}{dt} \underline{v} = 0$, et l'équation de transport (22) sur S s'écrit avec v remplacé par \underline{v} . Les données initiales du système limite sont les moyennes spatiales des données initiales du système d'origine (voir le théorème 11.1 de [52]).

4.5. Conclusion et problèmes ouverts

Nous avons présenté dans ce paragraphe les résultats mathématiques les plus récents à ce jour concernant la limite incompressible non visqueuse. Le cas où les équations sont posées dans \mathbb{R}^d est bien compris puisqu'on démontre ([51]) un résultat de convergence forte vers un système limite connu.

Dans le cas périodique en revanche, l'article [52] est un véritable travail de pionnier qui ouvre la voie sur beaucoup de recherches. Une question cruciale est de déterminer si la condition (G) de la définition 4.3 est ou non générique — pour l'instant seul le cas monodimensionnel est résolu, toujours dans [52]. Une autre question fondamentale est de comprendre le système limite obtenu dans le théorème 4.5 ; en particulier ce système est-il bien posé ? On ne connaît pour l'instant que des estimations *a priori* H^s sur ce système (voir [52] Section 12). En outre la convergence vers ce système limite est faible, et en particulier on n'a pas de description assez précise des oscillations pour pouvoir en déduire un résultat du même type que dans le cas isentropique : dans ce cas après action du groupe d'oscillations on a pu montrer la convergence forte vers un système limite dont une partie est celui vérifié par la limite faible, mais qui contient aussi une équation couplée gardant trace des oscillations.

Toutes ces questions restent largement ouvertes pour les équations non isentropiques, sans parler du cas où les équations sont posées dans un domaine plus général que \mathbb{R}^d ou \mathbb{T}^d , qui semble pour l'instant hors d'atteinte.

BIBLIOGRAPHIE

- [1] T. Alazard – Incompressible limit of the non-isentropic Euler equations with solid wall boundary conditions, *soumis*.
- [2] V. Arnold – Sur la géométrie différentiable des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, *Annales de l'Institut Fourier*, **16** (1966), 319–361.
- [3] H. Bahouri et J.-Y. Chemin – Équations d'ondes quasilineaires et estimation de Strichartz, *American Journal of Mathematics*, **121** (1999), 1337–1377.
- [4] H. Beirao da Veiga – On the sharp singular limit for slightly compressible fluids, *Math. Methods Appl. Sci.*, **18** (1995), no. 4, 295–306.
- [5] H. Beirao da Veiga – Singular limits in compressible fluid dynamics, *Arch. Rational Mech. Anal.*, **128** (1994), no. 4, 313–327.
- [6] D. Bresch, B. Desjardins, E. Grenier et C.K. Lin – Low Mach number limit of viscous polytropic flows: formal asymptotics in the periodic case, *Stud. Appl. Math.*, **109** (2002), no. 2, 125–149.

- [7] G. Browning et H.-O. Kreiss – Problems with different time scales for nonlinear partial differential equations, *SIAM Journal of Applied Mathematics*, **42** (1982), 4, 704–718.
- [8] N. Burq – Mesures semi-classiques et mesures de défaut, *Sém. Bourbaki* (1996/97), Exp. No. 826, Astérisque **245** (1997), 167–195.
- [9] J.-Y. Chemin – *Fluides parfaits incompressibles*, Astérisque **230**, 1995.
- [10] R. Danchin – Fluides légèrement compressibles et limite incompressible, *Séminaire Équations aux Dérivées Partielles de l'École polytechnique* 2000–2001, Exp. No. III, 19 pages, 2001.
- [11] R. Danchin – Zero Mach number limit for compressible flows with periodic boundary conditions, *Amer. J. Math.* **124** (2002), no. 6, 1153–1219.
- [12] R. Danchin – Zero Mach number limit in critical spaces for compressible Navier-Stokes equations, *Ann. Sci. École Norm. Sup. (4)* **35** (2002), no. 1, 27–75.
- [13] B. Desjardins et E. Grenier – Low Mach number limit of viscous compressible flows in the whole space, *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, **455** (1999), no. 1986, 2271–2279.
- [14] B. Desjardins, E. Grenier, P.-L. Lions et N. Masmoudi – Incompressible limit for solutions of the isentropic Navier-Stokes equations with Dirichlet boundary conditions, *J. Math. Pures Appl.*, (9) **78** (1999), no. 5, 461–471.
- [15] B. Desjardins et C.-K. Lin – A survey of the compressible Navier-Stokes equations, *Taiwanese J. Math.*, **3** (1999), no. 2, 123–137.
- [16] A. Dutrifoy et T. H'midi – The incompressible limit of solutions of the two-dimensional compressible Euler system with degenerating initial data, *C. R. Math. Acad. Sci. Paris* **336** (2003), no. 6, 471–474.
- [17] W. E – Propagation of oscillations in the solutions of 1-D compressible fluid equations, *Comm. Partial Differential Equations*, **17** (1992), no. 3-4, 347–370.
- [18] C. Fermanian et P. Gérard – Mesures semi-classiques et croisements de mode, *Bull. S.M.F.*, **130** 1 (2002), 123–168.
- [19] I. Gallagher – Asymptotics of the solutions of hyperbolic equations with a skew-symmetric perturbation, *Journal of Differential Equations* **150** (1998), 363–384, 1998.
- [20] I. Gallagher – A remark on smooth solutions of the weakly compressible Navier-Stokes equations, *Journal of Mathematics of Kyoto University*, **40** (2000), 3, 525–540.
- [21] P. Gérard – Microlocal defect measures, *Communications in Partial Differential Equations*, **16** (1991), 1761–1794.
- [22] J. Ginibre et G. Velo – Generalized Strichartz inequalities for the wave equation, *Journal of Functional Analysis*, **133** (1995), 50–68.

- [23] E. Grenier – Oscillatory perturbations of the Navier–Stokes equations, *Journal de Mathématiques Pures et Appliquées*, **76** (1997), 477–498.
- [24] G. Hagedorn – Proof of the Landau-Zener formula in an adiabatic limit with small eigenvalue gap, *Comm. Math. Phys.*, **110** (1987), 519–524.
- [25] T. Hagstrom et J. Lorenz – All-time existence of classical solutions for slightly compressible flows, *SIAM Journal of Mathematical Analysis*, **28** (1998), 3, 652–672.
- [26] T. Hagstrom et J. Lorenz – On the stability of approximate solutions of hyperbolic-parabolic systems and the all-time existence of smooth, slightly compressible flows, *Indiana Univ. Math. J.*, **51** (2002), no. 6, 1339–1387.
- [27] T. Iguchi – The incompressible limit and the initial layer of the compressible Euler equation in \mathbb{R}_+^n , *Math. Methods Appl. Sci.*, **20** (1997), no. 11, 945–958.
- [28] H. Isozaki – Wave operators and the incompressible limit of the compressible Euler equation, *Comm. Math. Phys.*, **110** (1987), no. 3, 519–524.
- [29] H. Isozaki – Singular limits for the compressible Euler equation in an exterior domain. *J. Reine Angew. Math.*, **381** (1987), 1–36.
- [30] H. Isozaki – Singular limits for the compressible Euler equation in an exterior domain. II. Bodies in a uniform flow, *Osaka J. Math.*, **26** (1989), no. 2, 399–410.
- [31] A. Joye – Proof of the Landau-Zener formula, *Asymptotic Analysis*, **9** (1994), 209–258.
- [32] T. Kato – *Perturbation Theory for Linear Operators*, Springer–Verlag, Berlin, 1980.
- [33] S. Klainerman – Global existence for nonlinear wave equations, *Communications in Pure and Applied Mathematics*, **33**, (1980) 43–101.
- [34] S. Klainerman et M. Machedon – Remark on Strichartz type inequalities (with an appendix of J. Bourgain and D. Tataru), *International Math. Research Notices*, **5** (1996), 201–220.
- [35] S. Klainerman et A. Majda – Singular limits of quasilinear hyperbolic systems with large parameters, and the incompressible limit of compressible fluids, *Communications on Pure and Applied Mathematics*, **34** (1981), 481–524.
- [36] S. Klainerman et A. Majda – Compressible and incompressible fluids, *Communications on Pure and Applied Mathematics*, **35** (1982), 629–651.
- [37] R. Klein – Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics, I, One-dimensional flow, *J. Comput. Phys.*, **121** (1995), no. 2, 213–237.
- [38] H.-O. Kreiss, J. Lorenz et M. J. Naughton – Convergence of the solutions of the compressible to the solutions of the incompressible Navier-Stokes equations, *Adv. in Appl. Math.*, **12** (1991), no. 2, 187–214.
- [39] P. Lax – Hyperbolic systems of conservation laws II. *Communications on Pure and Applied Mathematics*, **10** (1957), 537–566.

- [40] P.-L. Lions – *Mathematical Topics in Fluid Mechanics*, Vol. I, Incompressible Models, Oxford Science Publications, 1997.
- [41] P.-L. Lions – *Mathematical Topics in Fluid Mechanics*, Vol. II, Compressible Models, Oxford Science Publications, 1997.
- [42] P.-L. Lions et N. Masmoudi – Une approche locale de la limite incompressible, *C. R. Acad. Sci. Paris Sér. I Math.* **329** (1999), no. 5, 387–392.
- [43] P.-L. Lions et N. Masmoudi – Incompressible limit for a viscous compressible fluid, *J. Math. Pures Appl.* (9) **77** (1998), no. 6, 585–627.
- [44] A. Majda – *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*, Applied Mathematical Sciences, **53**, Springer-Verlag, New York, 1984.
- [45] A. Majda et S. Osher – Initial-boundary value problems for hyperbolic equations with uniformly characteristic boundary, *Comm. Pure Appl. Math.*, **28** (1975), 607–675.
- [46] C. Marchioro et M. Pulvirenti – *Mathematical Theory of Incompressible Nonviscous Fluids*, Applied Mathematical Sciences, **96**, Springer-Verlag, New York, 1994.
- [47] N. Masmoudi – Incompressible, inviscid limit of the compressible Navier–Stokes system, *Annales de l’Institut H. Poincaré, Anal. non linéaire*, **18** (2001), 2, 199–224.
- [48] N. Masmoudi – Asymptotic problems and compressible-incompressible limit, *Advances in mathematical fluid mechanics* (Paseky, 1999), 119–158, Springer, Berlin, 2000.
- [49] A. Meister – Asymptotic single and multiple scale expansions in the low Mach number limit, *SIAM J. Appl. Math.*, **60** (1999), no. 1, 256–271.
- [50] G. Métivier et S. Schochet – Limite incompressible des équations d’Euler non isentropiques, *Séminaire Équations aux Dérivées Partielles de l’École polytechnique 2000–2001*, Exp. No. X, 17 pages, 2001.
- [51] G. Métivier et S. Schochet – The incompressible limit of the non-isentropic Euler equations, *Arch. Ration. Mech. Anal.* **158** (2001), no. 1, 61–90.
- [52] G. Métivier et S. Schochet – Averaging theorems for conservative systems and the weakly compressible Euler equations, *J. Differential Equations*, **187** (2003), no. 1, 106–183.
- [53] C.D. Munz – Computational fluid dynamics and aeroacoustics for low Mach number flow, *Hyperbolic partial differential equations* (Hamburg, 2001), 269–320, Vieweg, Braunschweig, 2002.
- [54] T. Schneider, N. Botta, K.J. Geratz et R. Klein – Extension of finite volume compressible flow solvers to multi-dimensional, variable density zero Mach number flows, *J. Comput. Phys.*, **155** (1999), no. 2, 248–286.

- [55] S. Schochet – The compressible Euler equations in a bounded domain: existence of solutions and the incompressible limit, *Comm. Math. Phys.*, **104** (1986), 49–75.
- [56] S. Schochet – Fast singular limits of hyperbolic PDEs, *Journal of Differential Equations*, **114** (1994), 476–512.
- [57] P. Secchi – On the incompressible limit of inviscid compressible fluids, Navier-Stokes equations and related nonlinear problems (Ferrara, 1999), *Ann. Univ. Ferrara Sez. VII (N.S.)*, **46** (2000), 21–33.
- [58] P. Secchi – On the singular incompressible limit of inviscid compressible fluids, *J. Math. Fluid Mech.*, **2** (2000), no. 2, 107–125.
- [59] T. Sideris – Formation of singularities in three-dimensional compressible fluids, *Comm. Math. Phys.*, **101** (1985), no. 4, 475–485.
- [60] L. Tartar – H-measures, a new approach for studying homogenization, oscillations and concentration effects in partial differential equations, *Proc. Roy. Soc. Edinburgh, Sect. A*, **115** (1990), 193–230.
- [61] S. Ukai – The incompressible limit and the initial layer of the compressible Euler equation. *Journal of Mathematics of Kyoto University*, **26** (1986), no. 2, 323–331.
- [62] V. Yudovitch – Non stationary flows of an ideal incompressible fluid, *Zhurnal Vych Matematika*, **3** (1963), 1032–1066.
- [63] R. Zeytounian – *Theory and applications of nonviscous fluid flows*, Springer-Verlag, Berlin, 2002.
- [64] R. Zeytounian – *Asymptotic modelling of fluid flow phenomena*, Fluid Mechanics and its Applications, **64**, Kluwer Academic Publishers, Dordrecht, 2002.

Isabelle GALLAGHER

ÉCOLE POLYTECHNIQUE

Centre de Mathématiques

UMR 7640 du CNRS

F-91128 PALAISEAU Cédex

E-mail: Isabelle.Gallagher@math.polytechnique.fr

DÉVIATIONS DE MOYENNES ERGODIQUES

[d'après Forni, Kontsevich, Zorich...]

par Raphaël KRIKORIAN

1. INTRODUCTION

Considérons un espace topologique X sur lequel agit un flot ϕ et supposons que μ soit une mesure de probabilité invariante ergodique pour le flot ϕ (c'est-à-dire que les seuls ensembles μ mesurables ϕ -invariants sont de mesure 0 ou 1). On sait d'après le théorème de Birkhoff que pour toute observable μ -intégrable $f : X \rightarrow \mathbf{R}$ on a pour μ -presque tout $x \in X$

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\phi^s(x)) ds = \int_X f d\mu.$$

En particulier si f est d'intégrale nulle, les moyennes ergodiques précédentes convergent vers 0 et il est naturel d'étudier dans ce cas le comportement asymptotique des intégrales ergodiques $\int_0^t f(\phi^s(x)) ds$. Lorsque l'on suppose que le système dynamique est hyperbolique, comme c'est le cas par exemple du flot géodésique sur des surfaces compactes de courbure négative, le comportement des intégrales ergodiques est décrit par le théorème central limite (comme dans le cas de variables aléatoires indépendantes de même loi). Pour des flots non hyperboliques, les situations dans lesquelles on peut espérer énoncer des résultats intéressants sont peu nombreux mais existent. Ainsi, pour mentionner l'exemple le plus simple, le comportement des intégrales ergodiques des flots (linéaires car le cas général s'y ramène) sur des tores est bien compris. Soit ϕ le flot du champ de vecteurs $X = \alpha \frac{\partial}{\partial x} + \beta \frac{\partial}{\partial y}$ sur le tore $\mathbf{T}^2 = \mathbf{R}^2 / \mathbf{Z}^2$; supposons que

- (hypothèse sur l'arithmétique) (α, β) vérifie une condition diophantienne de la forme

$$|k\alpha + l\beta| \geq \frac{K^{-1}}{(|k| + |l|)^\tau}, \quad \forall (k, l) \in \mathbf{Z}^2 - \{0, 0\}$$

condition qui, dès que $\tau > 1$, est de mesure de Lebesgue positive pour $K > 0$ assez grand et de mesure totale si l'on prend l'union sur tous les K positifs ;

- (hypothèse sur la régularité de f) f soit suffisamment dérivable (p. ex $C^{\tau+2}$).

Dans ce cas il est facile de voir, en utilisant l'analyse de Fourier (bien définie sur le tore !) et la condition diophantienne, que les intégrales ergodiques sont bornées pour toute f régulière et tout $x \in \mathbf{T}^2$. Dans le cas où (α, β) n'est pas diophantien mais seulement irrationnel et f est à variations bornées, les intégrales ergodiques ne sont plus bornées mais admettent une borne de type logarithmique (en fonction de t). Il faut alors utiliser le développement en fractions continues de (α/β) (que l'on suppose irrationnel) et la propriété de Denjoy-Koksma.

Le cas qui nous intéresse ici est celui des flots des champs de vecteurs sur des surfaces de genre plus grand que 2 qui préservent une forme volume et dont les singularités sont de type selle. La dynamique dans ce cas n'est ni hyperbolique (comme c'est le cas du flot géodésique où deux points proches ont tendance à se séparer de façon exponentielle sous l'effet du flot), ni elliptique (comme dans le cas des flots sur le tore où deux points proches le restent sous l'effet de la dynamique) mais plutôt de type parabolique (comme dans le cas du flot horocyclique où deux points proches se séparent à vitesse au plus polynomiale). Ce type de flot intervient naturellement dans l'étude d'au moins deux problèmes importants de dynamique : les billards rationnels et les échanges d'intervalles.

Considérons un billard plan polygonal P dont les angles sont des multiples rationnels de 2π ; on peut réduire la dynamique des points $(x, v) \in P \times \mathbf{R}^2$, v étant de direction fixée, à celle d'un flot sur une surface de genre g obtenue en recollant des copies de la table (qui sont les images du polygone par le groupe engendré par les symétries par rapport à ses côtés). On peut ainsi voir que, pour presque toute direction v , le flot (dans l'espace des configurations) est uniquement ergodique (cf. [6] utilisant des résultats antérieurs de Masur [11], Veech [15]) et ceci permet de démontrer qu'il existe un G_δ -dense de billards (non rationnels) uniquement ergodiques (dans l'espace des phases). Je renvoie au séminaire Bourbaki de P. Arnoux [1] pour un exposé très clair de ces travaux.

Un échange d'intervalles sur n intervalles (I_1, \dots, I_n) qui forment une partition de $[0, 1]$ est déterminé par une paire (l, π) où $l = (|I_1|, \dots, |I_n|)$ (longueurs des intervalles) et π est une permutation de l'ensemble $\{1, \dots, n\}$. On suppose que la permutation est irréductible c'est-à-dire qu'il n'existe pas $k < n$ pour lequel $\pi\{1, \dots, k\} = \{1, \dots, k\}$. Notons S_n^0 l'ensemble des permutations irréductibles. L'ensemble des échanges d'intervalles irréductibles est donc paramétré par $\Delta^{n-1} \times S_n^0$ où Δ^{n-1} est le simplexe standard de dimension $n - 1$. Une rotation est un cas particulier d'échange d'intervalles où $n = 2$. On peut définir pour les échanges d'intervalles des procédures de *renormalisation* G (cf. Rauzy, Veech, Zorich, Marmi-Moussa-Yoccoz...) qui sont des analogues de l'algorithme de Gauss pour les rotations et qui permettent une analyse fine des propriétés ergodiques des échanges d'intervalles. Nous renvoyons le lecteur à [13] pour de plus amples renseignements. L'ensemble des échanges d'intervalles (irréductibles) $\Delta^{n-1} \times S_n^0$ se décompose sous l'action de G en sous-ensembles invariants de la forme $\Delta^{n-1} \times R$ et on appelle les

ensembles R ainsi obtenus les classes de Rauzy. Zorich [18] montre qu'une de ces applications de Gauss, l'algorithme de Zorich (qui est une version accélérée de l'algorithme de Rauzy), est ergodique par rapport à une mesure de probabilité μ_R absolument continue sur chaque classe de Rauzy. C'est l'analogie d'un théorème de Veech et Masur ([15],[11]) sur l'ergodicité du flot de Teichmüller sur les composantes connexes de l'espace des modules (cf. section 4). Ces liens entre flots et échanges d'intervalles sont en fait naturels : on peut voir un échange d'intervalles comme la dynamique induite par l'application de retour d'un flot sur un intervalle (transverse au flot) d'une surface de genre g . Considérons un échange d'intervalles T sur n intervalles I_1, \dots, I_n et notons,

$$S_i(x, N) = \sum_{k=0}^{N-1} \chi_i(T^k(x)),$$

la N -ième somme ergodique de la fonction caractéristique χ_i de l'intervalle I_i sous l'action de T (on compte le nombre de retours dans I_i entre les temps 0 et N). Formons le vecteur $S(x, N) = (S_1(x, N), \dots, S_n(x, N)) \in \mathbf{R}^n$. Si l'échange d'intervalles est uniquement ergodique (ce qui est une condition de mesure pleine) alors pour tout x

$$\lim_{N \rightarrow \infty} \frac{S(x, N)}{N} = l,$$

Zorich démontre le résultat suivant :

THÉORÈME 1.1 (Zorich [19]). — *Soit R une classe de Rauzy. Il existe $\theta_1 > \theta_2 \geq 0$ tels que pour presque tout échange d'intervalle et presque tout x*

$$\limsup_{N \rightarrow \infty} \frac{\log |S(x, N) - Nl|}{\log N} = \frac{\theta_2}{\theta_1} < 1.$$

Les réels θ_1, θ_2 sont en fait les deux plus grands exposants de Lyapunov d'un cocycle de matrices défini au-dessus de $\Delta^{n-1} \times R$ (cf. section 5.1) introduit par Zorich.

Revenons aux champs de vecteurs sur les surfaces de genre $g \geq 2$. Soient M une surface orientable compacte de genre $g \geq 2$ et ω une forme volume sur M . Considérons un champ de vecteurs X (on note $\phi_X(\cdot, t)$ son flot) dont les singularités sont de type selle (Σ, ι) où $\Sigma = \{p_1, \dots, p_\sigma\}$ est l'ensemble des singularités du champ X et $\iota_k, 1 \leq k \leq \sigma$ est l'indice de X au point p_k . La notion de distribution sur M (au sens de Schwartz et De Rham) a bien un sens ; par ailleurs on peut définir de façon habituelle l'espace de Sobolev $H^1(M)$ et son dual $H^{-1}(M)$. Une distribution \mathcal{D} est dite X -invariante si $X\mathcal{D} = 0$ et d'ordre 1 si elle est dans $H^{-1}(M)$. Nous noterons $\mathcal{I}_X^1(M)$ l'ensemble des distributions X -invariantes et d'ordre 1.

Le résultat que démontre Giovanni Forni est le suivant :

THÉORÈME 1.2 (Forni [5]). — *Pour « presque tout » champ de vecteurs X qui préserve ω et dont les singularités sont de type selle il existe des nombres réels strictement positifs $\lambda'_1(X) > \dots > \lambda'_s(X)$ et des espaces de distributions $\mathcal{I}_X^1(\lambda'_i)$, $1 \leq i \leq s$, X -invariantes et d'ordre 1 tels que*

(1) on ait la décomposition

$$\mathcal{I}_X^1 = \mathcal{I}_X^1(\lambda'_1) \oplus \cdots \oplus \mathcal{I}_X^1(\lambda'_s),$$

(2) si $i < s$ et si une observable f à support dans $M - \Sigma$ et dans l'espace de Sobolev $H_0^1(M - \Sigma)$ vérifie

$$\forall \mathcal{D} \in \mathcal{I}_X^1(\lambda'_1) \oplus \cdots \oplus \mathcal{I}_X^1(\lambda'_i), \quad \mathcal{D}(f) = 0$$

alors pour presque tout point $p \in M$

$$\limsup_{T \rightarrow \infty} \frac{\log \left| \int_0^T f(\phi_X(p, \tau)) d\tau \right|}{\log T} \leq \lambda'_{i+1};$$

en outre il existe une distribution $\mathcal{D}_{i+1} \in \mathcal{I}_X(\lambda'_{i+1})$ telle que si $\mathcal{D}_{i+1}(f) \neq 0$ alors l'inégalité précédente est une égalité ;

(3) si $i = s$ et sous les mêmes hypothèses sur f qu'en (2),

$$\limsup_{T \rightarrow \infty} \frac{\log \left| \int_0^T f(\phi_X(p, \tau)) d\tau \right|}{\log T} = 0.$$

Le « presque tout » signifie la chose suivante : à tout champ de vecteurs X préservant la forme volume ω on associe la classe de cohomologie $[i_X \omega] \in H^1(M, \Sigma, \mathbf{R})$ (cf. section 6.1). Un ensemble de champs de vecteurs préservant ω est négligeable si son image par l'application précédente est de mesure de Lebesgue nulle dans $H^1(M, \Sigma, \mathbf{R})$.

La preuve de ce résultat repose sur l'analyse de deux types de dynamiques différents. Il y a, d'une part, la dynamique du champ de vecteurs X sur la surface M que l'on peut ramener à celle du feuilletage horizontal d'une différentielle quadratique q . D'autre part, il existe des dynamiques de « renormalisation » qui ont pour but d'accélérer la dynamique sur M . Les espaces dans lesquels travaillent ces dynamiques sont des espaces de modules (cf. section 2) et pour les comprendre il est important de savoir démontrer que les exposants de Lyapunov de certains cocycles, le cocycle de Kontsevich-Zorich (cf. section 3.2) et le cocycle de Forni (cf. section 6.3 et section 8) ont des exposants de Lyapunov non nuls (cf. section 5 et section 7). Pour cela Forni utilise des outils d'Analyse qu'il a développés dans [4] et que l'on peut voir comme une extension de la théorie de Hodge (section 4).

Remarque : Il existe des liens importants entre le problème des déviations des moyennes ergodiques pour des flots sur des surfaces de genre supérieur à 2 ou des échanges d'intervalles, et celui de la résolution des équations cohomologiques. Dans le cas des flots, par exemple, il s'agit de résoudre l'équation $\mathcal{L}_X f = g$ où g est une fonction (ou une distribution) donnée (\mathcal{L}_X est la dérivation de Lie). Ces équations n'ont pas toujours de solution, mais on connaît dans certains cas les obstructions. Nous ne parlerons pas de ce sujet dans la suite mais nous renvoyons à l'article fondateur de Forni [4] et mentionnons les résultats nouveaux de Marmi-Moussa-Yoccoz [13] dans le cadre des échanges d'intervalles.

Remerciements : Je tiens à remercier Giovanni Forni, qui a répondu très patiemment à mes nombreuses questions, pour son aide, Jean-Paul Thouvenot, Stefano Marmi et Jean-Christophe Yoccoz pour leurs commentaires éclairés.

2. LES ESPACES DE MODULES

2.1. L'espace de Teichmüller

L'espace de Teichmüller T_g d'une surface M de genre $g \geq 2$ est l'ensemble des structures complexes de M à isotopies près, c'est-à-dire l'ensemble obtenu en décrétant que deux structures complexes sont équivalentes si l'on peut passer de l'une à l'autre par un élément de $\text{Diff}_0(M)$, l'ensemble des difféomorphismes de M isotopes à l'identité. C'est aussi l'ensemble des métriques de courbure -1 à isotopies près. On peut munir T_g d'une distance : si $[C_1]$ et $[C_2]$ sont deux classes de structures conformes, $d([C_1], [C_2])$ est l'infimum sur les $h \in \text{Diff}_0(M)$ des $(1/2) \ln K(h)$, $K(h)$ étant la constante de quasi-conformalité de h (relativement à C_1, C_2). Muni de cette distance, T_g est homéomorphe à la boule ouverte d'un espace euclidien réel de dimension $6g - 6$. On peut également munir T_g d'une structure complexe qui en fait un domaine borné d'holomorphicité de \mathbf{C}^{3g-3} .

2.2. Différentielles quadratiques

Une structure complexe étant choisie sur M , une différentielle quadratique holomorphe q sur M est la donnée, pour chaque carte d'un atlas conforme $(U_i, \zeta_i)_i$, d'une expression $q_i(\zeta_i) d\zeta_i^2$ (q_i holomorphe sur U_i) où l'élément différentiel signifie que sur l'ouvert $U_i \cap U_j$ on a la relation de compatibilité

$$q_j(\zeta_j) \left(\frac{d\zeta_j}{d\zeta_i} \right)^2 = q_i(\zeta_i).$$

Le quotient de deux différentielles quadratiques est une fonction méromorphe (qui a donc autant de zéros que de pôles) et par conséquent toutes les différentielles quadratiques ont le même nombre de zéros qui est le double du nombre de zéros d'une 1-forme holomorphe : une différentielle quadratique a donc $4g - 4$ zéros (comptés avec multiplicité). L'espace des différentielles quadratiques sur M est un espace vectoriel de dimension (complexe) finie égale à $3g - 3$ (Riemann-Roch) si $g \geq 2$ et de dimension 1 sur le tore. Chaque différentielle quadratique définit une structure plate sur M avec des singularités (Gauss-Bonnet) aux zéros de q : en un point $x \in M$ où $q(x) \neq 0$ il existe un paramètre local z sur un voisinage U de x où $q = dz^2$ que l'on obtient de la façon suivante : si dans la carte (U, ζ) on a $q = q(\zeta) d\zeta^2$, on pose

$$z(y) = \int_{[x,y]} (q(\zeta))^{1/2} d\zeta.$$

Puisqu'un tel z est invariant modulo des translations et un changement de signe, M a localement une structure plate au voisinage de tout point non singulier. En un point où

q admet un zéro d'ordre k on peut trouver un paramètre local pour lequel $q = z^k dz^2$. On peut ainsi définir grace à q deux objets différentiels importants sur M et un objet de nature plus géométrique (voir pour la partie géométrique la référence qu'est [3]) :

- une métrique plate en dehors des zéros de q que nous noterons R_q : $R_q = (dx^2 + dy^2)^{1/2}$ au voisinage d'un point régulier et $R_q = |z|^{k/2}(dx^2 + dy^2)^{1/2}$ au voisinage d'un zéro d'ordre k ;
- une forme volume $\omega_q = (i/2)h \wedge \bar{h}$ où $h = q^{1/2}$: $\omega_q = dx \wedge dy$ au voisinage d'un point régulier et $\omega_q = |z|^k dx \wedge dy$ au voisinage d'un zéro d'ordre k ;
- une paire de feuilletages mesurés transverses $(\mathcal{F}_q, \mathcal{F}_{-q})$ définis par $\mathcal{F}_q = \{\text{Im}(q^{1/2}) = 0\}$ (le feuilletage horizontal) et $\mathcal{F}_{-q} = \{\text{Re}(q^{1/2}) = 0\}$ (le feuilletage vertical).

Nous noterons Q_g l'ensemble des différentielles quadratiques modulo l'action des difféomorphismes de M isotopes à l'identité. Remarquons que Q_g se projette sur T_g (puisque une différentielle quadratique « porte » sa propre structure complexe). On peut démontrer que Q_g est naturellement l'espace cotangent de T_g (donc de dimension complexe $6g - 6$). En fait nous ne considérerons dans la suite (comme le fait Forni) que des différentielles quadratiques qui sont le carré de différentielles holomorphes (abéliennes) $q = h^2$, $h = h(z)dz$ (ce qui correspond au cas où les feuilletages sont orientables) car ce sont celles qui interviennent en dynamique ; par ailleurs, le cas général s'y ramène. On a alors $h = \alpha + i\beta$ où α, β sont deux 1-formes réelles fermées et transverses (i.e $\alpha \wedge \beta > 0$ sauf en un nombre fini de points). Réciproquement si α, β sont deux 1-formes fermées réelles transverses avec des singularités canoniques (de la forme $\text{Re}(z^k dz) = 0$), il existe une unique structure conforme telle que $\alpha + i\beta$ soit une 1-forme holomorphe (c'est clair en dehors des points où $\alpha \wedge \beta$ s'annule et en ces points on invoque le théorème de Riemann sur les singularités isolées des fonctions holomorphes). Cette remarque permet de faire agir le groupe des difféomorphismes $\text{Diff}^+(M)$ sur l'ensemble des différentielles quadratiques orientables : si $f \in \text{Diff}^+(M)$ et $q = h^2$ avec $h = \alpha + i\beta$, $d\alpha = d\beta = 0$, l'action de f sur q est $f^*q = (f^*\alpha + if^*\beta)^2$.

Afin d'obtenir une paramétrisation plus agréable des points de Q_g , il est nécessaire de se restreindre aux *strates* de Q_g définies de la façon suivante. Fixons $\kappa = (k_1, \dots, k_\sigma)$ un σ -uplet d'entiers positifs pairs dont la somme vaut $4g - 4$ et considérons l'ensemble des différentielles quadratiques qui sont le carré de différentielles holomorphes et dont les zéros distincts ont pour ordre (k_1, \dots, k_σ) . Il est possible de voir que ces propriétés sont invariantes par isotopies et on note Q_κ l'ensemble obtenu en quotientant par $\text{Diff}_0(M)$; c'est un sous-ensemble complexe analytique de Q_g de dimension complexe $2g + \sigma - 1$ et quand $\sigma = 2g - 2$ (on parle alors de la strate principale) de dimension $2g + 2g - 2 - 1 = 4g - 3 = 3g - 3 + g$.

2.3. Les espaces de modules

Notons $\Gamma_g = \text{Diff}^+(M)/\text{Diff}_0(M)$ (le «pure mapping class group») le groupe des difféomorphismes de M préservant l'orientation à isotopies près. Ce groupe agit par pull-back sur T_g, Q_g, Q_κ et on peut définir les espaces de modules de structures complexes et de différentielles quadratiques suivants : $R_g = T_g/\Gamma_g$, $\mathcal{M}_g = Q_g/\Gamma_g$, $\mathcal{M}_\kappa = Q_\kappa/\Gamma_g$. Ce sont des espaces complexes analytiques. Par ailleurs les strates \mathcal{M}_κ de \mathcal{M}_g peuvent être munies d'une structure complexe affine dont le modèle est $H^1(M, \Sigma_k, \mathbf{C})$ au moyen de l'application de périodes. Fixons $a_1, \dots, a_g, b_1, \dots, b_g$ des générateurs de l'homologie $H_1(M, \mathbf{Z})$ et $\gamma_1, \dots, \gamma_{\sigma-1}$ des chemins reliant une singularité fixée de q (un point de Σ_q) aux autres singularités de q (les autres points de Σ_q). Localement, deux 1-formes holomorphes $q_1^{1/2}$ et $q_2^{1/2}$ définissent le même point de $H^1(M, \Sigma_\kappa, \mathbf{C})$ si et seulement si elles sont isotopes, ce qui identifie localement Q_κ à $H^1(M, \Sigma_\kappa, \mathbf{C})$. L'intégration de la 1-forme holomorphe $q^{1/2}$ le long des a_i, b_i, γ_i définit un point $(x_1, \dots, x_g, y_1, \dots, y_g, t_1, \dots, t_{\sigma-1})$ de $\mathbf{C}^{2g+\sigma-1}$ et on peut vérifier que les changements de cartes (obtenus par changement des générateurs (a_i, b_i, γ_i)) sont holomorphes, ce qui munit Q_κ d'une structure affine complexe. Ces changements de cartes préservent par ailleurs le volume, ce qui permet de tirer en arrière la forme $dx_1 \wedge d\bar{x}_1 \wedge \dots \wedge dx_g \wedge d\bar{x}_g \wedge dy_1 \wedge d\bar{y}_1 \wedge \dots \wedge dy_g \wedge d\bar{y}_g \wedge dt_1 \wedge d\bar{t}_1 \wedge \dots \wedge dt_{\sigma-1} \wedge d\bar{t}_{\sigma-1}$ et de définir une mesure absolument continue μ_κ sur Q_κ (normalisée en demandant que la mesure de Lebesgue du tore complexe obtenu en quotientant par le réseau entier $\mathbf{C} \otimes H^1(M_q, \Sigma_q, \mathbf{Z})$ soit égale à 1). Cette mesure se projette sur \mathcal{M}_κ . Notons qu'à ce niveau il n'est pas clair que la mesure μ_κ soit finie sur \mathcal{M}_κ . D'autre part, comme cela a été mis en évidence par P. Arnoux et W.A. Veech ces espaces de modules ne sont pas toujours connexes et on devra parfois se restreindre dans la suite à leurs composantes connexes.

Pour terminer la description de ces espaces \mathcal{M}_κ , notons qu'il existe

- une fonction (la fonction d'aire) $A : \mathcal{M}_\kappa \rightarrow [0, \infty)$ définie par

$$A(q) = \int_M \omega_q$$

(c'est bien invariant par l'action de $\text{Diff}^+(M)$)

- une action du groupe $GL_+(2, \mathbf{R})$ sur \mathcal{M}_κ définie par

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} \text{Re}(q^{1/2}) \\ \text{Im}(q^{1/2}) \end{pmatrix} = \begin{pmatrix} a\text{Re}(q^{1/2}) + b\text{Im}(q^{1/2}) \\ c\text{Re}(q^{1/2}) + d\text{Im}(q^{1/2}) \end{pmatrix}.$$

Cette action sur Q_κ commute à celle de Γ_g .

3. DYNAMIQUES SUR LES ESPACES DE MODULES

3.1. Le flot de Teichmüller

Par définition c'est l'action du sous-groupe \mathbf{A} de $SL(2, \mathbf{R})$ constitué des matrices diagonales $G_t = \text{diag}(e^t, e^{-t})$ sur l'espace Q_g . Ce flot s'identifie au flot géodésique sur

l'espace de Teichmüller : si $\pi(q)$ désigne la structure complexe définie par q , on a $d(\pi(q), \pi(G_t(q))) = t$. Par ailleurs, ce flot laisse invariante chacune des strates Q_κ et la fonction d'aire A . Par conséquent les ensembles $Q_\kappa^{(1)} = Q_\kappa \cap A^{-1}(1)$ sont également invariants par la dynamique du flot de Teichmüller. Puisque l'action de $GL_+(2, \mathbf{R})$ commute à celle du groupe modulaire Γ_g , on peut définir le flot de Teichmüller sur chacun des espaces de modules, $\mathcal{M}_g, \mathcal{M}_\kappa, \mathcal{M}_g^{(1)} = \mathcal{M}_g \cap A^{-1}(1), \mathcal{M}_\kappa^{(1)} = \mathcal{M}_\kappa \cap A^{-1}(1)$. On notera $\mu_\kappa^{(1)}$ la mesure induite par μ_κ sur les ensembles $\mathcal{M}_g^{(1)}, \mathcal{M}_\kappa^{(1)}$.

THÉORÈME 3.1 (Masur [11], Veech [16]). — *Pour chaque strate $\mathcal{M}_\kappa^{(1)}$:*

- *Le volume total de $\mathcal{M}_\kappa^{(1)}$ pour la mesure $\mu_\kappa^{(1)}$ est fini.*
- *Le flot de Teichmüller G_t est ergodique sur chaque composante connexe de $\mathcal{M}_\kappa^{(1)}$ pour la mesure invariante $\mu_\kappa^{(1)}$ et est non uniformément hyperbolique (Veech).*

3.2. Le cocycle de Kontsevich et Zorich

3.2.1. *Première description.* Soit $\mathcal{W}_\kappa \subset \mathcal{M}_\kappa^{(1)}$ une petite section transverse au flot de Teichmüller et soit $\tilde{\mathcal{W}}_\kappa$ un représentant de \mathcal{W}_κ dans $Q_\kappa^{(1)}$ (l'action de Γ_g est discrète et propre sur $Q_\kappa^{(1)}$ et ses points fixes constituent un ensemble de $\mu_\kappa^{(1)}$ -mesure nulle). Pour $\mu_\kappa^{(1)}$ -presque tout point q de $\mathcal{M}_\kappa^{(1)}$ et tout $t \in \mathbf{R}_+$ (resp. $t \in \mathbf{R}_-$) notons $0 \leq t_0 < \dots < t_r \leq t$ (resp. $t \leq t_r < \dots < t_0 \leq 0$) les temps tels que $G_{t_i}(q) \in \mathcal{W}_\kappa$. Soit $\tilde{q} \in Q_\kappa$ un représentant de $q \in \mathcal{M}_\kappa^{(1)}$ tel que $G_{t_0}(\tilde{q}) \in \tilde{\mathcal{W}}_\kappa$ (l'action de Γ_g commute à celle de G_t). Pour chaque $0 \leq i \leq r-1$, il existe un unique $f_i \in \Gamma_g$ tel que $G_{t_{i+1}}(\tilde{q}) = f_i^* \cdot G_{t_i}(\tilde{q})$. L'action de $f_{r-1}^* \circ \dots \circ f_0^*$ sur $H^1(M, \mathbf{R})$ est symplectique et on l'identifie à une matrice de $Sp(2g, \mathbf{R})$. On pose $G_t^{KZ}(q) = f_{r-1}^* \circ \dots \circ f_0^*$. On a alors la relation de cocycle

$$G_{t+s}^{KZ}(q) = G_t^{KZ}(G_s(q))G_s^{KZ}(q).$$

On a ainsi une action de \mathbf{R} sur $\mathcal{M}_\kappa^{(1)} \times H^1(M, \mathbf{R})$: pour $(q, v) \in \mathcal{M}_\kappa^{(1)} \times H^1(M, \mathbf{R})$

$$t.(q, v) = (G_t(q), G_t^{KZ}(q).v).$$

Le cocycle $G^{KZ} : \mathbf{R} \times \mathcal{M}_\kappa^{(1)} \rightarrow Sp(2g, \mathbf{R})$ est (une version) du cocycle de Kontsevich-Zorich. Kontsevich et Zorich donnent une définition intrinsèque de ce cocycle en utilisant le transport parallèle que fournit la connexion de Gauss-Manin (cf. [10] ou [5]).

Si l'on choisit une norme sur $H^1(M, \mathbf{R})$, il n'est pas difficile de voir que pour tout $t \in \mathbf{R}$

$$\int_{\mathcal{M}_\kappa^{(1)}} \log \|G_t^{KZ}(q)\| d\mu_\kappa(q) < \infty,$$

si bien que l'on peut appliquer le théorème d'Oseledec (cf. section 5.1). La conjecture de Kontsevich et Zorich ([9], [10]) est la suivante :

CONJECTURE 3.2 (Kontsevich-Zorich). — *Les exposants de Lyapunov du cocycle de Kontsevich-Zorich sont tous non nuls et simples.*

Kontsevich et Zorich esquissent dans leur article [10] la démonstration du théorème suivant :

THÉORÈME 3.3 (Kontsevich-Zorich). — 1) *Le plus grand exposant de Lyapunov est 1 et est simple (ce qui découle de la non hyperbolicité uniforme du flot géodésique prouvée par Veech) ; l'espace instable associé est $[\operatorname{Re}(q^{1/2})]$.*

2) *Le deuxième exposant de Lyapunov gouverne le terme d'erreur dans le théorème ergodique pour les échanges d'intervalles.*

3) *Les espaces instables au-dessus du point q ne dépendent que de la classe de cohomologie de $\operatorname{Im}(q^{1/2})$ dans $H^1(M, \Sigma, \mathbf{R})$.*

Ils fournissent également une formule donnant la somme des exposants de Lyapunov positifs et conjecturent que cette somme est toujours un nombre rationnel.

3.2.2. Seconde description. Au lieu de considérer le cocycle $G_t^{KZ}(q)$ défini sur $\mathbf{R} \times \mathcal{M}_\kappa^{(1)}$ et de fixer une norme sur $H^1(M, \mathbf{R})$, Forni procède de la façon suivante. Il considère le produit trivial $Q_\kappa^{(1)} \times H^1(M, \mathbf{R})$ et considère le flot \tilde{G}_t sur ce produit qui agit par le flot de Teichmüller sur Q_κ et de façon triviale sur $H^1(M, \mathbf{R})$, i.e. $\tilde{G}_t(q, v) = (G_t(q), v)$. En revanche, au lieu de fixer une norme qui est la même sur tous les $H^1(M, \mathbf{R})$, au-dessus de chaque $q \in Q_\kappa^{(1)}$, il munit $H^1(M, \mathbf{R})$ de la norme de Hodge

$$(\alpha, \beta)_q = \int_M \alpha \wedge * \beta,$$

où l'opérateur $*$ est défini par la structure complexe induite par q (i.e dans une carte conforme $*dx = dy$, $*dy = -dx$, $x + iy$ étant le paramètre local). Une propriété évidente mais importante de cette norme est qu'elle est Γ_g -équivariante, c'est-à-dire $(\alpha, \beta)_q = (f^*\alpha, f^*\beta)_{f^*q}$ pour tout difféomorphisme f qui préserve l'orientation ; par conséquent le spectre de Lyapunov du cocycle $G_t^{KZ}(x)$ étudié précédemment et celui du cocycle que définissent \tilde{G} et les normes $(\cdot, \cdot)_q$ sont les mêmes (cf. section 5.1). Nous noterons à nouveau G_t^{KZ} le cocycle \tilde{G}_t .

Forni démontre une partie de la conjecture de Kontsevich et Zorich :

THÉORÈME 3.4 (Forni [5]). — *Les exposants du cocycle de Kontsevich-Zorich sont tous non nuls et le plus grand est simple (la simplicité du plus grand exposant étant vraie pour n'importe quelle mesure invariante ergodique par le flot de Teichmüller).*

3.3. Un exemple

Nous illustrons sur un exemple simple les notions précédentes. Supposons que $f : M \rightarrow M$ soit un difféomorphisme pseudo-Anosov dont les feuilletages invariants sont orientables, c'est-à-dire qu'il existe deux feuilletages mesurés orientables transverses \mathcal{F} et \mathcal{F}^\perp et un réel non nul Λ tels que

$$f_*(\mathcal{F}) = \Lambda \mathcal{F}, \quad f_*(\mathcal{F}^\perp) = \Lambda^{-1} \mathcal{F}^\perp.$$

De façon équivalente, on peut supposer qu'il existe une différentielle quadratique orientable q telle que

$$f^*(\operatorname{Re}(q^{1/2})) = \Lambda \operatorname{Re}(q^{1/2}), \quad f^*(\operatorname{Im}(q^{1/2})) = \Lambda^{-1} \operatorname{Im}(q^{1/2}).$$

Les équations précédentes admettent des interprétations distinctes dans des espaces différents :

- sur M : les classes de $\operatorname{Re}(q^{1/2})$, $\operatorname{Im}(q^{1/2})$ dans $H^1(M, \mathbf{R})$ sont valeurs propres de l'application linéaire symplectique $f^* : H^1(M, \mathbf{R}) \rightarrow H^1(M, \mathbf{R})$ (qu'on identifie à un élément de $Sp(2g, \mathbf{R})$) ;
- sur \mathcal{M}_κ : $[q] \in \mathcal{M}_\kappa$ est un point périodique de période $T = \log |\Lambda|$ du flot de Teichmüller G_t .

Par définition du flot de Kontsevich-Zorich on a

$$G_T^{KZ} = f^*,$$

et donc les valeurs propres de f^* , $\Lambda_i^{\pm 1}$ $1 \leq i \leq g$ ($|\Lambda_1| \geq \dots \geq |\Lambda_g|$) et les exposants de Lyapunov $\lambda_1 \geq \dots \geq \lambda_g \geq -\lambda_g \geq \dots \geq -\lambda_1$ du cocycle de Kontsevich-Zorich sont reliés par

$$\lambda_i = \frac{\log |\Lambda_i|}{\log |\Lambda|}.$$

On peut démontrer que $1 = \lambda_1 > \lambda_2$ (trou spectral, cf. section 4.3) et donc $\Lambda = \Lambda_1$ et $|\Lambda_1| > |\Lambda_2|$.

On va s'intéresser à présent non pas à f mais à la dynamique du feuilletage horizontal (ou vertical) ; ces feuilletages sont minimaux (toute feuille ne passant pas par les singularités de q est dense). On note (cf. section 4) S et T les champs de vecteurs associés aux feuilletages horizontal et vertical de q et ϕ_q^t le flot du champ de vecteurs horizontal S . Supposons en outre que μ soit une mesure invariante par S et même ergodique. Il n'est alors pas difficile de voir en appliquant le théorème ergodique (voir section 6.1 pour plus de détails) que si on note $\gamma_t(p)$ le chemin $\phi_q^s(p)$, $0 \leq s \leq t$, on a pour μ -presque tout p et toute 1-forme fermée α

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{\gamma_t(p)} \alpha = \rho_\mu(\alpha),$$

où $\rho_\mu \in H_1(M, \mathbf{R})$ est le vecteur de rotation de la mesure μ (défini section 6.1). Remarquer que le facteur $(1/t)$ est l'inverse de la q -longueur $L_q(\gamma_t)$ du chemin γ_t . On a toujours

$$\frac{1}{t} \int_{f_* \gamma_t(p)} f^* \alpha = \frac{1}{t} \int_{\gamma_t(p)} \alpha$$

et, comme $L_q(f(\gamma_t)) = \Lambda L_q(\gamma_t)$, on a

$$\Lambda^{-1} \langle \rho_\mu, f^* \alpha \rangle = \langle \rho_\mu, \alpha \rangle$$

et donc

$$f_* \rho_\mu = \Lambda \rho_\mu.$$

Comme Λ est simple, toutes les mesures ergodiques ont donc le même vecteur de rotation. Mais comme l'application $\mu \mapsto \rho_\mu$ est injective (cf. section 6.1), il n'y a qu'une seule mesure ergodique : le champ de vecteurs S (le feuilletage horizontal) est uniquement ergodique.

Si γ_t est une q -trajectoire horizontale de q -longueur t , on peut la fermer pour certaines valeurs de t par des chemins verticaux de q -longueurs petites. Notons $\hat{\gamma}_t$ un tel chemin et étudions $f^n(\hat{\gamma}_t)$; c'est un cycle de q -longueur $L_n \approx \Lambda^n L_q(\hat{\gamma}_t)$ et si on projette $[f^n(\hat{\gamma}_t)]$ sur l'espace propre de f_* de valeur propre Λ_2 (on note π_2 la projection), on obtient un cycle de q -longueur à peu près égale à $|\Lambda_2|^n = L_n^{\lambda_2}$. Ainsi,

$$\limsup_{n \rightarrow \infty} \frac{\log |\pi_2([f^n(\hat{\gamma}_t)])|}{\log L_q(f^n(\hat{\gamma}_t))} \leq \lambda_2.$$

Comme $f^*q = (\Lambda, \Lambda^{-1}).q$, $f^n(\hat{\gamma}_t)$ est une trajectoire q -horizontale fermée par un petit chemin q -vertical ; ceci suggère que

$$\limsup_{t \rightarrow \infty} \frac{\log |\pi_2([\gamma_t(x)])|}{\log t} \leq \lambda_2.$$

pour μ -presque tout $x \in M$.

Si $F : M \rightarrow \mathbf{R}$ est une observable, ses moyennes ergodiques peuvent s'écrire

$$\frac{1}{t} \int_M F(\phi_S^t(x)) d\mu(x) = \frac{1}{t} \int_{\gamma_t} \alpha,$$

où α est la 1-forme $F\eta_T$ (pas forcément fermée) et on peut espérer obtenir des déviations de moyennes ergodiques si on projette α , donc F , sur des espaces convenables.

4. L'ANALYSE HARMONIQUE SUR M

Les résultats de cette partie sont démontrés dans [4].

Soit q un élément de $\mathcal{M}_k^{(1)}$. Notons η_{ver} et η_{hor} les 1-formes fermées

$$\eta_{\text{ver}} = \operatorname{Re}(q^{1/2}), \quad \eta_{\text{hor}} = \operatorname{Im}(q^{1/2}),$$

si bien que $\omega_q = \eta_{\text{ver}} \wedge \eta_{\text{hor}}$. Il existe des champs de vecteurs S, T qui commutent, définis sur $M - \Sigma_q$ tels que (i_X désignant la contraction avec le champ de vecteurs X)

$$\eta_{\text{ver}} = \eta_T = -i_T \omega_q, \quad \eta_{\text{hor}} = \eta_S = i_S \omega_q.$$

Ces champs sont C^∞ sur $M - \Sigma_q$, préservent la forme d'aire ω_q et explosent sur Σ_q . Au voisinage d'une singularité de q d'ordre $k = 2m$, ils s'écrivent ($z = x + iy$)

$$S = |z|^{-2m} \left(\operatorname{Re}(z^m) \frac{\partial}{\partial x} - \operatorname{Im}(z^m) \frac{\partial}{\partial y} \right),$$

$$T = |z|^{-2m} \left(\operatorname{Im}(z^m) \frac{\partial}{\partial x} + \operatorname{Re}(z^m) \frac{\partial}{\partial y} \right).$$

4.1. Espaces fonctionnels

Il faut distinguer plusieurs espaces fonctionnels. Tout d'abord, si ω (à ne pas confondre avec ω_q) est une forme volume non dégénérée, on peut définir l'espace $L^2(M)$ des fonctions telles que $\|u\|_0 < \infty$ pour la norme

$$\|u\|_0 = \left(\int_M |u|^2 \omega \right)^{1/2},$$

et les espaces de Sobolev standards $H^s(M)$ définis en cartes locales.

Par ailleurs, la 2-forme dégénérée ω_q définit également un espace $L_q^2(M)$ de fonctions L_q^2 -intégrable pour la norme

$$|u|_0 = \left(\int_M |u|^2 \omega_q \right)^{1/2}.$$

L'espace de Sobolev $H_q^1(M)$ est alors le complété de l'espace des fonctions $C^\infty(M)$ pour le produit scalaire

$$|u|_1 = \left(|u|_0^2 + |Su|_0^2 + |Tu|_0^2 \right)^{1/2}$$

et plus généralement les espaces $H_q^s(M)$ sont définis comme les complétés de $C^\infty(M)$ pour la norme

$$|u|_s = \left(\sum_{i+j \leq s} |S^i T^j u|_0^2 \right)^{1/2}.$$

Les espaces $H_q^s(M)$ sont naturellement des espaces de Hilbert et les opérateurs S, T vérifient pour $u, v \in H_q^1(M)$

$$(Su, v)_q = -(u, Sv)_q, \quad (Tu, v)_q = -(u, Tv)_q,$$

et la relation de commutation plus forte que $ST = TS$

$$(Su, Tv)_q = (Tu, Sv)_q,$$

ce qui se vérifie en utilisant les relations

$$(Su)\omega_q = du \wedge \eta_S = d(u\eta_S), \quad (Tu)\omega_q = -du \wedge \eta_T = -d(u\eta_T)$$

$$dv = (Tv)\eta_S + (Sv)\eta_T,$$

et la formule de Green.

Les relations entre les espaces de Sobolev standards et les $H_q^s(M)$ sont les suivantes :

- $L^2(M) \subset L_q^2(M)$
- $H^1(M) = H_q^1(M)$
- $H_q^s(M) \subset H^s(M)$ pour $s \geq 2$.

Les deuxième et troisième points se démontrent grâce à une inégalité de type Poincaré que Forni démontre dans ([4]). Remarquons que l'espace $C_0^\infty(M - \Sigma)$ n'est *pas* dense dans $H_q^s(M)$ pour $s \geq 2$.

Notons $L_q^2(M) = L^2(M, \omega_q)$ l'espace de Hilbert des fonctions de carré intégrable par rapport à la forme d'aire ; ces espaces de Hilbert *sont invariants sous l'action du flot de Teichmüller*, c'est-à-dire que $|u|_q = |u|_{G_t(q)}$ (attention à ne pas confondre le produit scalaire sur $L^2(M, \omega_q)$ et celui de Hodge pour les 1-formes). Notons ∂_q^\pm les opérateurs de Cauchy-Riemann

$$\partial_q^\pm = S \pm iT.$$

L'analyse harmonique sur M développée par Forni dans [4] peut en partie se résumer dans le théorème suivant

THÉORÈME 4.1 (Forni). — *Soit q une différentielle quadratique qui est le carré d'une forme abélienne :*

- 1) *Les opérateurs ∂_q^\pm sont fermés sur le domaine $H^1(M) \subset L_q^2(M)$ constitué des fonctions L_q^2 dont les dérivées au sens des distributions sont L_q^2 .*
- 2) *Les opérateurs ∂_q^\pm ont des images R_q^\pm fermées et de codimension finie.*
- 3) *$(\partial_q^\pm)^* = -\partial_q^\mp$.*
- 4) *On a les décompositions orthogonales suivantes :*

$$L_q^2(M) = R_q^- \oplus \mathcal{M}_q^+ = R_q^+ \oplus \mathcal{M}_q^-$$

(on notera π_q^\pm les projections orthogonales sur \mathcal{M}_q^\pm associées).

- 5) *Les espaces de fonctions méromorphes (antiméromorphes) \mathcal{M}_q^+ [\mathcal{M}_q^-] sont isomorphes à l'espace $H^1(M_q, \mathbf{R})$ muni de la norme de Hodge déterminée par q .*

4.2. Les équations d'évolution

Le dernier point du théorème permet de *paramétrer* $H^1(M, \mathbf{R})$ muni de la norme de Hodge $\|\cdot\|_q$ par les fonctions q -méromorphes ou q -antiméromorphes de la façon suivante : les applications \mathbf{R} -linéaires $c_q^\pm : \mathcal{M}_q^\pm \rightarrow H^1(M_q, \mathbf{R})$ définies par

$$c_q^+(m^+) = [\operatorname{Re}(m^+ q^{1/2})]$$

$$c_q^-(m^-) = [\operatorname{Re}(m^- \bar{q}^{1/2})]$$

sont des isomorphismes d'espaces vectoriels. En effet, d'après la théorie de Hodge, toute 1-forme fermée réelle est cohomologue à la partie réelle d'une forme abélienne h (une 1-forme harmonique). Or $h = (h/q^{1/2})q^{1/2}$, où $h/q^{1/2}$ est une fonction q -méromorphe L_q^2 -intégrable. On a en outre les relations

$$\|c_q^\pm(m^\pm)\|_q^2 = \int_M c_q^\pm(m^\pm) \wedge {}^*c_q^\pm(m^\pm) = |m^\pm|_0^2,$$

et pour m_1^\pm, m_2^\pm dans \mathcal{M}_q^\pm

$$c_q^\pm(m_1^\pm) \wedge c_q^\pm(m_2^\pm) = \text{Im}((m_1^\pm, m_2^\pm)_q).$$

La décomposition du théorème 4.1, (4.1)

$$u = \partial_q^+ v + \pi_q^-(u), \quad v \in H^1(M)$$

permet de définir un opérateur unitaire U_q sur L_q^2 par

$$U_q : u = \partial_q^+ v + \pi_q^-(u) \mapsto \partial_q^- v - \overline{\pi_q^-(u)}$$

($v \in H^1(M)$). Forni peut alors décrire l'action du cocycle de Kontsevich-Zorich. Rappelons que si $q_t = G_t(q)$

$$\eta_T(t) = \text{Re}(q_t^{1/2}) = e^t \text{Re}(q^{1/2}) = e^t \eta_T$$

$$\eta_S(t) = \text{Im}(q_t^{1/2}) = e^{-t} \text{Im}(q^{1/2}) = e^{-t} \eta_S.$$

On a alors

$$\partial_t^\pm = S_t \pm iT_t = e^{-t} S \pm ie^t T.$$

Une des clés dans l'approche de Forni est la remarque suivante : il existe une correspondance entre les solutions de l'équation différentielle

$$m'(t) = U_{q_t}(m(t))$$

et le flot de Kontsevich-Zorich, en particulier

PROPOSITION 4.2. — *Avec les notations précédentes $m^+(t)$ est solution de l'équation différentielle*

$$(1) \quad m'(t) = U_{q_t}(m(t))$$

avec condition initiale $m^+ \in \mathcal{M}_q^+$ (et alors $m^+(t)$ est q_t -méromorphe) si et seulement si

$$(2) \quad G_t^{KZ}(c_q^+(m^+)) = c_{q_t}^+(m_t^+).$$

4.3. Exemple d'application : le trou spectral

Forni utilise ces équations d'évolution pour démontrer le trou spectral $\lambda_1 > \lambda_2$ (la plus grande valeur propre est simple) pour toute mesure μ invariante (et ergodique) par le flot géodésique (pas seulement $\mu_\kappa^{(1)}$).

Par définition du cocycle de Kontsevich-Zorich, les espaces $E_1(q) = \mathbf{R}[\text{Im}(q^{1/2})]$, $E_{-1}(q) = \mathbf{R}[\text{Re}(q^{1/2})]$ sont invariants et ont des exposants de Lyapunov 1 et -1 respectivement. Définissons, pour $q \in \mathcal{M}_\kappa$, $E_0(q)$ de la façon suivante :

$$E_0(q) = \{c \in H^1(M, \mathbf{R}), \quad c \wedge [\text{Re}(q^{1/2})] = c \wedge [\text{Im}(q^{1/2})] = 0\}.$$

Une classe c appartient à $E_0(q)$ si et seulement si c peut s'écrire $c = \text{Re}(m^+ q^{1/2})$ où m^+ est méromorphe (dans $L_q^2(M)$) et d'intégrale nulle, $\int_M m^+ \omega_q = 0$. Les fibrés $E_0(\cdot)$ sont

invariants par le cocycle de Kontsevich-Zorich. Si $q_t = G_t(q)$, $c_t = G_t^{KZ}(c)$ avec $c \in E_0(q)$ et q μ -générique, on peut démontrer en utilisant les équations d'évolution que

$$\frac{d}{dt} \log |m_t^+|_0 = -\frac{\operatorname{Re}((m_t^+, \overline{m_t^+})_q)}{|m_t^+|_0^2}.$$

En intégrant par rapport au temps il vient

$$\frac{1}{t} \log \frac{|m_t^+|_0}{|m^+|_0} = \frac{1}{t} \int_0^t -\frac{\operatorname{Re}((m_t^+, \overline{m_t^+})_q)}{|m_t^+|_0^2} dt,$$

et si on applique le théorème d'Oseledec (cf. 5.1) on a

$$\lambda_2 \leq \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t -\frac{\operatorname{Re}((m_t^+, \overline{m_t^+})_q)}{|m_t^+|_0^2} dt.$$

Il est tentant d'utiliser pour le membre de droite le théorème de Birkhoff mais il faut trouver une fonction qui ne dépende que de q . À cet effet on introduit la quantité

$$\Lambda(q) = \max \left\{ -\frac{\operatorname{Re}((m^+, \overline{m^+})_q)}{|m^+|_0^2}, \quad m^+ \in \mathcal{M}_q^+ - \{0\}, \quad \int_M m^+ \omega_q = 0 \right\}$$

et on obtient

$$\lambda_2 \leq \int_M \Lambda(q) d\mu(q).$$

Mais pour tout q , $\Lambda(q) < 1$ car sinon, d'après Cauchy-Schwarz, on aurait $\overline{m^+} = \rho m^+$, $|\rho| = 1$, ce qui voudrait dire que m^+ est méromorphe et anti-méromorphe donc constante donc nulle (cf. la condition d'intégrale nulle). Un argument de compacité donne la conclusion.

5. FORMULES POUR LES SOMMES DES EXPOSANTS DE LYAPUNOV

5.1. Rappels sur le théorème d'Oseledec

Considérons un espace X muni d'une mesure de probabilité μ et d'un flot $(\phi^t)_{t \in \mathbf{R}}$ (c'est-à-dire d'une famille à un paramètre d'homéomorphismes de X tels que $\phi^{t+s} = \phi^t \circ \phi^s$ pour $t, s \in \mathbf{R}$). Soit par ailleurs $S : \mathbf{R} \times X \rightarrow Sp(2g, \mathbf{R})$ une application μ -intégrable vérifiant la relation de cocycle $S^{t+s}(x) = S^t(\phi^s(x))S^s(x)$ (on notera parfois dans la suite $S_t(x) = S^t(x) = S(t, x)$). On a ainsi une action de \mathbf{R} sur $X \times \mathbf{R}^{2g}$:

$$t.(x, v) = (\phi^t(x), S^t(x)v).$$

On peut aussi supposer que pour μ -presque tout x il existe une norme $\|\cdot\|_x$ sur \mathbf{R}^{2g} dépendant mesurablement de x et on note $\|S^t(x)\|_{t,x}$ la norme de l'application linéaire $S^t(x) : (\mathbf{R}^{2g}, \|\cdot\|_x) \rightarrow (\mathbf{R}^{2g}, \|\cdot\|_{\phi^t(x)})$. Le théorème d'Oseledec (que nous énonçons dans le cas ergodique et symplectique) dit en partie la chose suivante

THÉORÈME 5.1. — Si le flot $(\phi^t)_{t \in \mathbf{R}}$ est μ -ergodique et si pour tout $t \in \mathbf{R}$

$$\int_X \log \|S^t(x) d\mu(x)\|_{t,x} < \infty,$$

alors pour μ -presque tout x on peut trouver

- un entier $0 \leq h \leq 2g$
- des nombres réels $\lambda'_h < \dots < \lambda'_1$ tels que $\lambda'_i + \lambda'_j = 0$ si $i + j = h + 1$
- des espaces $E_1(x), \dots, E_h(x)$ tels que

$$\mathbf{R}^{2g} = \bigoplus_{i=1}^h E_i(x)$$

qui dépendent mesurablement de x et sont invariants par la dynamique (i.e. $\phi^t(E_i(x)) = E_i(\phi^t(x))$)

tels que pour tout vecteur $v \in E_i(x)$ non nul

$$\lim_{t \rightarrow \pm\infty} \frac{1}{|t|} \log \|S^t(x)v\|_x = \pm\lambda'_i.$$

Par ailleurs, les espaces $E_1 \oplus \dots \oplus E_i$ $1 \leq i \leq [h/2]$ sont isotropes (et pour $i = h/2$ lagrangien).

La donnée des λ'_i et des dimensions des espaces E_i ($1 \leq i \leq h$) s'appelle le spectre de Lyapunov du cocycle ; on dit que l'exposant λ'_i est simple si $\dim E_i$ égale 1 et que le spectre est simple si tous les exposants sont simples. La somme directe des $E_i(x)$ sur les $1 \leq i \leq h$ tels que $\lambda'_i > 0$ (resp. $\lambda'_i < 0$) s'appelle l'espace instable (resp. stable) en x .

Une astuce classique pour avoir accès à la somme des exposants $\lambda_1 + \dots + \lambda_k$ (comptés avec multiplicités) est la suivante. On considère l'espace vectoriel $\Lambda^k E$ des k -formes alternées sur E^* qu'on peut munir du produit scalaire suivant : pour deux k -vecteurs décomposables $v = v_1 \wedge \dots \wedge v_k$ et $w = w_1 \wedge \dots \wedge w_k$

$$(v_1 \wedge \dots \wedge v_k, w_1 \wedge \dots \wedge w_k) = \det((v_i, w_j)_{ij}).$$

Étant donnée T une application linéaire sur E on définit sur $\Lambda^k E$ l'application linéaire $T^{\wedge k}$ par son action sur les k -vecteurs décomposables

$$T^{\wedge k}(v_1 \wedge \dots \wedge v_k) = (Tv_1) \wedge \dots \wedge (Tv_k).$$

Si $v = v_1 \wedge \dots \wedge v_k$ et $\tilde{v} = \tilde{v}_1 \wedge \dots \wedge \tilde{v}_k$ ($\tilde{v}_i = Tv_i$),

$$\frac{\|T^{\wedge k}v\|}{\|v\|} = \frac{\|\tilde{v}_1 \wedge \dots \wedge \tilde{v}_k\|}{\|v_1 \wedge \dots \wedge v_k\|} = \left(\frac{\det((\tilde{v}_i, \tilde{v}_j)_{ij})}{\det((v_i, v_j)_{ij})} \right)^{1/2}$$

et cette dernière quantité ne dépend que de l'espace I engendré par les v_1, \dots, v_k . On la notera $\det A_T^k(I)$. Par ailleurs, comme la somme des k plus grands exposants $\lambda_1 + \dots + \lambda_k$ est égale à

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_X \log \|S_t^{\wedge k}(x)\| d\mu(x)$$

on a le lemme suivant :

LEMME 5.2. — Soit G_k la grassmannienne des k -plans isotropes I_k de \mathbf{R}^{2g} et σ_k la mesure canonique sur G_k . La somme $\lambda_1 + \dots + \lambda_k$ égale

$$\lambda_1 + \dots + \lambda_k = \lim_{T \rightarrow \infty} \frac{1}{T} \int_X \int_{G_k} \log |\det A_{S^t(x)}^k(I)|^{1/2} d\sigma^k(I) d\mu(x).$$

5.2. Disques de Teichmüller et laplacien hyperbolique

L'orbite d'une différentielle quadratique q sous l'action du groupe $SL(2, \mathbf{R})$ est métriquement isométrique au fibré unitaire cotangent du disque de Poincaré (muni de sa métrique de courbure -4). Le quotient $Q_g^{(1)}/SO(2, \mathbf{R})$ est feuilleté par des feuilles de dimension réelle 2 isométriques au disque de Poincaré qu'on appelle disques de Teichmüller. De façon plus explicite, soit $z = re^{i\theta}$, $|z| < 1$ un point du disque de Poincaré. Notons

$$q_z = G_t(q_\theta), \quad q_\theta = e^{i\theta} q,$$

où $t = (1/2) \log \frac{1+r}{1-r}$. On pose $j_q(z) = [q_z] \in Q_g^{(1)}/SO(2, \mathbf{R})$. Observons que $j_q(0) = [q]$. On peut projeter ce plongement sur $\mathcal{M}_g^{(1)}$ et on obtient un feuilletage \mathcal{T}_g de $\mathcal{M}_g^{(1)}/SO(2, \mathbf{R})$ qui est régulier pour presque toute feuille (cf. [10]).

Sur le disque hyperbolique muni de sa métrique $|dz|/(1 - |z|^2)$ (de courbure -4) le laplacien hyperbolique en coordonnées polaires hyperboliques (t, θ) s'écrit

$$\Delta_h = \frac{\partial^2}{\partial t^2} + 2\coth(2t) \frac{\partial}{\partial t} + \frac{4}{\sinh^2(2t)} \frac{\partial^2}{\partial \theta^2}.$$

Le lemme suivant est très utile pour la suite :

LEMME 5.3. — Soit $\Lambda : D \rightarrow \mathbf{R}$ une fonction bornée et C^∞ et $L : D \rightarrow \mathbf{R}$ une solution C^∞ de

$$\Delta_h L = \Lambda.$$

La relation suivante est alors vraie :

$$\frac{1}{2\pi} \frac{\partial}{\partial t} \int_0^{2\pi} L(t, \theta) d\theta = \frac{1}{2} \tanh(t) \frac{1}{|D_t|} \int_{D_t} \Lambda \omega_P,$$

où $|D_t|$ représente l'aire du disque de centre 0 et de rayon (hyperbolique) t et ω_P est la forme d'aire.

5.3. Utilisation des équations d'évolution

Reprenons les notations de la section 5.1 dans le cadre du cocycle de Kontsevich-Zorich. Notons $G_k(M, \mathbf{R})$ la grassmannienne des espaces isotropes de dimension k ($1 \leq k \leq g$) de l'espace symplectique $H^1(M, \mathbf{R})$ (pour la forme d'intersection) et soit $I_k \in G_k(M, \mathbf{R})$ un k -plan isotrope dont on a choisi une base (c_1, \dots, c_k) . Si q est une différentielle quadratique on peut définir (cf. section 4) des fonctions méromorphes m_1^+, \dots, m_k^+ telles que

$$c_i = \operatorname{Re}(m_i^+ q^{1/2}),$$

et introduire des fonctions $v_1, \dots, v_k \in H^1(M)$ telles que

$$m_i^+ = \partial_q^+ v_i + \pi_q^-(m_i^+),$$

où $\pi_q^- : L_q^2(M) \rightarrow \mathcal{M}_q^-$ est l'opérateur de projection orthogonal sur les fonctions anti-méromorphes défini précédemment.

D'après le lemme 5.2 la somme des exposants $\lambda_1 + \dots + \lambda_k$ du cocycle de Kontsevich-Zorich est égale à

$$\lambda_1 + \dots + \lambda_k = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{\mathcal{M}_k^{(1)}} \int_{G_k(M)} \log |\det A_T^k(q, I)|^{1/2} d\sigma^k(v) d\mu_k^{(1)}(x),$$

où $\det A_T^k(q, I) = \det A_{G_T(q)}^k(I)$ est le déterminant de la matrice dont le coefficient i, j est $(c_i(T), c_j(T))_q$ ou de façon équivalente $(m_i^+(T), m_j^+(T))_q$.

Pour clarifier un peu ce qui va suivre, plaçons-nous dans le cadre suivant. Soient X un espace «sympathique» sur lequel le groupe $SL(2, \mathbf{R})$ agit, ν une mesure invariante par l'action du sous-groupe $SO(2, \mathbf{R})$ et $f : X \rightarrow \mathbf{R}$ une fonction «régulière». Nous notons g_t la matrice diagonale $\text{diag}(e^t, e^{-t})$ et $\rho_\theta \in SO(2, \mathbf{R})$ la rotation d'angle θ . On veut évaluer le comportement de $f(g_t \cdot x)$ quand t tend vers l'infini. L'idée est d'introduire une complexification du temps : au lieu d'étudier $f(t \cdot x) := f(g_t \cdot x)$ on va étudier $f(z \cdot x)$ où z est dans le disque de Poincaré. Définissons pour $z = re^{i\theta}$ (avec $t = (1/2) \log \frac{1+r}{1-r}$)

$$f_x(z) = f(z \cdot x) = f((g_t \rho_\theta) \cdot x).$$

La formule du lemme précédent montre que si $h_x = \Delta_h f_x$ on a

$$\frac{1}{2\pi} \partial_t \int_0^{2\pi} f((g_t \rho_\theta) \cdot x) d\theta = \frac{1}{2} \tanh(t) \frac{1}{|D_t|} \int_{D_t} h(z \cdot x) \omega_P(z).$$

Comme ν est une mesure finie sur X invariante par l'action de $SO(2, \mathbf{R})$, on a en intégrant par rapport à ν

$$\partial_t \int_X f(g_t \cdot x) d\nu(x) = \frac{1}{2} \tanh(t) \frac{1}{|D_t|} \int_X \int_{D_t} h(z \cdot x) \omega_P(z) d\nu(x).$$

Si on suppose en plus que h vérifie pour tout t

$$\int_X h(g_t \cdot x) d\nu(x) = \int_X h d\nu,$$

c'est-à-dire si en plus de la $SO(2, \mathbf{R})$ -symétrie de ν , (h, ν) possède une symétrie (faible) par rapport au groupe diagonal, on obtient

$$\partial_t \int_X f(t \cdot x) d\nu(x) = \frac{1}{2} \tanh(t) \int_X h(x) d\nu(x),$$

et si on intègre en t ,

$$\int_X f(t \cdot x) d\nu(x) = \text{cste} + \log \cosh(t) \int_X h(x) d\nu(x).$$

Enfin si on divise par t et qu'on passe à la limite

$$(3) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \int_X f(t.x) d\nu(x) = \int_X h(x) d\nu(x).$$

Dans la suite l'espace X sera le produit $\mathcal{M}_\kappa^{(1)} \times G_k(M, \mathbf{R})$, $x = (q, I)$ et la mesure ν sera le produit de la mesure $\mu_\kappa^{(1)}$ par la mesure canonique σ_k sur $G_k(M, \mathbf{R})$. Notons que $\mu_\kappa^{(1)}$ est invariante par l'action de $SL(2, \mathbf{R})$ tandis que σ_k et donc ν ne sont invariante que par $SO(2, \mathbf{R})$. L'application $f_x(z)$ sera égale à $\log |\det A_z^k(q, I)|^{1/2}$. Il nous faut à présent identifier la fonction h et pour cela calculer le laplacien hyperbolique de f_x . La paramétrisation introduite par Forni que nous avons décrite en 4 se révèle alors très utile.

On a vu que l'évolution des $c_i(t)$ suivant le cocycle de Kontsevich-Zorich, c'est-à-dire l'évolution des c_i sous l'action du groupe diagonal \mathbf{A} se lisait dans la façon dont évoluent les $m_i^+(t)$ suivant la dynamique donnée par (1). De manière analogue Forni calcule le laplacien hyperbolique de f (il faut en plus dériver f par rapport à l'algèbre de Lie du groupe des rotations). Le résultat est alors le suivant : si on note

$$\begin{aligned} A_{ij}^k &= (m_i^+, m_j^+)_q, \\ H_{ij}^k &= (\pi_q^-(m_i^+), \pi_q^-(m_j^+))_q, \\ B_{ij}^k &= B_q(m_i^+, m_j^+) = (m_i^+, \overline{m_j^+})_q, \\ V_{ij}^k &= (\partial_q^+ v_i, \partial_q^+ \overline{v_j})_q = (\partial_q^- v_i, \partial_q^- \overline{v_j})_q \end{aligned}$$

on a

$$\Delta_h \log |\det(A_z^k)|^{1/2} = 4\text{tr}[(A_z^k)^{-1} H_z^k] - 2\text{tr}[(A_z^k)^{-1} B_z^k (A_z^k)^{-1} \overline{B_z^k}]$$

La fonction

$$\Phi_k = 2\text{tr}[(A^k)^{-1} H^k] - \text{tr}[(A^k)^{-1} B^k (A^k)^{-1} \overline{B^k}]$$

est en fait indépendante de la base choisie pour le k -plan isotrope I_k et définit une fonction $\Phi_k : Q_\kappa^{(1)} \times G_k(M, \mathbf{R}) \rightarrow \mathbf{R}_+$ invariante par l'action naturelle du groupe modulaire Γ_g . Soient $I_1 \subset \dots \subset I_g \subset H^1(M, \mathbf{R})$ une suite d'espaces isotropes, $q \in Q_\kappa^{(1)}$, et m_1^+, \dots, m_g^+ une base orthonormale de \mathcal{M}_q^+ tels que les c_1^+, \dots, c_k^+ représentent une base orthonormale de I_k (via $c_i = [\text{Re}(m_i^+ q^{1/2})]$) alors

- (1) $\Phi_1(q, I_1) = 2|\pi_q^-(m_1^+)|^2 - |B_q(m_1^+)|^2$,
- (2) $\Phi_k(q, I_k) = \Phi_g(q, I_g) - \sum_{i,j=k+1}^g |B_q(m_i^+, m_j^+)|^2$
- (3) $\Phi_g(q, I_g) = \Lambda_1(q) + \dots + \Lambda_g(q)$ où les $\Lambda_1(q) \geq \dots \geq \Lambda_g(q) \geq 0$ sont les valeurs propres de la matrice hermitienne positive H_q . Notons que dans le cas $k = g$ il n'y a pas de dépendance en I_g .

5.4. Conséquences

Revenons au problème de la somme des exposants de Lyapunov du cocycle de Kontsevich-Zorich. Nous avons vu précédemment que l'hypothèse sous laquelle la formule donnant la somme des exposants de Lyapunov est vraie et que la fonction

$h(q, I) = \Phi_k(q, I)$ a une symétrie supplémentaire sous l'action du groupe diagonal. Comme la mesure $\mu_\kappa^{(1)}$ est invariante pour l'action de $SL(2, \mathbf{R})$, ceci est assuré quand $k = g$ car dans ce cas $h(q, I) = \Phi_g(q, I) = \Lambda_1(q) + \dots + \Lambda_g(q)$ ne dépend *pas* de I . Par conséquent si $\mathcal{C}_\kappa^{(1)}$ est une composante connexe de $\mathcal{M}_\kappa^{(1)}$ (où la mesure $\mu_\kappa^{(1)}$ est ergodique ce qui permet d'appliquer la version ergodique du théorème d'Oseledec)

$$\lambda_1 + \dots + \lambda_g = \frac{1}{\mu_\kappa^{(1)}(\mathcal{C}_\kappa^{(1)})} \int_{\mathcal{C}_\kappa^{(1)}} (\Lambda_1(q) + \dots + \Lambda_g(q)) d\mu_\kappa^{(1)}(q).$$

Une formule explicite donnant la somme des exposants avait déjà été obtenue par Kontsevich et Zorich ([10]). Comme $\lambda_1 = 1$ et $\Lambda_1 \equiv 1$ la formule précédente donne

$$\lambda_2 + \dots + \lambda_g = \frac{1}{\mu_\kappa^{(1)}(\mathcal{C}_\kappa^{(1)})} \int_{\mathcal{C}_\kappa^{(1)}} (\Lambda_2(q) + \dots + \Lambda_g(q)) d\mu_\kappa^{(1)}(q),$$

et montre que $\lambda_2 > 0$ *pourvu que* l'on sache démontrer que (par exemple) Λ_g est *strictement* positive sur un ouvert de la composante $\mathcal{C}_\kappa^{(1)}$ (en fait Forni donne dans ce cas une démonstration plus simple).

Dans le cas où $1 \leq k < g$ et en supposant que l'on sache démontrer que $\Lambda_g(q)$ est *strictement* positive sur un ouvert de la composante $\mathcal{C}_\kappa^{(1)}$, l'argument précédent ne fonctionne pas directement mais dans le cas où k est inférieur ou égal à $g/2$ Forni démontre de façon analogue et par un argument de dimension que la somme $\lambda_1 + \dots + \lambda_k$ est strictement positive. Dans les autres cas, la conclusion de l'analyse est la suivante :

PROPOSITION 5.4. — *Si pour $1 \leq k \leq g - 1$ on a $\lambda_k > \lambda_{k+1} \geq 0$, alors*

$$\lambda_1 + \dots + \lambda_k = \frac{1}{\mu_\kappa^{(1)}(\mathcal{C}_\kappa^{(1)})} \int_{\mathcal{C}_\kappa^{(1)}} \Phi_k(q, E_k^+(q)) d\mu_\kappa^{(1)}(q),$$

où $E_k^+(q)$ est le sous-espace de $H^1(M, \mathbf{R})$ de dimension k correspondant aux exposants $\{\lambda_1, \dots, \lambda_k\}$.

5.5. Le «determinant locus»

Comme on vient de le voir, pour obtenir des résultats de positivité d'exposants de Lyapunov il est important de savoir que $\Lambda_g(q)$ est *strictement* positive sur un ouvert de la composante $\mathcal{C}_\kappa^{(1)}$ et il est donc important de connaître le lieu des q où $\Lambda_g(q)$ s'annule. La matrice des périodes va jouer un rôle important dans cette analyse géométrique. Si $\{a_1, \dots, a_g, b_1, \dots, b_g\}$ est une base canonique de l'homologie (donnant le marquage de M) et $\theta_1, \dots, \theta_g$ la base duale des différentielles holomorphes (i.e. telle que $\theta_i(a_j) = \delta_{ij}$) la matrice des périodes Π est la matrice complexe $g \times g$ d'éléments $\Pi_{ij} = \theta_i(b_j)$ qui vit dans l'espace de Siegel des matrices complexes symétriques et de partie réelle définie positive. Cette matrice Π dépend de la structure complexe choisie. Si $(M_t, q_t) = G_t(M, q)$, posons $\Pi_t = \Pi(M_t)$. Forni définit $\mathcal{D}_\kappa^{(1)}$ l'ensemble des $[q] \in \mathcal{M}_\kappa^{(1)}$ où

$$\det \left(\left(\frac{d\Pi}{dt} \right) \Big|_{t=0} \right) = 0$$

(c'est bien invariant par le groupe modulaire à cause du déterminant). C'est une surface analytique réelle dans $\mathcal{M}_\kappa^{(1)}$ de codimension 2. Le lemme clé est alors

LEMME 5.5. — L'ensemble des q où $\Lambda_g(q) = 0$ est égal à $\mathcal{D}_\kappa^{(1)}$.

Si on pose $\phi_i^+ = \theta_i/q^{1/2}$, $\{\phi_1^+, \dots, \phi_g^+\}$ est une base de \mathcal{M}_q^+ et il est possible de voir (formule de Rauch) que

$$\begin{aligned} \left(\frac{d\Pi}{dt}\right)|_{t=0} &= \int_M \phi_i^+ \phi_j^+ \omega_q \\ &= (\phi_i^+, \overline{\phi_j^+})_q \end{aligned}$$

c'est-à-dire

$$\left(\frac{d\Pi}{dt}\right)|_{t=0} = B_q(\phi_i^+, \phi_j^+).$$

Comme $\{\phi_1^+, \dots, \phi_g^+\}$ et $\{m_1^+, \dots, m_g^+\}$ sont deux bases, le déterminant $\det B_q(m_i^+, m_j^+)$ est nul si et seulement si $\det B_q(\phi_i^+, \phi_j^+)$ l'est. Comme $|\det H_q(m_i^+, m_j^+)| = |\det B_q(m_i^+, m_j^+)|^2$, on a bien la conclusion du lemme.

Ainsi, pour pouvoir démontrer que $\Lambda_g(q) > 0$ sur un ouvert de la composante $\mathcal{C}_\kappa^{(1)}$, il faut démontrer le théorème suivant :

THÉORÈME 5.6. — *Les composantes $\mathcal{C}_\kappa^{(1)}$ de $\mathcal{M}_\kappa^{(1)}$ ne sont jamais contenues dans le «determinant locus».*

La preuve de ce résultat nécessite une analyse fine de ce qui se passe au bord de l'espace des modules, c'est-à-dire quand on «pince» certaines courbes sur la surface. Il se trouve que dans cette situation les calculs sont plus faciles, car pincer revient à remplacer certains morceaux de cylindres sur M en des cylindres beaucoup plus longs (mais de même diamètre) ce qui a tendance à «concentrer la masse» en des endroits (les cylindres) où les calculs sont plus explicites.

6. CYCLES ASYMPTOTIQUES ET COURANTS

6.1. Cycles asymptotiques

Considérons, sur une surface compacte M , le flot ϕ_X^t d'un champ de vecteurs X dont les singularités sont de type selle ; en genre $g \geq 2$ un tel champ de vecteurs a des singularités $\Sigma = \{p_1, \dots, p_\sigma\}$ d'indices n_1, \dots, n_σ (nombres de séparatrices moins 1) qui vérifient $n_1 + \dots + n_\sigma = 2g - 2$ (Poincaré-Hopf). Les entiers σ, p_i, n_i ($1 \leq i \leq \sigma$) déterminent le type de singularité de X . Si μ est une mesure invariante pour le flot ϕ_X^t on peut définir pour toute 1-forme fermée α la quantité,

$$\int_M i_X \alpha d\mu,$$

qui dépend linéairement de α et on vérifie facilement que si α est exacte cette quantité est nulle. On a ainsi construit un élément $\rho_{X,\mu}$ dans le dual de $H^1(M, \mathbf{R})$ (que l'on identifie à $H_1(M, \mathbf{R})$) :

$$\rho_{X,\mu}([\alpha]) = \int_M i_X \alpha d\mu,$$

que l'on appelle le vecteur de rotation de μ . Quand μ est une mesure ergodique pour le flot, on sait d'après le théorème de Birkhoff que pour μ -presque tout $x \in M$

$$\begin{aligned} \rho_{X,\mu}([\alpha]) &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t (i_X \alpha)(\phi_X^t(x)) dt \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_{\gamma_t} \alpha, \end{aligned}$$

où γ_t est le morceau d'orbite de x pour $0 \leq s \leq t$ suivant le flot ϕ_t . Si on choisit t de façon que x et $\phi_X^t(x)$ soient proches, on peut fermer l'orbite par une petite transversale au flot pour obtenir un chemin fermé $\tilde{\gamma}_t$. On a alors dans $H_1(M, \mathbf{R})$

$$\lim_{t \rightarrow \infty} \frac{1}{t} [\tilde{\gamma}_t] = \rho_{X,\mu}.$$

On peut aussi définir ρ d'une autre manière quand μ est définie par une forme volume ω : puisque $0 = i_X(\alpha \wedge \omega) = i_X \alpha \omega + \alpha \wedge i_X \omega$

$$\rho_{X,\mu}(\alpha) = \int_M i_X \omega \wedge \alpha$$

et, via la dualité de Poincaré, $\rho_{X,\mu}$ s'identifie à $i_X \omega$. (Dans le cas général : Si γ est un chemin fermé sur M , on laisse évoluer γ suivant le flot entre 0 et t et on calcule l'aire balayée au cours du mouvement.) Puisque $i_X \circ d + d \circ i_X = \mathcal{L}_X$, on voit que la 1-forme $i_X \omega$ est fermée si et seulement si ω est invariante par le flot de X (la mesure que définit ω est donc invariante par le flot). On peut associer deux classes de cohomologie à $i_X \omega$ qui ont des rôles distincts :

- (1) d'une part sa classe de cohomologie dans $H^1(M, \Sigma, \mathbf{R})$ (de dimension $2g + \sigma - 1$) qui intervient dans la classification des *champs de vecteurs* X dont les singularités sont de type prescrit et qui préservent une forme volume fixée ω : un théorème de Katok (cf. [7] chap. 14) dit que deux tels champs de vecteurs qu'on peut relier par un chemin continu de champs de vecteurs sont orbite-équivalents (ou équivalents modulo changement de temps) si et seulement si leurs classes de cohomologie dans $H^1(M, \Sigma, \mathbf{R})$ sont colinéaires (la constante de colinéarité étant positive). La conjugaison peut être choisie Lipschitz sur M et C^∞ en dehors des singularités ;
- (2) d'autre part sa classe de cohomologie dans $H^1(M, \mathbf{R})$ (de dimension $2g$) qui intervient dans la classification des *mesures* μ -invariantes par X : un théorème de Katok (cf. *loc. cit.*) dit que deux mesures μ_1 et μ_2 invariantes par X et supportées sur des composantes transitives sont égales si et seulement si $\rho_{X,\mu_1} = \rho_{X,\mu_2}$ dans

$H^1(M, \mathbf{R})$. Par ailleurs l'image de ces mesures par ρ_X , dans $H^1(M, \mathbf{R})$ est un espace *isotrope*, donc de dimension inférieure à g .

Détaillons un peu ce dernier point. Si μ_1, μ_2 sont deux mesures invariantes que l'on suppose régulières (i.e données par des 2-formes C^∞ ω_1, ω_2) dont les nombres de rotation sont les mêmes, on a

$$i_X \omega_1 - i_X \omega_2 = df,$$

où f est une fonction C^∞ . En contractant l'égalité précédente suivant X on obtient $i_X df = 0$, c'est-à-dire $\mathcal{L}_X f = 0$: f est invariante par le flot. Si on suppose ce flot (quasi-) minimal alors f est constante et donc $df = 0$ soit

$$i_X \omega_1 = i_X \omega_2,$$

et en prenant le produit extérieur avec n'importe quelle 1-forme α on voit que les fonctions $(i_X \alpha)$ sont d' ω -intégrales nulles ($\omega = \omega_1 - \omega_2$) donc $\omega_1 = \omega_2$.

Si on ne fait aucune hypothèse de régularité sur les mesures μ_1, μ_2 , hormis l'hypothèse qu'elles sont sans atomes, on a encore

$$i_X \mu_1 - i_X \mu_2 = df,$$

mais cette fois-ci au sens des distributions (ou des courants) et on voit (intégrer l'équation précédente sur un chemin) que f est continue et donc, si X est quasi-minimal, constante.

Ceci suggère qu'il est important de travailler dans l'espace des courants.

Dans le cas «générique» le flot de ϕ_X^t est (quasi-) minimal et il n'y a donc, en dehors des mesures de Dirac aux singularités qu'une seule mesure invariante (ergodique) par X : l'image des mesures dans $H^1(M, \mathbf{R})$ est donc «génériquement» de dimension 1. Nous verrons par la suite que si au lieu de considérer des mesures invariantes on considère des *distributions* invariantes on pourra atteindre, en fonction de la régularité de ces distributions, des espaces de dimension g (lagrangiens) ou même $2g - 1$.

6.2. Courants

Un courant (cf. [14], [2]) est une forme linéaire C sur $\Lambda_0^* M$ (les formes C^∞ à support compact) continue au sens des distributions, i.e $C(\phi_j)$ tend vers 0 pour toute suite de formes à support compact ϕ_j tendant vers 0 en topologie C^∞ . Il est homogène de dimension p s'il est nul sur toute k -forme $k \neq p$. Les opérations usuelles de bord, contractions... se définissent par dualité (en général via Stokes) comme pour les distributions et un point important est le théorème de De Rham : tout courant peut être régularisé, plus précisément : tout courant est cohomologue (au sens des courants) à un courant lisse.

Dans $\mathcal{D}'(M - \Sigma)$ une distribution \mathcal{D} est un courant de dimension 0 et de degré 2 et on peut lui associer canoniquement un courant $\tilde{\mathcal{D}}$ de dimension 2 et de degré 0 tel que $\mathcal{D} = \tilde{\mathcal{D}}\omega_q$; tout courant de dimension 0 peut se représenter de façon unique sous cette forme. Afin de ne pas trop alourdir les notations, nous identifierons les courants de

dimension 0 et 2 et parlerons plus simplement de distributions. Un courant C homogène de degré 1 dans $\mathcal{D}'(M - \Sigma)$ est dit *basique* par rapport à un feuilletage \mathcal{F} si pour tout champ de vecteurs X tangent au feuilletage \mathcal{F}

$$i_X C = 0, \quad \mathcal{L}_X C = 0.$$

Dans la suite il sera important de travailler avec des courants qui se comportent bien aux singularités. En particulier un courant est dit *q-tempéré* s'il est dans le dual topologique de l'espace des formes α qui sont C^∞ et qui au voisinage de chaque singularité $p \in \Sigma$ d'ordre m (i.e $q(z) = z^{2m} dz^2$ au voisinage de p) s'écrivent $\alpha = \pi_p^*(\lambda_p)$ où λ est C^∞ et π_p s'écrit localement $\pi_p(z) = z^{m+1}/(m+1)$. On définit aussi des espaces de Sobolev : d'abord d'ordre $s \geq 0$ pour les 1-formes : une 1-forme est dans l'espace de Sobolev $\mathcal{H}_q^s(M)$ si ses contractions avec les champs de vecteurs S, T sont des fonctions dans $H_q^s(M)$; puis par dualité on définit les espaces $\mathcal{H}_q^{-s}(M)$, $s \geq 0$. L'espace de Sobolev d'ordre $s \in \mathbf{N}$, $\mathcal{B}_q^s(M)$ (resp. $\mathcal{B}_{-q}^s(M)$) est l'ensemble des courants basiques q -tempérés C tels que $i_S C = 0$ dans $H_q^{-s}(M)$ et $\mathcal{L}_S C = 0$ dans $\mathcal{H}_q^{-s-1}(M)$ (resp. $i_T C = 0$ dans $H_q^{-s}(M)$ et $\mathcal{L}_T C = 0$ dans $\mathcal{H}_q^{-s-1}(M)$).

En fait tout courant basique tempéré $C \in \mathcal{B}_q^s(M)$ peut se représenter sous la forme $\mathcal{D}\eta_S$ (i.e $\mathcal{D}.i_S\omega_q$) où \mathcal{D} est une distribution S -invariante (i.e $\mathcal{L}_S \mathcal{D} = 0$) dans $H_q^{-s}(M)$ définie par

$$\mathcal{D}\omega_q = -C^+ \wedge \eta_T.$$

et réciproquement si \mathcal{D} est une distribution invariante tempérée $\mathcal{D}\eta_S$ est un courant basique tempéré.

Tout courant basique C réel est fermé et définit un élément du dual $H_c^1(M - \Sigma, \mathbf{R})^*$ des formes à support compact ; par la dualité de Poincaré $H_c^1(M - \Sigma, \mathbf{R})^* = H^1(M - \Sigma, \mathbf{R})$. Si le courant est q -tempéré on peut en fait lui associer une classe de cohomologie dans $H^1(M, \mathbf{R})$. Nous notons j_q cette application. Les images par j_q des espaces \mathcal{B}_q^s ne dépendent que de la classe de cohomologie de $\text{Im}q^{1/2}$ dans $H^1(M, \Sigma, \mathbf{R})$ ce qui découle d'une version du théorème de Katok de la section 6.1 (1). À la différence du vecteur de rotation des mesures, la classe de cohomologie d'un courant basique tempéré ne détermine pas le courant de façon unique. D'autre part, il n'est pas clair que l'image par j_q de l'ensemble des courants basiques tempérés soit isotrope dans $H^1(M, \mathbf{R})$. En revanche, pour *presque tout* q (dans un sens à préciser) Forni apporte les réponses suivantes :

- (1) Soit q une différentielle quadratique. Alors, pour presque tout θ , l'application $\mathcal{B}_{e^{i\theta}q}^1(M) \rightarrow H^1(M, \mathbf{R})$ qui à un courant basique $qe^{i\theta}$ -tempéré associe sa classe de cohomologie est injective (pour des courants d'ordre $s \geq 1$ il existe en fait une suite exacte qui généralise ce résultat).
- (2) Il existe un entier $l > 1$ tel que pour tout $s \geq l$ et presque toute différentielle quadratique l'image par j_q dans $H^1(M, \mathbf{R})$ des courants basiques q -tempérés d'ordre s est de dimension $2g - 1$: c'est l'ensemble des classes c telles que $c \wedge [\text{Im}q^{1/2}] = 0$.

(3) Dans le cas où $s = 1$, l'image par j_q des courants basiques q -tempérés est pour presque tout q un espace lagrangien de $H^1(M, \mathbf{R})$.

Le cas où $s = 1$ est celui qui intervient dans l'étude du cocycle de Kontsevich-Zorich. On peut d'ailleurs identifier ce sous-espace lagrangien : c'est l'espace instable du cocycle de Kontsevich-Zorich au-dessus du point q . Nous reviendrons sur ce point plus tard.

6.3. Espaces stables/instables, courants basiques et cocycle de Forni

À ce stade nous ne savons pas si les exposants de Lyapunov du cocycle de Kontsevich-Zorich sont tous non nuls. Plaçons-nous dans l'hypothèse de la proposition 5.4 : pour $1 \leq k \leq g - 1$ on a $\lambda_k > \lambda_{k+1} = 0$ (on prend k maximal), c'est-à-dire que pour presque tout $q \in \mathcal{C}_\kappa^{(1)}$ on peut parler des espaces stables et instables $E^- = E_k^-$, $E^+ = E_k^+$ qui sont de dimension k .

THÉORÈME 6.1. — *Pour $\mu_\kappa^{(1)}$ -presque tout q les espaces $E^+(q)$, $E^-(q)$ coïncident avec $\mathcal{B}_q^1(M)$, $\mathcal{B}_{-q}^1(M)$ les espaces de courants basiques tempérés (c'est-à-dire que l'image dans $H^1(M, \mathbf{R})$ de $\mathcal{B}_{\pm q}^1(M)$ par j_q égale $E^\pm(q)$).*

A) Esquissons la preuve du fait que toute classe $[c]$ dans $E^+(q)$ peut être représentée par un courant basique tempéré d'ordre 1.

1) Déjà, d'après la théorie de Hodge c est cohomologue à une 1-forme harmonique fermée de la forme $\text{Re}(m^+ q^{1/2})$ où m^+ est méromorphe dans $L_q^2(M)$. Notons $c_t = G_t^{KZ}(c) = \text{Re}(m_t^+ q_t^{1/2})$ son image sous l'évolution du cocycle de Kontsevich-Zorich. Puisque c_t et c sont cohomologues, il existe une fonction U_t qu'on peut choisir de moyenne nulle telle que

$$(4) \quad dU_t = \text{Re}(m^+ q^{1/2}) - \text{Re}(m_t^+ q_t^{1/2}).$$

Contractons l'équation précédente suivant S ; nous obtenons

$$SU_t = \text{Re}(m^+) - e^t \text{Re}(m_t^+).$$

Nous allons démontrer que $|U_t|_{L_q^2}$ est bornée. On pourra ainsi extraire de U_t une sous-suite faiblement convergente dans L^2 vers un U et comme $e^t \text{Re}(m_t^+)$ tend vers 0 quand $t \rightarrow -\infty$ on aura au sens des distributions

$$SU = \text{Re}(m^+).$$

Pour cela, rappelons que

$$(5) \quad m_t^+ = \partial_t^+ v_t + \pi_t^-(m_t^+),$$

(où v_t est définie à une constante près) vérifie l'équation différentielle

$$\frac{dm_t^+}{dt} = \partial_t^- v_t - \overline{\pi_t^-(m_t^+)},$$

si bien qu'en dérivant (4) on obtient

$$\frac{d}{dt}(dU_t) = \operatorname{Re}(dv_t),$$

et en intégrant

$$(6) \quad \frac{d}{dt}U(t) = \operatorname{Re}(v_t).$$

Comme la décomposition (5) est orthogonale, on a

$$|\partial^+ v_t|_0 \leq |m_t^+|_0.$$

Forni démontre et utilise à présent une inégalité de Poincaré avec estimation de la constante :

$$|v_t|_0 \leq C(q_t)|\partial^+ v_t|_0$$

où la constante $C(q)$ est majorée par l'inverse de la longueur $l(q)$ de la plus courte géodésique entre les singularités de q_t (les éléments de Σ_{q_t}). Mais d'après un résultat de H. Masur [12] on sait que presque sûrement

$$\limsup_{t \rightarrow \pm\infty} \frac{-\log l(q_t)}{\log |t|} = \frac{1}{2}.$$

Par conséquent

$$|v_t|_0 \leq Ct|m_t^+|_0,$$

et d'après (6) ($U_0 = 0$)

$$|U_t|_0 \leq C \int_{-\infty}^0 s|m_s^+|_0 ds,$$

ce qui montre que $|U_t|_0$ est bornée pour $t \rightarrow -\infty$ puisque par définition q est dans $E^+(q)$.

2) Écrivons

$$dU = SU\eta_T + TU\eta_S$$

et utilisons l'égalité que nous venons de démontrer

$$SU = \operatorname{Re}(m^+).$$

Nous obtenons,

$$dU = (1/2)[(m^+(\eta_T + i\eta_S) + \overline{m^+}((\eta_T - i\eta_S))] + \mathcal{D}\eta_S$$

avec

$$\mathcal{D} = TU - im^+ + i\overline{m^+},$$

c'est-à-dire

$$dU = \operatorname{Re}(m^+q^{1/2}) + \mathcal{D}\eta_S,$$

ce que l'on voulait démontrer puisque $\mathcal{D}\eta_S$ est un courant basique tempéré d'ordre 1.

B) Passons à la réciproque. Il s'agit de démontrer qu'un courant basique tempéré d'ordre 1 pour le feuilletage $\mathcal{F}_{\pm q}$ a pour image dans $H^1(M, \mathbf{R})$ un élément de $E^\pm(q)$. Forni introduit un cocycle G_t^F (que nous appellerons le cocycle de Forni) sur l'espace

\mathcal{Z}'_κ des courants tempérés fermés d'ordre 1 qu'il définit de façon analogue au cocycle de Kontsevich-Zorich : ce cocycle envoie de façon triviale (c'est l'identité) la fibre $\{q\} \times \mathcal{H}_q^{-1}(M)$ (munie de sa q -norme d'espace de Hilbert) au-dessus de $q \in Q_\kappa^{(1)}$ sur la fibre $\{G_t(q)\} \times \mathcal{H}_{G_t(q)}^{-1}(M)$ (munie de sa $G_t(q)$ -norme d'espace de Hilbert) au-dessus de $G_t(q) \in Q_\kappa^{(1)}$. Ce cocycle vérifie

$$j_\kappa \circ G_t^F = G_t^{KZ} \circ j_\kappa,$$

et ses fibrés stable et instable sont en fait les fibrés \mathcal{B}^\mp . Pour démontrer ce point Forni utilise une construction de Burns-Katok ([8]), inspirée d'un article de Wojtkowski ([17]), et construit des fonctions de Lyapunov \mathcal{L}^\pm pour les fibrés \mathcal{B}_q^1 qui vérifient

$$\frac{d}{dt} \mathcal{L}^+ \circ G_t^F(C^+) > 0,$$

$$\frac{d}{dt} \mathcal{L}^- \circ G_t^F(C^+) < 0.$$

Leur construction se fait de la façon suivante (par exemple pour \mathcal{L}^+) : Pour un courant basique $C^+ \in \mathcal{B}_q^1(M)$ la distribution $\mathcal{D}^+ = C^+ \wedge \eta_T$ est dans l'espace de Sobolev $H_q^{-1}(M)$. D'après les travaux précédents de Forni ([4]) on peut écrire

$$\mathcal{D}^+ = \partial^+ U^+ + A^+,$$

où U^+ est une fonction de $L_q^2(M)$ orthogonale aux fonctions méromorphes \mathcal{M}_q^+ et $A^+ = \mathcal{D}(1)$ («l'intégrale» de \mathcal{D}). La fonction de Lyapunov \mathcal{L}^+ que Forni définit est

$$\mathcal{L}^+(C^+) = |U^+|_0^2 + (A^+)^2.$$

7. LE COCYCLE DE KONTSEVICH-ZORICH EST NON UNIFORMÉMENT HYPERBOLIQUE (FIN)

Pour notre propos, ce qu'il faut retenir de la section précédente c'est que pour presque tout q l'espace instable $E^+(q)$ (resp. $E^-(q)$) ne dépend que (de la classe de cohomologie dans $H^1(M, \Sigma_q, \mathbf{R})$) du feuilletage horizontal $\text{Im}(q^{1/2}) = 0$ (resp. vertical $\text{Re}(q^{1/2}) = 0$). Revenons à la formule de la proposition 5.4.

$$\lambda_1 + \dots + \lambda_k = \frac{1}{\mu_\kappa^{(1)}(\mathcal{C}_\kappa^{(1)})} \int_{\mathcal{C}_\kappa^{(1)}} \Phi_k(q, E_k^+(q)) d\mu_\kappa^{(1)}(q),$$

où $E_k^+(q)$ est l'espace $H^1(M, \mathbf{R})$ de dimension k correspondant aux exposants $\{\lambda_1, \dots, \lambda_k\}$. Comme on suppose $\lambda_{k+1} = \dots = \lambda_g = 0$ on a donc d'après la proposition 5.4 que pour $\mu_\kappa^{(1)}$ -presque tout q

$$\Phi_k(q, E^+(q)) = \Lambda_1(q) + \dots + \Lambda_g(q).$$

Si on choisit $\{c_1, \dots, c_g\} \in H^1(M_q, \mathbf{R})$ une base orthonormale (pour la norme de Hodge) telle que $\{c_1, \dots, c_k\}$ soit une base de $E_k^+(q)$ et si on pose $c_i = \text{Re}(m_i^+ q^{1/2})$ on aura alors presque partout

$$(7) \quad \sum_{i,j=k+1}^g |B_q(m_i^+, m_j^+)|^2 = 0.$$

Nous allons démontrer que si $k < g$ cette égalité est violée sur un ensemble de q de mesure positive. La belle idée de Forni est la suivante :

- A) comme la dépendance de $E^\pm(q)$ par rapport à q est mesurable, on peut trouver un ensemble de $\mu_\kappa^{(1)}$ -mesure positive proche de 1 tel que sur cet ensemble $q \mapsto E^\pm(q)$ soit continue et générique pour le théorème d'Oseledec ;
- B) il existe des différentielles quadratiques (que Forni appelle lagrangiennes)
 - i) dont toutes les feuilles verticales sont fermées (en dehors de celles qui passent par les singularités) ;
 - ii) dont les classes d'homologies engendrent un sous-espace lagrangien Λ de $H_1(M, \mathbf{R})$;
 - iii) telles que le dual de Poincaré $\mathcal{L}_0 = P(\Lambda)$ de ce sous-espace lagrangien soit transverse à $E^+(q)$;
 - iv) pour lesquelles on peut choisir g feuilles régulières verticales $\gamma_1, \dots, \gamma_g$ telles que $M - \cup\{\gamma_1, \dots, \gamma_g\}$ soit une sphère moins $2g$ trous. Choisissons-en une q_0 .
- C) on peut trouver proche de q_0 une différentielle quadratique q dans l'ensemble défini en A) et des courbes $\bar{\gamma}_1, \dots, \bar{\gamma}_g$ qui sont des q -trajectoires (des géodésiques pour la métrique plate R_q) presque verticales pour q et qui engendrent le même sous-espace lagrangien de $H_1(M, \mathbf{R})$. Appelons a_1, \dots, a_g ces courbes $\bar{\gamma}_1, \dots, \bar{\gamma}_g$ et b_1, \dots, b_g une base duale symplectique de a_1, \dots, a_g (pour la forme d'intersection). Le dual de Poincaré $\bar{\mathcal{L}}'_0$ de l'espace lagrangien engendré par les b_i peut être choisi transverse à $E^-(q)$.
- D) On laisse agir le flot géodésique jusqu'au temps T grand, de façon que les espaces $E^+(q_T)$ de dimension k et l'espace lagrangien $\bar{\mathcal{L}}'_0(T)$ soient proches. On a ainsi au-dessus de q_T deux espaces $E^+(q_T)$ et $\bar{\mathcal{L}}'_0(T)$ qui sont proches (on a utilisé le fait que $\bar{\mathcal{L}}'_0$ et $E^-(q)$ sont transverses).
- E) Forni effectue une déformation de (M, q_T) qui l'amène au bord de l'espace des modules (à la limite on obtient une surface pincée) tout en préservant le feuilletage horizontal. Pour cela :
 - i) par un élément du flot horocyclique on peut faire en sorte que les q -trajectoires $a_i = \bar{\gamma}_i(T)$ (qui font de petits angles non nuls avec le feuilletage vertical) deviennent des trajectoires verticales sans que l'on change le feuilletage horizontal ;
 - ii) on effectue ensuite un pincement (M, \tilde{q}_T) le long des courbes a_1, \dots, a_g (il faut imaginer que les bouts de cylindre dont les génératrices horizontales qui sont orthogonales aux courbes a_i qui les entourent deviennent très longs, les longueurs des courbes a_i restant constantes) ;
 - iii) on revient par l'inverse du flot horocyclique : on a effectué une conjugaison. Les courbes a_i et le feuilletage horizontal font toujours le même angle (mesuré par la nouvelle

différentielle quadratique \tilde{q}_T). Comme on a préservé le feuilletage horizontal et que $E^+(q)$ ne dépend que de la classe de cohomologie dans $H^1(M, \Sigma, \mathbf{R})$ du feuilletage horizontal, $E^+(q_T)$ et $E^+(\tilde{q}_T)$ sont les mêmes. On a donc au-dessus du point \tilde{q}_T , qui est près du bord de l'espace des modules, un espace instable $E^+(\tilde{q}_T) = E^+(q_T)$ et un plan lagrangien $\tilde{\mathcal{L}}_0(\tilde{q}_T)$ (qui est le dual de Poincaré de l'espace engendré par les courbes b_i) qui sont proches.

- F) Mais dans cette situation les calculs sont plus faciles à faire et on peut voir que si on a suffisamment pincé

$$|B_q(m_i^+, m_j^+)| \geq (1/2)$$

où les m_i sont associés aux courbes b_i (via la dualité de Poincaré) : si $\{c_1, \dots, c_g\}$ est une base orthonormale pour la forme de Hodge du dual de Poincaré de l'espace engendré par les b_i , on a posé $c_i = \text{Re}(m_i^+ q^{1/2})$.

- G) En utilisant la continuité de B_q , Forni trouve ainsi un ensemble de mesure positive où (7) n'est pas satisfaite.

8. LE COCYCLE DE FORNI

Le théorème fondamental qui géométrise les résultats précédents est le suivant :

THÉORÈME 8.1. — *Le fibré $\mathcal{Z}_\kappa^1 \subset \mathcal{H}_\kappa^{-1}$ des courants fermés d'ordre 1 admet la décomposition suivante qui est G_t^F -invariante :*

$$\mathcal{Z}_\kappa^1 = \mathcal{B}_{\kappa,+}^1 \oplus \mathcal{B}_{\kappa,-}^1 \oplus \mathcal{E}_\kappa^1.$$

- *Le fibré \mathcal{E}_κ^1 est défini au-dessus de tout point $q \in \mathcal{M}_\kappa$ et sa fibre \mathcal{E}_q^1 (de dimension infinie) au-dessus du point q est l'ensemble des courants exacts de \mathcal{Z}_q^1*

$$\mathcal{E}_q^1 = \{dU, \quad U \in L_q^2(M)\}.$$

- *Pour $\mu_\kappa^{(1)}$ -presque tout q les fibres $\mathcal{B}_{\pm q}^1$ de $\mathcal{B}_{\kappa,\pm}^1$ sont de dimensions finies et égales.*
- *La restriction du cocycle de Forni G_t^F au fibré*

$$\mathcal{B}_\kappa^1 = \mathcal{B}_{\kappa,+}^1 \oplus \mathcal{B}_{\kappa,-}^1,$$

est mesurablement isomorphe au cocycle de Kontsevich-Zorich (défini sur le fibré $\mathcal{H}_\kappa^1(M, \mathbf{R})$ de fibre $H^1(M_q, \mathbf{R})$) et il a donc le même spectre de Lyapunov. Les fibrés $\mathcal{B}_{\kappa,\pm}^1$ correspondent à E_κ^\pm .

- *L'exposant de Lyapunov du cocycle de Forni sur \mathcal{E}_κ^1 est zéro.*

9. DÉVIATIONS DES MOYENNES ERGODIQUES

9.1. Comment évaluer une moyenne ergodique

Considérons le cas où le champ de vecteurs X que l'on étudie est le champ S introduit précédemment. Pour une fonction $f : M \rightarrow \mathbf{R}$ «régulière» on veut évaluer pour un $p \in M$ générique le comportement quand $t \rightarrow \infty$ de

$$\frac{1}{t} \int_0^t f(\Phi_S^s(p)) ds$$

où Φ_S est le flot associé à S . L'intégrale précédente se réécrit

$$(8) \quad \int_0^t f(\Phi_S^s(p)) ds = \int_{\gamma_+^t(p)} \alpha$$

où $\gamma_+^t(p)$ est le chemin (non fermé) $\Phi_S^s(p)$ pour $0 \leq s \leq t$ et α est la 1-forme (non fermée)

$$\alpha = f\eta_T.$$

Si on identifie $\gamma_+^t(p)$ à un courant on a

$$\int_0^t f(\Phi_S^s(p)) ds = (\gamma_+^t(p) \wedge \alpha)(1)$$

Soit $I_{+q}(p)$ (resp. $I_{-q}(p)$) le segment vertical (resp. horizontal) centré en p et de longueur $l(q)$ où $l(q)$ est la longueur de la plus petite connexion géodésique entre deux singularités de q . Nous dirons que T est un *temps de retour horizontal* (resp. *vertical*) de p si $\Phi_q(p, T) \in I_{+q}(p)$ (resp. $\Phi_{-q}(p, T) \in I_{-q}(p)$) et le segment d'orbite $\gamma_q^T(p)$ (resp. $\gamma_{-q}^T(p)$) est appelé une trajectoire de retour horizontale (resp. verticale). Le *premier temps de retour horizontal* (resp. *vertical*) est défini comme le plus petit temps de retour horizontal (resp. vertical). Pour tout temps T (qui n'est pas forcément un temps de retour) on peut fermer la trajectoire horizontale (resp. verticale) $\gamma_q^T(p)$ (resp. $\gamma_{-q}^T(p)$) par le plus petit segment géodésique joignant les extrémités de $\gamma_{\pm}^T(p)$ et on note $\hat{\gamma}_{\pm}^T(p)$ le chemin fermé ainsi obtenu. Si T est un temps de retour, le segment qui fait la jonction est évidemment vertical (resp. horizontal).

Puisque

$$(\mathcal{F}_{q_t}, \mathcal{F}_{-q_t}) = (e^{-t}\mathcal{F}_q, e^t\mathcal{F}_{-q}),$$

toute trajectoire de retour horizontale pour \mathcal{F}_{q_t} est une trajectoire de retour horizontale pour \mathcal{F}_q (au moins pour $t < 0$, $|t|$ suffisamment grand).

9.2. Temps de retour principaux

Considérons une suite de temps négatifs $-t_k$ tels que q_{-t_k} soient très proches de q_0 dans $\mathcal{M}_\kappa^{(1)}$. L'ergodicité du flot de Teichmüller sur $\mathcal{M}_\kappa^{(1)}$ pour $\mu_\kappa^{(1)}$ garantit l'existence d'une telle suite. Les propriétés métriques des q_{-t_k} sont donc essentiellement les mêmes que celles de q , en particulier les longueurs $l(q_{-t_k})$ des plus petits segments géodésiques

entre les singularités de q_{-t_k} sont pratiquement les mêmes et par conséquent, puisque les longueurs des intervalles verticaux $I_{q_{-t_k}}(p)$ dans la métrique définie par q sont très petits, on a une suite décroissante d'intervalles emboîtés $I_q(p) \supset \dots \supset I_{q_{-t_k}}(p) \supset I_{q_{-t_{k+1}}}(p)$. Si $T = T_{q_{-t_k}}^{(1)}(p)$ est un temps de premier retour pour q_{-t_k} alors Te^{t_k} est un temps de retour pour q (mais pas de premier retour). On dit que les

$$(9) \quad T_q^{(k)}(p) = Te^{t_k} = T_{q_{-t_k}}^{(1)}(p)e^{t_k}$$

sont des temps de retour principaux pour q (en p). Tels que nous les avons définis les temps de retour ne sont pas vraiment canoniques. Par contre près d'un q «générique» pour Oseledec (et pour d'autres propriétés) on peut choisir sur un petit bout d'hypersurface contenant q et transverse au flot géodésique un ensemble \mathcal{P} de mesure positive (de densité proche de 1) et définir les $-t_k$ comme les temps de retour sur \mathcal{P} de $G_t(q)$ ($t < 0$).

Si on note $\gamma_q^{(k)}(p)$ la trajectoire de retour horizontale correspondant au temps $T_q^{(k)}(p)$ et $\hat{\gamma}_q^{(p)}$ la trajectoire fermée associée on a

- en homologie

$$\hat{\gamma}_q^{(k)}(p) = G_{t_k}^{KZ}(\hat{\gamma}_{G_{-t_k}^{KZ}(q)}^{(1)}(p)),$$

(on a noté G^{KZ} le cocycle de Kontsevich-Zorich agissant en homologie) ;

- dans \mathcal{Z}^1 les courants tempérés fermés d'ordre 1

$$\hat{\gamma}_q^{(k)}(p) = G_{t_k}^F(\hat{\gamma}_{G_{-t_k}^F(q)}^{(1)}(p));$$

- et dans \mathcal{H}^{-1} les courants tempérés d'ordre 1

$$\gamma_q^{(k)}(p) = G_{t_k}^F(\gamma_{G_{-t_k}^F(q)}^{(1)}(p)).$$

Comme nous avons besoin d'estimer des quantités de la forme (8) ce sont plutôt les deux dernières égalités qui sont importantes. Si on note $\Pi_q^{\pm i} : \mathcal{Z}_q^1 \rightarrow E_i^{\pm}$ la projection sur l'espace stable/instable (de dimension finie) correspondant à l'exposant de Lyapunov $\pm\lambda'_i$ on a par exemple une estimée de la forme

$$(10) \quad |\Pi_q^{\pm i}(\hat{\gamma}_q^{(k)}(p))|_{-1} \leq \text{cste. exp}(\lambda'_i |t_k|)$$

pour tous $\lambda'_i{}^+ > \lambda'_i > 0$, $-\lambda'_i < \lambda'_i{}^- < 0$

9.3. Découper une orbite en trajectoires de retour principales

Le lemme «arithmétique» suivant est pour cela crucial. C'est l'analogue du lemme fondamental de la preuve du théorème de Denjoy-Koksma (cf. aussi Zorich [19]).

LEMME 9.1. — Sous les hypothèses précédentes, pour tout $T > 0$ il existe une suite finie de points $(p_j^{(k)})$ de l'orbite $\gamma_q^T(p)$, $1 \leq k \leq n$, $1 \leq j \leq m_k$ tels que les trajectoires de retour principales $\gamma_q^{(k)}(p_j^{(k)})$ ne s'intersectent pas et telles que

$$(11) \quad \gamma_q^T(p) = \sum_{k=1}^n \sum_{j=1}^{m_k} \gamma_q^{(k)}(p_j^{(k)}) + b_q^T(p),$$

avec

$$(12) \quad m_k \leq \text{cste.} \exp(|t_{k+1}| - t_k),$$

et la longueur du reste vérifie

$$(13) \quad L(b_q^T(p)) \leq \text{cste.}$$

Il n'est pas complètement évident, mais Forni le démontre, que l'inégalité (13) entraîne (pour q dans un compact \mathcal{P})

$$(14) \quad |b_q^T(p)|_{-1} \leq \text{cste.}$$

Voyons à présent comment on utilise les résultats précédents.

Déjà, puisque l'ensemble \mathcal{P} de la section 9.2 est de mesure positive, disons μ , on a du fait de l'ergodicité du flot de Teichmüller sur \mathcal{M}_κ

$$\lim_{k \rightarrow \infty} \frac{|t_k|}{k} = \frac{1}{\mu},$$

ce qui donne $|t_{k+1} - t_k| \leq C_\epsilon \epsilon k$ pour tout $\epsilon > 0$ et par conséquent d'après (10), (11), (12), (14) et le théorème 8.1 on obtient

$$\begin{aligned} \limsup_{T \rightarrow \infty} \frac{\log |\Pi_q^{+i}(\gamma_q^T(p))|_{-1}}{\log T} &\leq \lambda'_i, \\ \limsup_{T \rightarrow \infty} \frac{\log |\Pi_q^{-i}(\gamma_q^T(p))|_{-1}}{\log T} &= 0, \\ \limsup_{T \rightarrow \infty} \frac{\log |\Pi_q^\varepsilon(\gamma_q^T(p))|_{-1}}{\log T} &= 0. \end{aligned}$$

On a utilisé en outre la définition (9) des temps de retour principaux et un résultat que démontre Forni qui dit qu'on peut choisir la section \mathcal{P} (définie section 9.2) de façon que les temps de *premier* retour soient bornés.

Enfin notons que la première inégalité est une égalité pour presque tout p (mais il faut travailler un peu plus).

9.4. Moyennes ergodiques pour le champ de vecteurs S

Comme précédemment notons $E_i^+(q)$ le sous-espace de $\mathcal{B}_q^1(M)$ (l'ensemble des courants basiques tempérés d'ordre 1) associé à l'exposant $\lambda'_i > 0$ de multiplicité m_i (rappelons que c'est l'image de la projection Π_q^{+i}) et soit $C_{i,j}^+$, $1 \leq j \leq m_i$ une base de courants basiques tempérés d'ordre 1 de $E_i^+(q)$. Les distributions tempérées d'ordre 1

$$\mathcal{D}_{i,j}^S = C_{i,j}^+ \wedge \eta_T,$$

sont S -invariantes. De plus, pour toute fonction dans $H^1(M)$

$$\begin{aligned} C_{i,j}^+ \wedge (f\eta_T) &= fC_{i,j}^+ \wedge \eta_T \\ &= \mathcal{D}_{i,j}^S(f). \end{aligned}$$

Par conséquent si on note $\mathcal{I}_S^1(\lambda'_i)$ l'espace des distributions S -invariantes engendrées par les $\mathcal{D}_{i,j}$, $1 \leq j \leq m_i$, on a bien, en utilisant le résultat de la fin de la section précédente, que si f dans $H^1(M)$ est annulée par les distributions $\mathcal{D} \in \mathcal{I}_S^1(\lambda'_1) \oplus \cdots \oplus \mathcal{I}_S^1(\lambda'_i)$ alors

$$\limsup_{T \rightarrow \infty} \frac{\log \left| \int_0^T f(\Phi_S(p, \tau)) d\tau \right|}{\log T} \leq \lambda'_{i+1}.$$

La démonstration de l'existence de distributions tempérées d'ordre 1, $\mathcal{D}_{i+1} \in \mathcal{I}_S^1(\lambda'_{i+1})$, dont la non annulation par f assure l'égalité dans l'équation précédente est un peu plus délicate.

9.5. Le cas général

Supposons à présent que X soit un champ de vecteurs dont le flot est quasi-minimal avec des singularités de type selle d'indice ι_k aux points p_k , $k = 1, \dots, \sigma$, et préserve une forme volume ω (non singulière). On sait alors que l'on peut réaliser le feuilletage \mathcal{F}_X en orbites de X par le feuilletage horizontal d'une différentielle quadratique q (qui est le carré d'une forme holomorphe) avec $\kappa = -2\iota$. Il existe alors une fonction W avec des zéros d'ordre fini aux points de Σ telle que $\omega = W^{-1}\omega_q$ et $X = WS$. On considère l'espace de Sobolev

$$H_W^1(M) = \{f, \quad W^{-1}f \in H^1(M)\},$$

et on pose pour $f \in H_W^1(M)$

$$\mathcal{D}^X(f) = \mathcal{D}^S(W^{-1}f);$$

la distribution $\mathcal{D}^X \in H_W^{-1}(M)$ est X -invariante. Notons $\gamma_X^T(p)$ l'orbite de p du temps 0 au temps T sous l'action de X . Il existe un temps $\tilde{T}(T)$ tel que

$$\gamma_S^{\tilde{T}(T)}(p) = \gamma_X^T(p).$$

Pour toute fonction $f \in H_W^1(M)$ on a

$$\begin{aligned} \int_0^T f(\Phi_X(p, \tau)) d\tau &= \int_{\gamma_X^T(p)} W^{-1}f \eta_T, \\ &= \int_{\gamma_S^{\tilde{T}(T)}(p)} W^{-1}f \eta_T. \end{aligned}$$

Mais la moyenne $(1/T) \int_0^T W(\Phi_X(p, \tau)) d\tau$ qui est égale à $(1/T)L_q(\gamma_X^T(p))$ (où $L_q(\cdot)$ est la q -longueur), c'est-à-dire $(\tilde{T}(T)/T)$, converge d'après le théorème ergodique vers $\int_M W\omega = \int_M \omega_q = 1$. On obtient donc la conclusion du théorème.

BIBLIOGRAPHIE

- [1] P. Arnoux, Ergodicité générique des billards polygonaux [d'après Kerchhoff, Masur, Smillie], Sémin. Bourbaki (1987/88), *Astérisque* **161-162** (1988), 202-221.
- [2] G. De Rham, *Variété différentiables, formes, courants, formes harmoniques, Actualités Scientifiques et Industrielles*, **1222b**, Hermann, Paris, 1973.
- [3] A. Fathi, F. Laudenbach, V. Poénaru, Travaux de Thurston sur les surfaces, Séminaire Orsay, *Astérisque* **66-67** (1979).
- [4] G. Forni, Solution of the cohomological equation for area-preserving flows on compact surfaces of higher genus, *Ann. of Math.* **146** (1997), 295-344.
- [5] G. Forni, Deviation of ergodic averages for area preserving flows on surfaces of higher genus, *Ann. of Math.* **155** (2002), 1-103.
- [6] S. Kerchhoff, H. Masur, J. Smillie, Ergodicity of Billiard flows and Quadratic Differentials, *Ann. of Math.* **124** (1986), 293-311.
- [7] A. Katok, B. Hasselblat, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge Univ. Press, Cambridge, 1995.
- [8] A. Katok, Infinitesimal Lyapunov functions, invariant cones families and stochastic properties of smooth dynamical systems (with the collaboration of K. Burns), *Erg. Theo. Dyn. Sys.* **14** (1994), 757-785.
- [9] M. Kontsevich, Lyapunov exponents and Hodge Theory, dans *The Mathematical Beauty of Physics* (Saclay, 1996), *Adv. Ser. Math. Phys.* **24**, 318-332, World Sci. Publ., River Edge, NJ, 1997.
- [10] M. Kontsevich, A. Zorich, Lyapunov exponents and Hodge Theory, arXiv:hep-th/9701164v1 28 Jan 1997.
- [11] H. Masur, Interval exchange transformations and measured foliations, *Ann. of Math.* **115** (1982), 169-200.
- [12] H. Masur, Logarithmic law for geodesics in moduli space, in *Mapping Class Groups and Moduli Spaces of Riemann Surfaces* (Göttingen, 1991/ Seattle, WA, 1991), *Contemp. Math.* **150**, 229-245, A.M.S., Providence, RI, 1993.
- [13] S. Marmi, P. Moussa, J.-C. Yoccoz, *The cohomological equation for Roth type interval exchange map*, preprint.
- [14] L. Schwartz, *Théorie des distributions*, Hermann, Paris, 1966.
- [15] W.A. Veech, Gauss measures for transformations on the space of interval exchange maps, *Ann. of Math.* **115** (1982), 201-242.
- [16] W.A. Veech, The Teichmüller geodesic flow, *Ann. of Math.* **124** (1986), 441-530.
- [17] M. Wojtkowski, Invariant family of cones and Lyapunov exponents, *Erg. Theo. Dyn. Sys.* **5** (1985), 145-161.
- [18] A. Zorich, Finite Gauss measure on the space of interval exchange transformations. Lyapunov exponents. *Ann. Inst. Fourier* **46** (1996), 325-370.

- [19] A. Zorich, Deviation for interval exchange transformation, *Erg. Theo. Dyn. Sys.* **17** (1997), 1477-1499.

Raphaël KRIKORIAN

Laboratoire de Probabilités
et Modèles Aléatoires

Université Pierre et Marie Curie

Boîte courrier 188

F-75252 PARIS Cédex 05

E-mail : `krikoria@moka.ccr.jussieu.fr`

INÉGALITÉ DE BRUNN-MINKOWSKI-LUSTERNIK, ET AUTRES INÉGALITÉS GÉOMÉTRIQUES ET FONCTIONNELLES

par **Bernard MAUREY**

INTRODUCTION

La théorie des corps convexes a commencé à la fin du 19ème siècle avec l'inégalité de Brunn, généralisée ensuite en inégalité de Brunn-Minkowski-Lusternik qui s'applique à des ensembles non nécessairement convexes. Ce thème a depuis longtemps des contacts avec les problèmes isopérimétriques et avec des inégalités d'Analyse, telles que celles qui traduisent les plongements de Sobolev. Nous allons développer quelques aspects plus récents des inégalités géométriques, dont certains sont liés à la technique du transport de mesure, notamment le transport dit «de Brenier». On ne trouvera pas ici l'approche d'un résultat faramineux, mais des pistes vers un ensemble convergent de techniques qui ont prouvé leur applicabilité.

L'essentiel de notre travail sera fait dans l'espace euclidien \mathbb{R}^n ; on notera le produit scalaire par $x \cdot y = \sum_{i=1}^n x_i y_i$ et la norme par $|x| = (x \cdot x)^{1/2}$. On notera $\int_{\mathbb{R}^n} f(x) dx$ ou bien $\int_{\mathbb{R}^n} f$ l'intégrale d'une fonction f pour la mesure de Lebesgue sur \mathbb{R}^n . On notera $\mathbf{1}_A$ la fonction indicatrice d'un sous-ensemble A de \mathbb{R}^n , et $A + B$ la somme de Minkowski de deux sous-ensembles A, B de \mathbb{R}^n , égale à $\{a + b : a \in A, b \in B\}$.

1. BRUNN-MINKOWSKI ET INÉGALITÉS GÉOMÉTRIQUES

L'inégalité de Brunn-Minkowski *sans dimension* se formule ainsi : si A, B sont deux compacts non vides de \mathbb{R}^n , et si on note $|A|, |B|$ leurs volumes (pour la mesure de Lebesgue), on a

$$\left| \frac{A + B}{2} \right| \geq |A|^{1/2} |B|^{1/2}.$$

En réalité, l'inégalité $|(1-t)A + tB| \geq |A|^{1-t} |B|^t$ est valable pour tout $t \in [0, 1]$, et elle se transforme facilement par des arguments d'homogénéité en la forme classique pour l'inégalité de Brunn-Minkowski dans \mathbb{R}^n ,

$$|A + B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

Si A est convexe et si B est un translaté d'un homothétique de A , l'inégalité devient une égalité. Appliquée en prenant pour B une boule euclidienne de rayon tendant vers 0, cette inégalité conduit à une démonstration de l'inégalité isopérimétrique dans \mathbb{R}^n . Démontrée d'abord par Brunn (1887) en dimension 2 ou 3 pour deux convexes, reprise par Minkowski au tout début du 20ème siècle, l'inégalité a été étendue par Lusternik à des compacts quelconques de \mathbb{R}^n ; curieusement, le nom de Lusternik est rarement associé de nos jours à cette extension considérable (voir [Lus], [HeM]). Hadwiger et Ohmann [HaO] en donnent une démonstration assez simple, en approchant A et B par des réunions finies de rectangles, et en réalisant une association bien choisie entre parties de A et B de même mesure ; la preuve finit par l'inégalité entre moyennes arithmétiques et géométriques. Ces deux ingrédients se retrouvent dans la plupart des autres preuves.

L'article récent de Richard Gardner [Gar] couvre presque tous les thèmes traités dans cet exposé, et bien d'autres qui ne seront pas abordés ici.

1.1. Prékopa-Leindler

On attribue généralement à Prékopa le résultat suivant : si $\varphi(x, t)$ est convexe sur $\mathbb{R}^n \times \mathbb{R}$, la fonction Φ définie sur \mathbb{R} par

$$e^{-\Phi(t)} = \int_{\mathbb{R}^n} e^{-\varphi(x,t)} dx$$

est convexe (on admettra les valeurs infinies). L'inégalité de Prékopa-Leindler a un rapport plus direct avec Brunn-Minkowski-Lusternik : on se donne $0 < \theta < 1$ et trois fonctions réelles s.c.i. f_0, f_θ, f_1 sur \mathbb{R}^n qui vérifient pour tous $x_0, x_1 \in \mathbb{R}^n$ l'inégalité

$$(H_{PL}) \quad f_\theta((1-\theta)x_0 + \theta x_1) \leq (1-\theta)f_0(x_0) + \theta f_1(x_1)$$

et on obtient l'énoncé qui suit.

THÉORÈME 1.1 (Prékopa, Leindler). — *Si les fonctions f_0, f_θ, f_1 satisfont (H_{PL}) , on a*

$$\int_{\mathbb{R}^n} e^{-f_\theta(x)} dx \geq \left(\int_{\mathbb{R}^n} e^{-f_0(x)} dx \right)^{1-\theta} \left(\int_{\mathbb{R}^n} e^{-f_1(x)} dx \right)^\theta.$$

En attribuant ce résultat aux seuls Prékopa [Pre] et Leindler [Lei], nous faisons un choix de simplicité qui n'est sans doute pas historiquement tout à fait correct ; en effet, la paternité de ce type de résultat a été revendiquée par deux groupes différents (voir Das Gupta [DaG] pour une autre vision de l'histoire). Le théorème 1.1 redonne immédiatement Brunn-Minkowski-Lusternik en prenant les fonctions f_0, f_θ, f_1 égales à 0 sur les ensembles $A_0, A_\theta = (1-\theta)A_0 + \theta A_1, A_1$, et égales à $+\infty$ en dehors, de sorte que pour $j = 0, \theta, 1$ on ait l'égalité $\mathbf{1}_{A_j} = e^{-f_j}$.

Pour éviter ces valeurs infinies, il est souvent plus agréable d'écrire l'hypothèse sous la forme suivante : trois fonctions réelles positives s.c.s. g_0, g_θ, g_1 sur \mathbb{R}^n vérifient pour tous $x_0, x_1 \in \mathbb{R}^n$ l'inégalité

$$(H_{PL2}) \quad g_\theta((1-\theta)x_0 + \theta x_1) \geq g_0(x_0)^{1-\theta} g_1(x_1)^\theta$$

et on déduit que $\int g_\theta \geq (\int g_0)^{1-\theta} (\int g_1)^\theta$. Cette inégalité est adaptée à l'étude des mesures log-concaves. Une mesure μ sur \mathbb{R}^n à densité log-concave s'écrit $d\mu(x) = e^{-\varphi(x)} dx$, avec φ convexe sur \mathbb{R}^n ; on voit facilement que le triplet $g_j = \mathbf{1}_{A_j} e^{-\varphi}$, $j = 0, \theta, 1$ avec $A_\theta = (1-\theta)A_0 + \theta A_1$ vérifie l'hypothèse (H_{PL2}) , ce qui conduit à l'inégalité

$$(1) \quad \mu((1-\theta)A_0 + \theta A_1) \geq \mu(A_0)^{1-\theta} \mu(A_1)^\theta.$$

Appliquée à une mesure log-concave symétrique (c'est-à-dire invariante par $x \rightarrow -x$), cette inégalité donne le résultat de T.W. Anderson (en fait, le résultat d'Anderson [And] est un peu plus général que le théorème 1.2, et il a été montré directement à partir de Brunn-Minkowski) : si $\theta = 1/2$, si $A_0 = C + v$ et $A_1 = C - v$, où C est un convexe symétrique et v un vecteur quelconque, on trouve que $(A_0 + A_1)/2 = C$, pendant que $\mu(A_0) = \mu(A_1)$ par la symétrie de μ ; ainsi, on obtient

$$\mu(C) = \mu((A_0 + A_1)/2) \geq \mu(A_0)^{1/2} \mu(A_1)^{1/2} = \mu(C + v),$$

ce qui signifie ceci : parmi les translatés d'un convexe symétrique, le convexe centré a la plus grande mesure.

THÉORÈME 1.2 (Anderson). — *Si μ est une mesure log-concave symétrique sur \mathbb{R}^n , si C est un convexe symétrique et v un vecteur quelconque, on a*

$$\mu(C + v) \leq \mu(C).$$

Un cas particulier important est celui de la mesure gaussienne. Désignons par γ_n la mesure gaussienne standard sur \mathbb{R}^n , dont la densité est $(2\pi)^{-n/2} \exp(-|x|^2/2)$. Elle est bien sûr log-concave, symétrique. Le principe précédent s'applique donc à γ_n , et il joue un rôle important dans certaines questions de statistique.

1.2. Isopérimétrie gaussienne

Le problème isopérimétrique gaussien a été résolu par Christer Borell [Bo1], Vladimir Sudakov et Boris Tsirelson [SuT]. Plus tard, Antoine Ehrhard [Ehr] a donné une autre démonstration et apporté d'autres informations, que nous discuterons plus loin. Si A est un sous-ensemble fermé de \mathbb{R}^n , on désigne par A_ε son épaissement de taille $\varepsilon > 0$ pour la distance euclidienne,

$$A_\varepsilon = \{x \in \mathbb{R}^n : \text{dist}(x, A) \leq \varepsilon\}.$$

THÉORÈME 1.3. — *Parmi les ensembles fermés A de mesure gaussienne $\gamma_n(A) = a$ fixée, les demi-espaces affines minimisent l'accroissement de mesure $\gamma_n(A_\varepsilon) - \gamma_n(A)$, ou ce qui revient au même, minimisent la mesure $\gamma_n(A_\varepsilon)$.*

Par exemple, lorsque $\gamma_n(A) = a = 1/2$, un demi-espace affine B de même mesure que A est donné par $B = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_1 \leq 0\}$, et on peut énoncer en passant aux

complémentaires

$$\gamma_n(A_\varepsilon^c) \leq \gamma_n(B_\varepsilon^c) = \gamma_n(\{x_1 > \varepsilon\}) = \int_\varepsilon^{+\infty} d\gamma_1(t).$$

Si f est une fonction 1-lipschitzienne sur \mathbb{R}^n (vérifiant $|f(x) - f(y)| \leq |x - y|$ pour tous x, y) possédant une *valeur médiane* m , c'est-à-dire une valeur telle que $\gamma_n(f \leq m) = 1/2$, on voit que f reste inférieure ou égale à $m + \varepsilon$ sur l'épaisi A_ε de $A = \{f \leq m\}$, donc

$$\gamma_n(\{f > m + \varepsilon\}) \leq \gamma_n(A_\varepsilon^c) \leq \int_\varepsilon^{+\infty} d\gamma_1(t);$$

on a la même majoration pour $\gamma_n(\{f < m - \varepsilon\})$, ce qui donne une bonne borne pour la probabilité $\gamma_n(\{|f - m| > \varepsilon\})$ que f dévie de plus de ε de sa valeur médiane m ,

$$\gamma_n(\{|f - m| > \varepsilon\}) \leq \gamma_1(\{t \in \mathbb{R} : |t| > \varepsilon\}) \leq e^{-\varepsilon^2/2}.$$

Cette propriété de concentration a eu de nombreuses applications en *théorie asymptotique*, la théorie des espaces normés de dimension tendant vers l'infini, voir par exemple le fameux théorème de Dvoretzky dans le livre de Gilles Pisier [Pis]. Cependant, il est suffisant pour ce type d'applications d'avoir un résultat moins précis, de la forme

$$\gamma_n(\{|f - m_1| > \varepsilon\}) \leq 2 e^{-\varepsilon^2/4}$$

par exemple, où m_1 peut désigner la moyenne de f au lieu de sa médiane, et ce type d'inégalité peut être obtenu de bien des manières (intégrale stochastique, méthodes d'espace gaussien, voir [Led]), mais aussi à partir de Prékopa-Leindler comme ci-dessous.

PROPOSITION 1.4. — *Pour toute fonction f réelle 1-lipschitzienne sur \mathbb{R}^n et pour tout nombre réel t , on a*

$$\int e^{tf(x)-tf(y)} d\gamma_n(x)d\gamma_n(y) \leq e^{t^2}.$$

Sous cette forme le résultat est optimal, puisque l'inégalité ci-dessus est une égalité pour les fonctions linéaires. À partir de cette estimation de transformée de Laplace, on obtient les inégalités de concentration par Markov, comme il est habituel.

Preuve. — Posons

$$g_t(x) = \min\{tf(x+h) + |h|^2/4 : h \in \mathbb{R}^n\}.$$

Le triplet de fonctions $\varphi_0(x) = -g_t(x) + |x|^2/2$, $\varphi_{1/2}(z) = |z|^2/2$, $\varphi_1(y) = tf(y) + |y|^2/2$ vérifie l'hypothèse de Prékopa-Leindler (H_{PL}), car pour tous $x, y \in \mathbb{R}^n$ on a

$$\varphi_0(x) + \varphi_1(y) \geq -tf(y) - |x - y|^2/4 + |x|^2/2 + tf(y) + |y|^2/2 = 2\varphi_{1/2}\left(\frac{x+y}{2}\right).$$

On en déduit que

$$\left(\int e^{g_t(x)} d\gamma_n(x)\right)\left(\int e^{-tf(y)} d\gamma_n(y)\right) \leq \left(\int d\gamma_n(z)\right)^2 = 1.$$

Comme $tf(x+h) \geq tf(x) - |th|$ par la condition de Lipschitz, et que $tf(x) - |th| + |h|^2/4 \geq tf(x) - t^2$ pour tout h , on a $tf(x) \leq g_t(x) + t^2$, donc

$$\int e^{tf(x)-tf(y)} d\gamma_n(x)d\gamma_n(y) = \left(\int e^{tf(x)} d\gamma_n(x) \right) \left(\int e^{-tf(y)} d\gamma_n(y) \right) \leq e^{t^2}.$$

□

Avec une preuve apparentée à la précédente mais nettement plus subtile, Sergueï Bobkov et Michel Ledoux ont obtenu une démonstration d'une inégalité Log-Sobolev pour γ_n ([BoL]).

1.3. Une inégalité de Brascamp-Lieb

Une inégalité de Brascamp et Lieb [BrL] a trouvé des applications nombreuses en convexité ; on va l'énoncer sous une forme particulière due à Keith Ball [Ba2]. On se place dans \mathbb{R}^n ou dans un espace euclidien E de dimension n . Désignons par $v \otimes v$ l'opérateur de rang un $x \rightarrow (x \cdot v)v$, et supposons que

$$(J) \quad \text{Id}_{\mathbb{R}^n} = \sum_{j=1}^N c_j v_j \otimes v_j,$$

avec des vecteurs v_j de norme un et des scalaires $c_j > 0$, qui satisfont $\sum_{j=1}^N c_j = n$ par un calcul de trace immédiat. Il en résulte que pour tous $x, y \in \mathbb{R}^n$

$$x \cdot y = \sum_{j=1}^N c_j (x \cdot v_j) (y \cdot v_j), \quad |x|^2 = \sum_{j=1}^N c_j (x \cdot v_j)^2.$$

Cette décomposition de l'identité est fortement liée à une notion géométrique importante, l'*ellipsoïde de John*, et à ses propriétés remarquables ; l'ellipsoïde de John pour un corps convexe symétrique C est l'ellipsoïde maximal contenu dans C ; Fritz John a montré que lorsque cet ellipsoïde est égal à la boule euclidienne unité B , l'identité de \mathbb{R}^n admet la décomposition (J) avec des v_j choisis parmi les points de contact de B et du bord de C .

THÉORÈME 1.5 (Brascamp-Lieb, version Ball). — *Si les vecteurs (v_j) et les nombres (c_j) vérifient la relation (J), on a pour toutes les fonctions positives (f_j) intégrables sur \mathbb{R}*

$$\int_{\mathbb{R}^n} \left(\prod_{j=1}^N f_j(x \cdot v_j)^{c_j} \right) dx \leq \prod_{j=1}^N \left(\int_{\mathbb{R}} f_j(t) dt \right)^{c_j}.$$

Le cas trivial est celui d'une base orthonormée v_1, \dots, v_n , avec des (c_j) égaux à 1 : on trouve simplement Fubini dans ce cas, et on a bien sûr égalité. La méthode utilisée par Brascamp et Lieb est assez compliquée, et consiste à montrer que, si on fixe $\int f_j = 1$, le maximum du terme de gauche est atteint pour des fonctions gaussiennes. On peut

remarquer que, si $f_j(t) = e^{-t^2}$, la relation

$$\prod_{j=1}^N f_j(x \cdot v_j)^{c_j} = \exp\left(-\sum_{j=1}^N c_j(x \cdot v_j)^2\right) = \exp\left(-\sum_{i=1}^n x_i^2\right)$$

montre qu'on a égalité dans Brascamp-Lieb. Le théorème 1.5 a été généralisé par Lieb [Lie] au cas où l'identité de \mathbb{R}^n est représentée sous la forme $\sum_{j=1}^N c_j P_j$, où les P_j sont des projecteurs orthogonaux sur \mathbb{R}^n .

Frank Barthe a trouvé une preuve simple du théorème 1.5, et sa preuve lui a permis de montrer en même temps une inégalité inverse, qui avait été conjecturée par Ball. Cette inégalité inverse a elle aussi des applications géométriques intéressantes.

THÉORÈME 1.6 (Barthe, [Bar]). — *Si les vecteurs (v_j) et les nombres (c_j) vérifient la relation (J), on a pour toutes les fonctions positives (f_j) intégrables sur \mathbb{R} et pour toute fonction mesurable F sur \mathbb{R}^n telle que*

$$F(x) \geq \prod_{j=1}^N f_j(a_j)^{c_j} \text{ chaque fois que } x = \sum_{j=1}^N a_j c_j v_j$$

l'inégalité

$$\int_{\mathbb{R}^n} F(x) dx \geq \prod_{j=1}^N \left(\int_{\mathbb{R}} f_j(t) dt \right)^{c_j}.$$

Si on applique le théorème précédent avec $n = 1$, $N = 2$, $v_1 = v_2 = 1$ et $c_1 = 1 - \theta$, $c_2 = \theta$, on retrouve Prékopa-Leindler sur \mathbb{R} (qui entraîne assez facilement le cas général par Fubini et itération).

En appliquant l'inégalité de Brascamp-Lieb, Ball a montré le résultat qui suit sur les sections des cubes N -dimensionnels. Le résultat n'est optimal que lorsque la dimension k de la section divise N , mais il donne toujours une information non triviale. Lorsque $k = N - 1$, on trouve une majoration des volumes des sections hyperplanes par $e^{1/2}$, alors que le résultat optimal (dû également à Ball, [Ba1]) est $\sqrt{2}$. Mentionnons la méthode d'unimodalité de Kanter, qui donne d'autres informations dans des problèmes de volumes de sections ([Kan], voir [MeP]). Le lecteur consultera avec profit le bel article de Ball [Ba4] dans le *Handbook of the Geometry of Banach spaces*.

THÉORÈME 1.7 (K. Ball, [Ba2]). — *Toutes les sections k -dimensionnelles d'un cube N -dimensionnel de volume 1 ont un volume (k -dimensionnel) majoré par*

$$\left(\frac{N}{k}\right)^{k/2}.$$

Preuve. — La démonstration va mettre en lumière la parfaite adéquation de la version de Ball de l'inégalité de Brascamp-Lieb. Prenons pour cube C de volume 1 l'ensemble $[-1/2, 1/2]^N$ dans \mathbb{R}^N ; soit $(e_j)_{j=1}^N$ la base canonique de \mathbb{R}^N , soit E un sous-espace de dimension k ; en projetant sur E l'identité évidente $\text{Id}_{\mathbb{R}^N} = \sum_{j=1}^N e_j \otimes e_j$, on obtient

$\text{Id}_E = \sum_{j=1}^N c_j v_j \otimes v_j$, où $c_j^{1/2} v_j$ est la projection orthogonale de e_j sur E et $|v_j| = 1$. Les points x de $E \cap C$ sont caractérisés par les inégalités $|x \cdot (c_j^{1/2} v_j)| \leq 1/2$, $j = 1, \dots, N$, donc l'indicatrice de $E \cap C$ est le produit des fonctions $f_j(x \cdot v_j)$, où $f_j(t) = \mathbf{1}_{[-1/2, 1/2]}(c_j^{1/2} t)$. D'après le théorème 1.5,

$$|E \cap C|_k = \int_E \prod_{j=1}^N f_j(x \cdot v_j)^{c_j} dx \leq \prod_{j=1}^N \left(\int_{-1/2}^{1/2} f_j(t) dt \right)^{c_j} = \prod_{j=1}^N c_j^{-c_j/2}.$$

Sachant que $\sum_{j=1}^N c_j = k$, l'expression est maximale quand les c_j sont égaux, ce qui donne le résultat. \square

2. TRANSPORT OPTIMAL

On trouve Gaspard Monge à l'origine lointaine de la théorie du transport optimal ; dans [Mon], il pose le problème du transport optimal d'une masse de terre vers un emplacement de même volume ; pour Monge, le coût du transport est mesuré dans L_1 , c'est-à-dire que l'élément de coût dc du transport d'un élément de masse dm placé au point x et transporté au point Tx est $dc = |x - Tx| dm$. La théorie devient beaucoup plus agréable lorsqu'on cherche à minimiser le coût quadratique $|x - Tx|^2 dm$: étant données deux mesures finies μ et ν de même masse sur \mathbb{R}^n , on cherche à réaliser

$$\min \left\{ \int_{\mathbb{R}^n} |x - Tx|^2 d\mu(x) \right\},$$

l'inf étant pris sur toutes les transformations T telles que $T(\mu) = \nu$ (l'image de μ par T est ν). Au milieu du 20ème siècle, Kantorovitch exprime la question du transport optimal comme un problème de programmation convexe ; on parle alors du problème de Monge-Kantorovitch. Cette théorie a une histoire riche qui ne sera pas évoquée ici.

À la fin des années 90, Yann Brenier [Bre] montre que, sous certaines hypothèses, le transport optimal T d'une mesure finie à densité $f(x) dx$ sur \mathbb{R}^n vers une autre mesure à densité $g(y) dy$ de même masse est donné par le gradient d'une fonction convexe u , fonction unique à une constante près. Robert McCann a assoupli les conditions d'existence.

THÉORÈME 2.1 (McCann [MC1]). — *Si μ et ν sont deux probabilités sur \mathbb{R}^n , et si μ ne charge aucun ensemble borélien de dimension de Hausdorff $n - 1$, il existe une fonction u convexe sur \mathbb{R}^n telle que l'image de μ par l'application μ -presque partout définie ∇u soit égale à ν . Si u_1 et u_2 sont deux solutions, la différence $\nabla u_1 - \nabla u_2$ est nulle μ -presque partout.*

Quelques commentaires s'imposent ; la fonction convexe u est à valeurs dans $\mathbb{R} \cup \{+\infty\}$; elle est continue dans l'intérieur de l'ensemble convexe $\text{dom}(u) = \{x : u(x) < +\infty\}$, et on sait qu'elle y est différentiable, sauf sur un sous-ensemble de dimension de Hausdorff $\leq n - 1$; pour pouvoir définir la mesure image ν en posant $\nu(B) = \mu((\nabla u)^{-1}(B))$ pour

tout borélien B , il suffit que ∇u soit défini μ -presque partout. Le résultat de McCann est très général, mais il oblige à travailler avec des notions de calcul différentiel généralisé, telles que le Hessien au sens d'Alexandrov. Une autre approche, initiée par Luis Caffarelli immédiatement après l'article de Brenier, est d'utiliser des théorèmes de régularité de solutions d'une équation de Monge-Ampère pour pouvoir affirmer que le transport de Brenier est régulier sous certaines conditions (on est content quand u est C^2 , ce qui donne un transport C^1).

En dimension un, il n'y a qu'une façon raisonnable de définir ce transport optimal : pour transporter la probabilité $f(x) dx$ sur la probabilité $g(y) dy$, on introduit la fonction croissante $T(t)$ définie par

$$(T_1) \quad \int_{-\infty}^t f(x) dx = \int_{-\infty}^{T(t)} g(y) dy.$$

La fonction convexe u n'apparaît pas naturellement, mais il suffit de prendre une primitive quelconque de la fonction croissante T pour obtenir u , et l'étude de la régularité de u est évidente. Passons au cas multi-dimensionnel ; notons $\nabla_x u$ le gradient de u au point x , et de même pour le hessien $H_x u$. Supposons que la mesure $f(x) dx$ soit envoyée sur $g(y) dy$ par le changement de variable régulier $y = \nabla_x u$. Alors

$$g(\nabla_x u) \det(H_x u) = f(x),$$

donc u satisfait l'équation de Monge-Ampère

$$\det H_x u = \frac{f(x)}{g(\nabla_x u)}.$$

Si on a une propriété minimale de régularité pour ∇u , un argument de bootstrap et les résultats de régularité connus pour l'équation de Monge-Ampère entraînent que u est régulière quand f et g le sont, sous une certaine condition géométrique sur les supports (Caffarelli, [Ca1]).

Faisons un commentaire élémentaire sur cette condition géométrique. Il est évident que la régularité de l'application u est rompue si on doit transporter, par exemple, la mesure de $[-1, 1]$ sur la mesure de l'ensemble non connexe $[-2, -1] \cup [1, 2]$; mais c'est la convexité de l'image qui est le véritable enjeu. Si on transporte une mesure dont le support est \mathbb{R}^n sur une autre mesure par le gradient d'une fonction convexe régulière u , l'application gradient donnera de \mathbb{R}^n une image dont l'intérieur est convexe. Ainsi, il n'est pas possible de transporter de façon régulière la mesure de Gauss γ_n sur la mesure de Lebesgue d'un ouvert non convexe, par un transport de la forme ∇u , avec u convexe. Les bons résultats de régularité sont obtenus lorsque la mesure image est portée par un ensemble convexe C , et qu'elle possède une densité g , régulière et bornée inférieurement sur C .

McCann [MC2] a généralisé le transport de Brenier au cas des variétés riemanniennes. Il s'agit encore d'un transport optimal pour le coût quadratique. Pour appréhender cette généralisation, il faut reformuler le transport de Brenier par un gradient ∇u en termes de

déplacement : à chaque point x , on associe le vecteur $V_x = \nabla_x u - x$; on obtient le point Tx en se déplaçant à partir de x sur la géodésique tangente à V_x , dans le sens donné par V_x , et d'une longueur égale à $|V_x|$; on peut voir ce vecteur V_x comme le gradient d'une nouvelle fonction $v(x) = u(x) - |x|^2/2$. C'est sous cette forme que le résultat est généralisé : *il existe une fonction v sur la variété telle que $Tx = \exp_x(\nabla_x v)$ pour μ -presque tout x* , mais nous n'entrerons pas dans les détails.

L'inégalité de Prékopa-Leindler a été généralisée dans [CMS] au cas des variétés riemanniennes, mais à ma connaissance, elle attend encore à ce jour des applications qui soient au niveau des applications de l'inégalité classique.

2.1. Quelques applications du transport

On va donner quelques exemples simples, où on mettra en avant la grande simplicité de l'idée, en laissant malhonnêtement de côté les précautions techniques nécessaires. On va aussi laisser de côté un grand nombre d'autres aspects, qu'on pourra trouver dans la monographie de Cédric Villani [Vil].

2.1.1. *Preuve de l'inégalité de Prékopa-Leindler par transport* (McCann). Supposons que le triplet (g_0, g_θ, g_1) satisfasse les hypothèses de Prékopa-Leindler (H_{PL2}), que $\int g_0 = \int g_1$ et que le gradient ∇u d'une fonction convexe transporte la mesure $g_0(x) dx$ sur la mesure $g_1(y) dy$. Alors, en effectuant le changement de variable $z = (1 - \theta)x + \theta \nabla_x u$, en utilisant l'équation de transport $g_1(\nabla_x u) \det H_x u = g_0(x)$, le fait que le Hessien de u est positif et la concavité de la fonction $A \rightarrow \ln \det A$ sur le convexe des matrices réelles symétriques définies positives, on obtiendra la chaîne d'inégalités

$$\begin{aligned} \int g_\theta(z) dz &\geq \int g_\theta((1 - \theta)x + \theta \nabla_x u) \det((1 - \theta)I_n + \theta H_x u) dx \geq \\ &\geq \int g_0(x)^{1-\theta} g_1(\nabla_x u)^\theta (\det H_x u)^\theta dx = \int g_0(x) dx = \left(\int g_0(x) dx \right)^{1-\theta} \left(\int g_1(y) dy \right)^\theta. \end{aligned}$$

Avec Knothe [Kno] on avait déjà en 1957 une preuve de Brunn-Minkowski par une sorte de transport, qui est certainement moins élégante que celle obtenue par le transport de Brenier, mais où les problèmes de régularité ne se posent pas. L'application de Knothe peut se décrire ainsi en dimension deux : on se donne deux probabilités $d\mu(x_1, x_2) = f(x_1, x_2) dx_1 dx_2$ et $d\nu(y_1, y_2) = g(y_1, y_2) dy_1 dy_2$; on transporte $f_1(x_1) dx_1$, où $f_1(x_1) = \int f(x_1, x_2) dx_2$ sur $g_1(y_1) dy_1$, définie de façon analogue, par un transport 1-dimensionnel T_1 ; on transporte ensuite la section en x_1 de μ , renormalisée en probabilité par produit avec $f_1(x_1)^{-1}$, sur la section en $T_1(x_1)$ de ν , renormalisée, par un transport 1-dimensionnel T_2 dépendant de x_1 ; la matrice jacobienne de l'application de Knothe est triangulaire.

2.1.2. *Preuve de l'inégalité de Sobolev par transport* (Gromov). Supposons que f soit une fonction ≥ 0 régulière à support compact sur \mathbb{R}^n , et que $\int f^{n/(n-1)} = |B_n|$, le volume de la boule unité de \mathbb{R}^n . La mesure $f(x)^{n/(n-1)} dx$ est envoyée sur la mesure de Lebesgue de

la boule unité de \mathbb{R}^n par une application de la forme ∇u , avec u convexe. Il en résulte que $|\nabla u| \leq 1$ sur le support de f . En utilisant l'équation de transport et l'inégalité arithmético-géométrique, on obtient

$$\begin{aligned} |B_n| &= \int f^{n/(n-1)} = \int f f^{1/(n-1)} = \int f (\det Hu)^{1/n} \leq \\ &\leq \frac{1}{n} \int f \Delta u = -\frac{1}{n} \int \nabla f \cdot \nabla u \leq \frac{1}{n} \int |\nabla f|. \end{aligned}$$

Si on rétablit l'homogénéité, on retrouve l'inégalité de Sobolev optimale

$$\left(\int_{\mathbb{R}^n} f^{n/(n-1)} \right)^{(n-1)/n} \leq \frac{1}{n|B_n|^{1/n}} \int_{\mathbb{R}^n} |\nabla f|.$$

On peut aussi obtenir par transport les inégalités de Sobolev L_p avec constantes optimales (voir [CNV]).

2.1.3. *Preuve de l'inégalité de Brascamp-Lieb par transport.* La preuve de Barthe pour le théorème 1.5 utilise le transport dans son cas élémentaire, celui de la dimension un, mais Barthe montre aussi dans [Bar] l'inégalité plus générale de Lieb, qui demande d'utiliser le transport en plusieurs dimensions. On considère $g(t) = e^{-t^2}$, on suppose que $\int f_j = \int g$ pour chaque $j = 1, \dots, N$ et on suppose aussi $f_j > 0$. On définit $T_j(t)$ par l'équation

$$\int_{-\infty}^t f_j(x) dx = \int_{-\infty}^{T_j(t)} g(y) dy;$$

on a $f_j(t) = g(T_j(t)) T_j'(t)$. Désignons par u_j la fonction (strictement) convexe sur \mathbb{R} telle que $u_j' = T_j$ et $u_j(0) = 0$. On définit une fonction convexe u sur \mathbb{R}^n par

$$\forall x \in \mathbb{R}^n, \quad u(x) = \sum_{j=1}^N c_j u_j(x \cdot v_j),$$

où les (v_j) , (c_j) vérifient la relation (J). On va s'intéresser à la transformation T de \mathbb{R}^n définie par

$$x \in \mathbb{R}^n \rightarrow Tx = \sum_{j=1}^N c_j T_j(x \cdot v_j) v_j = \nabla_x u$$

et l'utiliser comme changement de variable dans \mathbb{R}^n ; en effet, la stricte convexité de u et sa croissance à l'infini garantissent que ∇u est bijective sur \mathbb{R}^n ; la matrice jacobienne de T au point x est

$$T'x = \sum_{j=1}^N c_j T_j'(x \cdot v_j) v_j \otimes v_j;$$

on utilisera la majoration

$$(2) \quad \left| \sum_{j=1}^N a_j c_j v_j \right|^2 \leq \sum_{j=1}^N c_j a_j^2$$

et l'estimation

$$(3) \quad \det\left(\sum_{j=1}^N a_j c_j v_j \otimes v_j\right) \geq \prod_{j=1}^N a_j^{c_j},$$

où les (a_j) sont des réels positifs. On aura donc en posant $\xi_j = x \cdot v_j$ pour $x \in \mathbb{R}^n$, en utilisant les équations pour les transports T_j et les inégalités (2) et (3)

$$\begin{aligned} \int_{\mathbb{R}^n} \prod_{j=1}^N f_j(\xi_j)^{c_j} dx &= \int_{\mathbb{R}^n} \prod_{j=1}^N g(T_j(\xi_j))^{c_j} \prod_{j=1}^N T'_j(\xi_j)^{c_j} dx = \\ &= \int_{\mathbb{R}^n} \exp\left(-\sum_{j=1}^N c_j (T_j(\xi_j))^2\right) \prod_{j=1}^N T'_j(\xi_j)^{c_j} dx \leq \\ &\leq \int_{\mathbb{R}^n} \exp(-|Tx|^2) \det\left(\sum_{j=1}^N T'_j(\xi_j) c_j v_j \otimes v_j\right) dx = \\ &= \int_{\mathbb{R}^n} e^{-|Tx|^2} \det(T'x) dx = \int_{\mathbb{R}^n} e^{-|y|^2} dy = \left(\int_{\mathbb{R}} g\right)^n = \prod_{j=1}^N \left(\int_{\mathbb{R}} f_j\right)^{c_j}. \end{aligned}$$

Revenons aux inégalités (2) et (3) ; l'inégalité (2) résulte du fait que les vecteurs $c_j^{1/2} v_j$ peuvent être vus comme les projections d'une base orthonormée w_1, \dots, w_N de \mathbb{R}^N : si $(w_{j,k})_{k=1}^n$ désigne les coordonnées de $c_j^{1/2} v_j$, la relation (J) montre que les colonnes $(w_{j,k})_{j=1}^N$, $k = 1, \dots, n$ sont orthonormées, et on peut par conséquent les compléter en une matrice orthogonale qui fournit les vecteurs w_1, \dots, w_N .

Prouvons (3) : si $(e_i)_{i=1}^n$ est une base orthonormée de \mathbb{R}^n , si Det désigne le déterminant de n vecteurs par rapport à cette base et si $M = \sum_{j=1}^N M_j$ est une somme de matrices $n \times n$ de rang 1, on voit que $\det M$ est égal à

$$\text{Det}\left(\sum_{j=1}^N M_j e_1, \dots, \sum_{j=1}^N M_j e_n\right) = \sum_{j_1, \dots, j_n=1}^N \text{Det}(M_{j_1} e_1, \dots, M_{j_n} e_n) = \sum_I \det M_I$$

où la somme est étendue à tous les sous-ensembles I à n éléments de $\{1, \dots, N\}$, et où on a posé $M_I = \sum_{j \in I} M_j$; comme les M_j sont de rang 1, les j_1, \dots, j_n qui donnent une contribution non nulle dans le terme central sont des entiers deux à deux distincts ; on obtient la dernière égalité en regroupant selon les valeurs de $I = \{j_1, \dots, j_n\}$. Si $A = \sum_{j=1}^N a_j M_j$, le même développement montre que $\det A = \sum_I a^I \det M_I$, où on a posé $a^I = \prod_{j \in I} a_j$; si $M_j = c_j v_j \otimes v_j$, on a $M = I_n$ donc $\sum_I \det M_I = 1$. L'inégalité arithmético-géométrique donne alors

$$(4) \quad \det A \geq \prod_I (a^I)^{\det M_I} = \prod_{j=1}^N a_j^{\sum_{I: j \in I} \det M_I}.$$

Mais $\sum_{j \neq i} c_j v_j \otimes v_j$ est égal à $I_n - c_i v_i \otimes v_i$, dont le déterminant est

$$\sum_{I: i \notin I} \det M_I = 1 - c_i,$$

ce qui montre que $\sum_{I: i \in I} \det M_I = c_i$ pour tout $i = 1, \dots, N$, et donne le résultat voulu (3) à partir de l'inégalité (4).

2.1.4. *Questions gaussiennes.* Caffarelli [Ca2] a remarqué que le transport de Brenier est contractant quand on transporte la probabilité gaussienne γ_n sur une probabilité μ de la forme $e^{-|x|^2/2 - \varphi(x)} dx$, φ convexe. Le caractère contractant de T peut être utilisé pour transporter certaines inégalités vraies pour γ_n en inégalités pour μ . On peut aussi utiliser ce résultat dans le problème de la corrélation gaussienne ([Co1], voir aussi [Ha2] pour un résultat connexe). La conjecture de la *corrélation gaussienne* prévoit que

$$\gamma_n(A \cap B) \geq \gamma_n(A) \gamma_n(B)$$

lorsque A, B sont deux convexes symétriques dans \mathbb{R}^n . Elle est prouvée dans \mathbb{R}^2 [Pit], mais seuls quelques cas particuliers sont connus en dimension plus grande (voir par exemple Gilles Hargé [Ha1]).

3. QUELQUES RÉSULTATS EN GÉOMÉTRIE GAUSSIENNE

L'inégalité isopérimétrique gaussienne du théorème 1.3 peut se montrer à partir de l'isopérimétrie sur la sphère (Borell, Sudakov-Tsirelson), par des techniques de symétrisation (Ehrhard) ou par des méthodes de semi-groupes. Posons pour tout x réel

$$\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2} \quad \text{et} \quad \Phi(x) = \int_{-\infty}^x d\gamma_1 = \int_{-\infty}^x \varphi(t) dt;$$

la *fonction isopérimétrique gaussienne* I est définie par $I(t) = \varphi \circ \Phi^{-1}(t)$ pour tout $t \in [0, 1]$, c'est-à-dire qu'on a $I(t) = (2\pi)^{-1/2} e^{-s^2/2}$ exactement quand $t = \int_{-\infty}^s d\gamma_1$. Le résultat du théorème 1.3 se reformule ainsi : pour tout fermé $A \subset \mathbb{R}^n$, on a

$$(5) \quad \gamma_n^+(A) := \liminf_{\varepsilon \rightarrow 0^+} \varepsilon^{-1} (\gamma_n(A_\varepsilon) - \gamma_n(A)) \geq I(\gamma_n(A)),$$

où A_ε est le ε -épaississement de A pour la distance euclidienne. Autrement dit, si on définit $s \in \mathbb{R}$ par l'équation $\gamma_n(A) = \int_{-\infty}^s d\gamma_1$, on a $\gamma_n^+(A) \geq (2\pi)^{-1/2} e^{-s^2/2}$. Il est clair que cette inégalité est une égalité pour les demi-espaces affines de \mathbb{R}^n .

Bobkov [Bob] a montré une forme fonctionnelle pour l'isopérimétrie gaussienne : pour toute fonction f localement lipschitzienne de \mathbb{R}^n dans $[0, 1]$, on a

$$(6) \quad I\left(\int_{\mathbb{R}^n} f d\gamma_n\right) \leq \int_{\mathbb{R}^n} \sqrt{I^2(f) + |\nabla f|^2} d\gamma_n.$$

On voit que ce résultat contient l'information isopérimétrique précédente en prenant pour f une fonction égale à 1 sur A , et décroissant vers 0 autour de A , pour s'annuler hors de A_ε . Bobkov déduit ce résultat d'une *inégalité à deux points* : pour tous $a, b \in [0, 1]$,

$$(7) \quad I\left(\frac{a+b}{2}\right) \leq \frac{1}{2}\sqrt{I^2(a) + \left(\frac{b-a}{2}\right)^2} + \frac{1}{2}\sqrt{I^2(b) + \left(\frac{b-a}{2}\right)^2}.$$

Utilisant les propriétés de tensorisation de cette inégalité et le théorème central limite, Bobkov montre que (7) implique (6). Comme il est noté dans [Bob], l'inégalité (6) pour \mathbb{R}^n peut aussi se déduire de (5) pour \mathbb{R}^{n+1} en choisissant pour $A \subset \mathbb{R}^n \times \mathbb{R}$ le sous-graphe de $\Phi^{-1} \circ f$; en fait cette remarque est déjà chez Ehrhard, mais le point remarquable de l'article [Bob] est que (6) y est déduite de l'inégalité élémentaire pour deux points ; on peut aussi transcrire la preuve de Bobkov dans le langage des semi-groupes ou des martingales browniennes (voir [CHL]).

Ehrhard a montré dans [Ehr] un renforcement de l'inégalité de Brunn-Minkowski gaussienne (1) : si A, B sont deux ensembles convexes fermés non vides dans \mathbb{R}^n , on a

$$\Phi^{-1}(\gamma_n((1-t)A + tB)) \geq (1-t)\Phi^{-1}(\gamma_n(A)) + t\Phi^{-1}(\gamma_n(B))$$

pour tout $t \in [0, 1]$. Cette propriété est plus forte que celle de l'inégalité (1), car pour tous $a, b \in [0, 1]$ on a $(1-t)\Phi^{-1}(a) + t\Phi^{-1}(b) \geq \Phi^{-1}(a^{1-t}b^t)$; cette dernière inégalité provient par exemple de l'inégalité de Brunn-Minkowski gaussienne (1) dans \mathbb{R} , appliqué aux deux intervalles $]-\infty, \Phi^{-1}(a)]$ et $]-\infty, \Phi^{-1}(b)]$. Donnons une interprétation plus géométrique de l'inégalité d'Ehrhard : si H_A et H_B sont deux demi-espaces affines limités par deux hyperplans parallèles, et si $\gamma_n(H_A) = \gamma_n(A)$, $\gamma_n(H_B) = \gamma_n(B)$, la mesure gaussienne de $(1-t)A + tB$ est au moins celle du demi-espace $(1-t)H_A + tH_B$.

La question de savoir si la convexité de A et B est nécessaire dans l'inégalité d'Ehrhard est restée longtemps ouverte. Rafał Latała [Lat] a généralisé au cas où un seul des ensembles A, B est convexe et très récemment Borell [Bo2] a étendu le résultat à deux fermés quelconques ; sa méthode se formule en termes de semi-groupes ou d'équation d'évolution. Elle aura probablement d'autres retombées : on a déjà pu montrer [BaC] qu'elle donne une preuve stochastique de l'inégalité Brascamp-Lieb inverse (théorème 1.6).

Les mesures des dilatés d'un convexe symétrique C ont également un comportement intéressant ; il résulte de Prékopa-Leindler que $t \rightarrow \ln \gamma_n(tC)$ est concave sur $]0, +\infty[$. La *B-conjecture*, démontrée récemment [CFM], dit plus : la fonction $t \rightarrow \ln \gamma_n(e^t C)$ est concave sur \mathbb{R} . Latała et Oleszkiewicz [LaO] ont montré la *S-conjecture*, dont la seule preuve connue à ce jour est très compliquée ; elle emprunte dans son début des idées d'Ehrhard. Le résultat est le suivant : si on envisage l'accroissement de volume gaussien produit en passant d'un convexe symétrique C à un homothétique tC , $t > 1$, l'accroissement est minimal pour une bande symétrique $B = \{x \in \mathbb{R}^n : |x_1| \leq b\}$ telle que $\gamma_n(B) = \gamma_n(C)$,

$$\forall t \geq 1, \quad \gamma_n(tC) \geq \gamma_n(tB).$$

4. DÉVELOPPEMENTS PLUS RÉCENTS AUTOUR DE PRÉKOPA

Ball, Barthe et Naor [BBN] ont récemment donné une expression utile pour la dérivée seconde en 0 de la fonction Φ définie par

$$e^{-\Phi(t)} = \int_{\mathbb{R}^n} e^{-\varphi(x,t)} dx$$

sur un intervalle réel contenant 0. Ils montrent (dans un langage équivalent) le résultat suivant : désignons par $\mathcal{H}_y\varphi$ la forme quadratique sur $\mathbb{R}^n \times \mathbb{R}$ définie par le Hessien de φ au point $\mathbf{y} = (y, 0)$ et par $H_y v$ la matrice hessienne au point y d'une fonction v définie sur \mathbb{R}^n ; on a

$$(8) \quad e^{-\Phi(0)} \Phi''(0) = \inf \left\{ \int_{\mathbb{R}^n} e^{-\varphi(y,0)} \left((\mathcal{H}_y\varphi)(\nabla_y v, 1) + \text{tr}((H_y v)^2) \right) dy \right\},$$

où l'inf est pris sur l'ensemble des fonctions v .

Si φ est convexe du couple (x, t) , le résultat sera bien sûr ≥ 0 , ce qui redonne le résultat de Prékopa. On a déjà pu trouver une application de cette formule, où la présence du terme positif $\text{tr}((H_y v)^2)$ permet d'obtenir la convexité de Φ dans des cas où φ n'est pas convexe [CFM].

Artstein, Ball, Barthe et Naor [ABBN] utilisent cet ingrédient pour montrer la croissance avec n de l'entropie de la suite des variables $Z_n = n^{-1/2}(X_1 + \dots + X_n)$ qui apparaissent dans le théorème central limite (les (X_i) sont des variables aléatoires indépendantes de carré intégrable, de même loi). Auparavant, on ne savait établir la croissance de l'entropie que sur la sous-suite (Z_{2^n}) .

Une approche heuristique pour le résultat sur $\Phi''(0)$ est la suivante : pour t tendant vers 0, on peut introduire la fonction convexe u_t sur \mathbb{R}^n , nulle en 0, dont le gradient transporte la probabilité $e^{\Phi(0)} e^{-\varphi(x,0)} dx$ sur $e^{\Phi(t)} e^{-\varphi(x,t)} dx$; la fonction $v = \dot{u}_0$, dérivée par rapport à t au point $t = 0$, est un bon candidat pour minimiser l'expression ci-dessus. En fait, cette approche est difficile à mettre en œuvre, et il est plus facile de traiter directement l'équation obtenue comme limite de l'équation de Monge-Ampère satisfaite par u_t , ce qui conduit à un opérateur linéaire du type Ornstein-Uhlenbeck.

Supposons que $\Phi(0) = 0$, de sorte que $d\mu(x) = e^{-\varphi(x,0)} dx$ soit une probabilité, et supposons qu'on puisse dériver deux fois sous le signe somme qui définit Φ . Le calcul direct de la dérivée seconde $\ddot{\Phi}(0)$ donne

$$(9) \quad \ddot{\Phi}(0) = \left(\int \dot{\varphi}_0 d\mu \right)^2 + \int (\ddot{\varphi}_0 - \dot{\varphi}_0^2) d\mu,$$

où $\dot{\varphi}_0$ représente la fonction $x \rightarrow \dot{\varphi}(x, 0)$; de même $\ddot{\varphi}_0, \varphi_0$ représentent dans ce qui suit les fonctions sur \mathbb{R}^n obtenues en fixant $t = 0$. La proposition qui suit ne prétend pas donner le résultat le plus général. On remarquera que si on veut prouver la convexité de Φ dans un intervalle autour de 0, on peut se permettre de modifier φ de façon à vérifier

les hypothèses ci-dessous, et d'obtenir Φ comme limite simple de fonctions convexes, si les fonctions modifiées $\tilde{\varphi}$ donnent des $\tilde{\Phi}$ à dérivée seconde positive.

PROPOSITION 4.1. — On suppose que φ_0 est de classe C^2 sur \mathbb{R}^n , que $\dot{\varphi}_0 \in L_2(\mu)$ et $\ddot{\varphi}_0 \in L_1(\mu)$, et que $\ddot{\Phi}(0)$ est donné par la formule (9). Dans ce cas, $\ddot{\Phi}(0)$ est donné aussi par la formule (8), où l'inf est pris sur l'ensemble des fonctions v de classe C^2 à support compact.

Preuve. — Considérons l'opérateur différentiel $L = \Delta - (\nabla\varphi_0) \cdot \nabla$, agissant sur l'espace C_{comp}^2 des fonctions à support compact de classe C^2 sur \mathbb{R}^n . Il est bien connu et facile à vérifier que la transformation isométrique U de $L_2(\mu)$ sur $L_2(\mathbb{R}^n)$ définie par $Uf = e^{-\varphi_0/2} f$ transforme l'opérateur $-L$ en l'opérateur $S = -\Delta + V$, avec $V = \frac{1}{4}|\nabla\varphi_0|^2 - \frac{1}{2}\Delta\varphi_0$; si $g = Uf$, on voit facilement que

$$\langle -Lf, f \rangle = \int |\nabla f|^2 d\mu = \int_{\mathbb{R}^n} \left(|\nabla_x g|^2 + V(x) |g(x)|^2 \right) dx = \langle -\Delta g + Vg, g \rangle.$$

On va montrer que l'image par L de C_{comp}^2 est dense dans le sous-espace H_0 de $L_2(\mu)$ formé des fonctions h telles que $\int h d\mu = 0$. Sinon, il existerait une fonction $f \in H_0$ non nulle orthogonale à toutes les Lh , ce qui implique que $g = Uf$ est orthogonale à toutes les $-\Delta\psi + V\psi$, $\psi \in \mathcal{D}(\mathbb{R}^n)$. On a donc $\Delta g = Vg$ dans $\mathcal{D}'(\mathbb{R}^n)$, ce qui entraîne que $g \in H_{\text{loc}}^2(\mathbb{R}^n)$ par la théorie classique. Si $\theta \in \mathcal{D}(\mathbb{R}^n)$, on a $\theta g \in H^2(\mathbb{R}^n)$ et on note (en utilisant la nullité de l'intégrale de la divergence de $\theta^2 g \nabla g$) que

$$\int_{\mathbb{R}^n} \left(|\nabla(\theta g)|^2 + \theta^2 g \Delta g \right) = \int_{\mathbb{R}^n} |\nabla\theta|^2 g^2.$$

En supposant que $\theta = 1$ dans un voisinage de 0, en introduisant $\theta_k(x) = \theta(x/k)$ pour tout $k \geq 1$, et en utilisant la relation $\Delta g = Vg$, on trouve que

$$\int_{\mathbb{R}^n} \left(|\nabla(\theta_k g)|^2 + V |\theta_k g|^2 \right) = \int_{\mathbb{R}^n} |\nabla\theta_k|^2 g^2 \xrightarrow[k]{} 0,$$

ce qui implique en revenant à f que $\int |\nabla f|^2 d\mu = 0$; la fonction f est par conséquent constante, donc nulle puisqu'elle est d'intégrale nulle. On atteint une contradiction qui montre que l'image de L est dense dans H_0 (on pourra trouver l'origine du calcul précédent chez Christian Simader [Sim]). On en déduit que

$$\left(\int \dot{\varphi}_0 d\mu \right)^2 = \min \left\{ \int (h - \dot{\varphi}_0)^2 d\mu : h \in H_0 \right\} = \inf_v \left\{ \int (Lv - \dot{\varphi}_0)^2 d\mu \right\},$$

où v varie dans C_{comp}^2 . On a déjà observé que $-\int (Lv) \dot{\varphi}_0 d\mu = \int \nabla v \cdot \nabla \dot{\varphi}_0 d\mu$; en intégrant par parties, et avec la relation formelle $\nabla L = L\nabla - \nabla^2\varphi_0 \cdot \nabla$, on trouve

$$\int (Lv)^2 d\mu = - \int (\nabla Lv) \cdot \nabla v d\mu = \int (\nabla^2 v \cdot \nabla^2 v + (\nabla^2\varphi_0 \cdot \nabla v) \cdot \nabla v) d\mu.$$

On voit que $\nabla^2 v \cdot \nabla^2 v$ est le carré de la norme Hilbert-Schmidt du Hessien Hv de v , égal à $\text{tr}((Hv)^2)$. Ainsi, on a montré que $\ddot{\Phi}(0)$ est l'inf en v de

$$\int (Lv - \dot{\varphi}_0)^2 d\mu + \int (\ddot{\varphi}_0 - \dot{\varphi}_0^2) d\mu = \int ((Lv)^2 - 2(Lv)\dot{\varphi}_0 + \ddot{\varphi}_0) d\mu =$$

$$\int ((\nabla^2 \varphi_0 \cdot \nabla v) \cdot \nabla v + 2 \nabla \dot{\varphi}_0 \cdot \nabla v + \ddot{\varphi}_0 + \text{tr}((Hv)^2)) d\mu,$$

qui coïncide avec l'expression attendue. □

4.1. Le cas complexe

Les ensembles pseudo-convexes jouent dans une certaine mesure en analyse complexe le rôle des ensembles convexes, et les fonctions pluri-sous-harmoniques celui des fonctions convexes. On pourrait se demander si la fonction Φ définie sur \mathbb{C} par

$$e^{-\Phi(z)} = \int_{\mathbb{C}^n} e^{-\varphi(x,z)} dx$$

est sous-harmonique lorsque φ est pluri-sous-harmonique sur $\mathbb{C}^n \times \mathbb{C}$, mais on sait depuis longtemps que cet énoncé trop général n'est pas vrai. Bo Berndtsson [Ber] a trouvé certaines conditions qui garantissent néanmoins ce résultat, et Dario Cordero [Co2] a remarqué que ce résultat de Berndtsson s'applique à certaines quantités géométriques liées à l'interpolation complexe en dimension finie. Par exemple, le volume de la boule unité de l'espace A_θ de la théorie de Calderón est log-concave par rapport à $\theta \in [0, 1]$. Ce résultat donne une preuve de l'inégalité de Santaló dans le cas d'espaces normés complexes.

L'inégalité de Santaló dit que le produit des volumes $|C| |C^\circ|$ d'un convexe symétrique C de \mathbb{R}^n et de son polaire C° est maximal pour la boule euclidienne. Si on se place dans \mathbb{C}^n , si C est la boule unité de l'espace normé complexe $A_0 = X = (\mathbb{C}^n, \|\cdot\|)$, on sait que la boule unité de X^* est le polaire C° , et que l'interpolé de Calderón $(X, X^*)_{1/2}$ est isométrique à \mathbb{C}^n muni du produit scalaire usuel. Le résultat précédent donne donc Santaló dans le cas de boules unité complexes, mais il donne aussi un résultat pour les volumes calculés avec certaines mesures log-concaves.

BIBLIOGRAPHIE

- [And] T. W. Anderson, *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Amer. Math. Soc. **6** (1955), 170–176.
- [ABBN] S. Artstein, K. Ball, F. Barthe et A. Naor, *Solution of Shannon's problem on the monotonicity of entropy*, preprint (2002).
- [Ba1] K. Ball, *Cube slicing in \mathbb{R}^n* , Proc. Amer. Math. Soc. **97** (1986), 465–473.

- [Ba2] K. Ball, *Volumes of sections of cubes and related problems*, Geometric aspects of functional analysis (1987–88), 251–260, Lecture Notes in Math. 1376, Springer.
- [Ba3] K. Ball, *Volume ratios and a reverse isoperimetric inequality*, J. London Math. Soc. **44** (1991), 351–359.
- [Ba4] K. Ball, *Convex geometry and functional analysis*, Handbook of the Geometry of Banach spaces, vol 1, 161–194, North Holland 2001.
- [BBN] K. Ball, F. Barthe et A. Naor, *Entropy jumps in the presence of a spectral gap*, Duke Math. J. **119** (2003), 41–63.
- [Bar] F. Barthe, *On a reverse form of the Brascamp-Lieb inequality*, Invent. Math. **134** (1998), 335–361.
- [BaC] F. Barthe et D. Cordero-Erausquin, *Inverse Brascamp-Lieb inequalities along the Heat equation*, à paraître dans GAFA seminar.
- [Ber] B. Berndtsson, *Prékopa’s theorem and Kiselman’s minimum principle for plurisubharmonic functions*, Math. Ann. **312** (1998), 785–792.
- [Bob] S. G. Bobkov, *An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space*, Ann. Proba. **25** (1997), 206–214.
- [BoL] S. G. Bobkov et M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal. **10** (2000), 1028–1052.
- [Bo1] C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), 207–216.
- [Bo2] C. Borell, *The Ehrhard inequality*, C. R. Acad. Sci. Paris **337** (2003), 663–666.
- [BrL] H. J. Brascamp et E. H. Lieb, *Best constants in Young’s inequality, its converse, and its generalization to more than three functions*, Advances in Math. **20** (1976), 151–173.
- [Bre] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), 375–417.
- [Ca1] L. A. Caffarelli, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc. **5** (1992), 99–104.
- [Ca2] L. A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities*, Comm. Math. Phys. **214** (2000), 547–563. Erratum. Comm. Math. Phys. **225** (2002), 449–450.
- [CHL] M. Capitaine, E. P. Hsu et M. Ledoux, *Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces*, Electron. Comm. in Proba. **2** (1997), 71–81.
- [Co1] D. Cordero-Erausquin, *Some applications of mass transport to Gaussian-type inequalities*, Arch. Ration. Mech. Anal. **161** (2002), 257–269.
- [Co2] D. Cordero-Erausquin, *Santaló’s inequality on \mathbb{C}^n by complex interpolation*, C. R. Acad. Sci. Paris **334** (2002), 767–772.

- [CFM] D. Cordero-Erausquin, M. Fradelizi et B. Maurey, *The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems*, à paraître dans *J. Funct. Anal.* (2004).
- [CMS] D. Cordero-Erausquin, R. McCann et M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, *Invent. Math.* **146** (2001), 219–257.
- [CNV] D. Cordero-Erausquin, B. Nazaret et C. Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities*, *Advances in Math.* **182** (2004), 307–332.
- [DaG] S. Das Gupta, *Brunn-Minkowski inequality and its aftermath*, *J. Multivariate Anal.* **10** (1980), 296–318.
- [Ehr] A. Ehrhard, *Symétrisation dans l'espace de Gauss*, *Math. Scand.* **53** (1983), 281–301.
- [Gar] R. J. Gardner, *The Brunn-Minkowski inequality*, *Bull. Amer. Math. Soc.* **39** (2002), 355–405.
- [HaO] H. Hadwiger et D. Ohmann, *Brunn-Minkowskischer Satz und Isoperimetrie*, *Math. Z.* **66** (1956), 1–8.
- [Ha1] G. Hargé, *A particular case of correlation inequality for the Gaussian measure*, *Ann. Probab.* **27** (1999), 1939–1951.
- [Ha2] G. Hargé, *Inequalities for the Gaussian measure and an application to Wiener space*, *C. R. Acad. Sci. Paris* **333** (2001), 791–794.
- [HeM] R. Henstock et A. M. Macbeath, *On the measure of sum-sets. I. The theorems of Brunn, Minkowski, and Lusternik*, *Proc. London Math. Soc.* **3** (1953), 182–194.
- [Kan] M. Kanter, *Unimodality and dominance for symmetric random vectors*, *Trans. Amer. Math. Soc.* **229** (1977), 65–85.
- [Kno] H. Knothe, *Contributions to the theory of convex bodies*, *Michigan Math. J.* **4** (1957), 39–52.
- [Lat] R. Latała, *A note on the Ehrhard inequality*, *Studia Math.* **118** (1996), 169–174.
- [LaO] R. Latała et K. Oleszkiewicz, *Gaussian measures of dilations of convex symmetric sets*, *Ann. Probab.* **27** (1999), 1922–1938.
- [Led] M. Ledoux, *The concentration of measure phenomenon*, American Mathematical Society, 2001.
- [Lei] L. Leindler, *On a certain converse of Hölder's inequality*, *Acta Sci. Math.* **33** (1972), 217–223.
- [Lie] E. H. Lieb, *Gaussian kernels have only Gaussian maximizers*, *Invent. Math.* **102** (1990), 179–208.
- [Lus] L. Lusternik, *Die Brunn-Minkowskische Ungleichung für beliebige messbare Mengen*, *Doklady Akad. SSSR*, 1935, No.3, 55–58.

- [MC1] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. **80** (1995), 309–323.
- [MC2] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.
- [MeP] M. Meyer et A. Pajor, *Sections of the unit ball of ℓ_n^p* , J. Funct. Anal. **80** (1988), 109–123.
- [Mon] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pages 666–704, 1781.
- [Pis] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics 94, Cambridge University Press, 1989.
- [Pit] L. D. Pitt, *A Gaussian correlation inequality for symmetric convex sets*, Ann. Probability **5** (1977), 470–474.
- [Pre] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. **34** (1973), 335–343.
- [Sim] C. G. Simader, *Essential self-adjointness of Schrödinger operators bounded from below*, Math. Z. **159** (1978), 47–50.
- [SuT] V. N. Sudakov et B. S. Tsirel'son, *Extremal properties of half-spaces for spherically invariant measures*, J. Soviet Math. **9** (1978), 9–18 ; traduit de Zap. Nauch. Sem. L.O.M.I. **41** (1974), 14–24.
- [Vil] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics **58** Amer. Math. Soc. 2003.

Bernard MAUREY

Université Paris 7

et

Laboratoire d'Analyse et Mathématiques Appliquées

UMR 8050 du CNRS

Université de Marne la Vallée

Cité Descartes – Champs sur Marne

F-77454 MARNE LA VALLÉE Cedex 2

E-mail : maurey@math.univ-mlv.fr

MOTIFS DE DIMENSION FINIE
[d'après S.-I. Kimura, P. O'Sullivan,...]

par Yves ANDRÉ

TABLE DES MATIÈRES

1.	Introduction : les groupes de Chow sont-ils « de dimension finie » ?	2
1.1.	Contre	2
1.2.	Pour	3
2.	Motifs purs	4
2.1.	Définitions	4
2.2.	Exemples	6
2.3.	Équivalences adéquates et \otimes -idéaux	7
2.4.	Le théorème de semi-simplicité de Jannsen	8
2.5.	Motifs de dimension finie. La conjecture de Kimura-O'Sullivan	9
2.6.	La conjecture de nilpotence de Voevodsky	10
2.7.	Filtration sur les groupes de Chow. La conjecture de Bloch-Beilinson	11
3.	Catégories de Kimura-O'Sullivan	12
3.1.	Foncteurs de Schur	12
3.2.	Objets pairs et objets impairs	12
3.3.	Propriétés de \otimes -nilpotence	15
3.4.	Propriétés de nilpotence	16
3.5.	Le théorème de scindage	19
3.6.	La stratégie de O'Sullivan	19
3.7.	Objets de dimension finie et super-fibrés vectoriels équivariants	21
4.	Applications	23
4.1.	Motifs de surfaces	23
4.2.	Motifs sur les corps finis	24
4.3.	Rationalité de fonctions zêta motiviques	26
4.4.	Construction inconditionnelle des groupes de Galois motiviques	27

4.5. Motifs de Chow vus comme super-fibrés vectoriels équivariants sous le groupe de Galois motivique. Hauteurs	28
RÉFÉRENCES	29

1. INTRODUCTION : LES GROUPES DE CHOW SONT-ILS « DE DIMENSION FINIE » ?

1.1. Contre

Soit X une variété projective lisse connexe de dimension d sur un corps k algébriquement clos. Soit $Z^*(X)$ le groupe des cycles algébriques sur X (groupe abélien libre engendré par les sous-variétés irréductibles de X), gradué par la codimension. Le groupe de Chow $CH^*(X)$ est le quotient de $Z^*(X)$ par l'équivalence rationnelle.

En codimension 1, sa structure est bien connue (le cas $d = 1$ et $k = \mathbb{C}$ remonte au 19ème siècle) : on a une suite exacte

$$0 \rightarrow CH^1(X)_0 \rightarrow CH^1(X) \rightarrow NS(X) \rightarrow 0$$

où $NS(X)$ est un groupe abélien de type fini (dit de Néron-Severi), ainsi qu'un isomorphisme (dit d'Abel-Jacobi)

$$AJ_X^1 : CH^1(X)_0 \cong Pic_X^0(k),$$

où Pic_X^0 est la variété abélienne de Picard attachée à X .

En codimension > 1 , l'espoir de comprendre ces groupes par dévissage en une partie discrète dénombrable et une partie continue contrôlée par les points d'une variété algébrique a donné naissance à la construction de théories de variétés de Picard et d'applications d'Abel-Jacobi intermédiaires.

À la fin des années 60, cet espoir s'est écroulé, même dans le cas le mieux compris après celui de la codimension 1, à savoir le cas de codimension maximale d . Dans ce cas, l'application d'Abel-Jacobi s'écrit

$$AJ_X^d : CH^d(X)_0 \rightarrow Alb_X(k),$$

où $CH^d(X)_0$ est le groupe des 0-cycles de degré 0, et Alb_X la variété abélienne d'Albanese attachée à X . Cette application est surjective.

En 1969, D. Mumford [37] a démontré que si X est une surface sur $k = \mathbb{C}$, AJ_X^2 n'est injective que si le genre géométrique $p_g(X)$ est nul (*i.e.* que si $H^2(X)$ est engendré par les classes de diviseurs). Par exemple, si X est le carré d'une courbe Y de genre ≥ 2 (d'où $p_g(X) \neq 0$), et en notant δ_Y l'application diagonale de Y , le 0-cycle $[D \times D] - \deg D \cdot (\delta_Y)_*[D]$ sur X est dans le noyau de AJ_X^2 pour tout diviseur D sur Y ; dans un article récent [12], M. Green et P. Griffiths montrent que ce 0-cycle n'est pas nul si D est « assez général ».

A. Roitman [42] a d'ailleurs démontré peu après Mumford que AJ_X^2 induit un isomorphisme sur la torsion, et est injective si et seulement si l'application différence

$$S^n(X) \times S^n(X) \rightarrow CH^2(X)_0, (\{x_1, \dots, x_n\}, \{x'_1, \dots, x'_n\}) \mapsto \sum x_i - \sum x'_i$$

est surjective pour n assez grand (où $S^n(X)$ désigne la puissance symétrique n -ième de X , c'est-à-dire X^n/\mathfrak{S}_n). Ces résultats impliquent qu'en un sens convenable, $CH^2(X)_0$ n'est pas contrôlé par une variété de dimension finie si $p_g(X) \neq 0$.

1.2. Pour

Dans l'article non encore publié [25], mais qui remonte à plusieurs années, S.-I. Kimura introduit un point de vue tout à fait nouveau sur cette question de la « dimension finie » des groupes de Chow. Étant donnés $\alpha_1, \dots, \alpha_n \in CH^*(X)_{\mathbb{Q}} := CH^*(X) \otimes \mathbb{Q}$, on peut former leur produit $\alpha_1 \times \dots \times \alpha_n \in CH^*(X^n)_{\mathbb{Q}}$ et définir les produits alternés et symétriques par la recette usuelle

$$\alpha_1 \wedge \dots \wedge \alpha_n = \sum_{\sigma \in \mathfrak{S}_n} \frac{\text{sgn}(\sigma)}{n!} \alpha_{\sigma(1)} \times \dots \times \alpha_{\sigma(n)}, \quad \beta_1 \bullet \dots \bullet \beta_n = \sum_{\sigma \in \mathfrak{S}_n} \frac{1}{n!} \beta_{\sigma(1)} \times \dots \times \beta_{\sigma(n)}.$$

Kimura montre que lorsque X est le produit de deux courbes projectives lisses, *il existe un entier n tel que, pour tout n -uplet de cycles $(\alpha_1, \dots, \alpha_n)$ dans le noyau de $AJ_X^2 \otimes \mathbb{Q}$, on ait $\alpha_1 \wedge \dots \wedge \alpha_n = 0$.*

Il montre par ailleurs que si X est une courbe, l'assertion analogue est vraie à condition de remplacer produits alternés par produits symétriques : *il existe un entier n tel que, pour tout n -uplet de cycles $(\beta_1, \dots, \beta_n)$ dans $CH^1(X)_0 \otimes \mathbb{Q}$, on ait $\beta_1 \bullet \dots \bullet \beta_n = 0$.*

Ces observations suggèrent que les groupes de Chow $\otimes \mathbb{Q}$ se comportent en un sens comme des *super-espaces vectoriels de dimension finie*.

S.-I. Kimura, et indépendamment P. O'Sullivan, ont avancé l'idée audacieuse qu'il s'agirait d'un phénomène tout à fait général : ils conjecturent que, pour toute variété projective lisse X , *il existe une décomposition (en général non canonique)*

$$CH^*(X)_{\mathbb{Q}} = CH^*(X)_+ \oplus CH^*(X)_-$$

et un entier naturel n , tels que pour tout n -uplet $(\alpha_1, \dots, \alpha_n)$ d'éléments de $CH^(X)_+$ et tout n -uplet $(\beta_1, \dots, \beta_n)$ d'éléments de $CH^*(X)_-$, on ait*

$$\alpha_1 \wedge \dots \wedge \alpha_n = \beta_1 \bullet \dots \bullet \beta_n = 0.$$

L'exposé est consacré à l'analyse des tenants et aboutissants de cette conjecture.

Plan de l'exposé. Le cadre naturel pour comprendre la nature de cette conjecture et ses implications est celui des motifs. Nous commencerons donc par rappeler la construction et diverses propriétés des motifs purs (pour une équivalence adéquate quelconque), notamment la semi-simplicité des motifs numériques conjecturée par A. Grothendieck et prouvée par U. Jannsen.

La forme générale de la conjecture de Kimura-O’Sullivan est que tout motif pour l’équivalence rationnelle est somme d’un motif pair (*i.e.* dont une puissance extérieure s’annule) et d’un motif impair (*i.e.* dont une puissance symétrique s’annule). C’est prouvé pour les motifs «de type abélien». Nous indiquons aussi les liens entre cette conjecture et deux autres conjectures fondamentales sur les groupes de Chow (Voevodsky, Bloch-Beilinson).

Nous passons ensuite à l’étude générale de la notion catégorique d’objet pair et d’objet impair, puis à celle des catégories F -tensorielles dans lesquelles tout objet est somme d’un pair et d’un impair. Cet axiome tout simple s’avère extrêmement fécond, comme l’ont mis en lumière Kimura et surtout O’Sullivan. Nous verrons qu’une telle catégorie se décrit d’une part comme «extension» d’une catégorie F -tensorielle abélienne semi-simple par un «radical» localement nilpotent, et d’autre part comme catégorie de super-fibrés vectoriels équivariants.

Revenant aux motifs et aux groupes de Chow, nous tirons pour terminer les fruits de ces considérations catégoriques.

Remerciements. Je remercie Bruno Kahn pour sa lecture critique attentive d’une première version de ce texte. Je remercie Shun-Ichi Kimura et Peter O’Sullivan d’avoir mis à ma disposition les versions les plus récentes de leurs papiers soumis, et d’avoir répondu diligemment à toutes mes questions concernant leurs travaux. Ce texte s’est nourri des nombreuses discussions que j’ai eues avec eux trois à propos des motifs de dimension finie.

Convention générale. Dans tout le texte, F désigne un corps de caractéristique nulle. Nous dirons *catégorie F -tensorielle* pour «catégorie F -linéaire monoïdale symétrique \mathcal{T} , essentiellement petite¹, **rigide** (tout objet M a un dual M^\vee), **pseudo-abélienne** (tout endomorphisme idempotent a un noyau), et vérifiant $\text{End } \mathbf{1} = F$ ($\mathbf{1}$ désignant l’objet unité)».

Dans toute catégorie F -tensorielle, on a la notion de *trace d’un endomorphisme*, et de *rang* (ou dimension) d’un objet : $\text{rg } M = \text{tr}(id_M) \in F$. Nous renvoyons à [43] pour les bases de la théorie des catégories monoïdales.

Étant donnés deux objets M, N , on note $\mathcal{T}(M, N)$ le F -espace des morphismes de M vers N . Les \otimes -foncteurs entre catégories F -tensorielles sont sous-entendus F -linéaires.

2. MOTIFS PURS

2.1. Définitions

La notion de motif pur a été introduite par Grothendieck dans les années soixante. Elle permet notamment de mettre en valeur les propriétés catégoriques du calcul des

¹*i.e.* les classes d’isomorphisme d’objets, et les morphismes entre deux objets quelconques, forment des ensembles appartenant à un univers fixé.

correspondances algébriques. Un exposé Bourbaki lui a été consacré il y a 35 ans [11]. La théorie s'étant beaucoup développée depuis une douzaine d'années, un bref tour d'horizon s'impose.

Soit $\mathcal{P}(k)$ la catégorie des variétés projectives lisses sur un corps k . Pour tout $X \in \mathcal{P}(k)$, notons $\mathcal{Z}^*(X)_F$ le F -espace vectoriel engendré par les sous-variétés irréductibles de X , gradué par la codimension. On fixe une relation d'équivalence *adéquate* \sim (en termes vagues, il s'agit d'une relation d'équivalence linéaire \sim sur $\mathcal{Z}^*(X)_F$ pour tout $X \in \mathcal{P}(k)$, telle que modulo \sim , les images directes et inverses des cycles sont définies ainsi que le produit d'intersection²). Muni du produit d'intersection, le quotient $\mathcal{Z}_\sim^*(X)_F = \mathcal{Z}^*(X)_F / \sim$ acquiert alors une structure d'anneau gradué. Les éléments de $\mathcal{Z}_\sim^{\dim X+r}(X \times Y)_F$ s'appellent correspondances algébriques (modulo \sim) entre X et Y , de degré r . La formule

$$g \circ f = pr_{XZ}^{XYZ}(pr_{XY}^{XYZ^*}(f) \cdot pr_{YZ}^{XYZ^*}(g))$$

définit une loi de composition associative pour les correspondances modulo \sim (les degrés s'additionnent). Les équivalences adéquates qui nous intéresseront sont les suivantes, de la plus fine à la plus grossière : - l'équivalence rationnelle : $\alpha \sim_{rat} 0$ dans $\mathcal{Z}^*(X)_F$ s'il existe $\beta \in \mathcal{Z}^*(X \times \mathbb{P}^1)_F$ tel que $\beta(0)$ et $\beta(\infty)$ soient bien définis et que $\alpha = \beta(0) - \beta(\infty)$; on écrit plutôt $CH^*(X)_F$ que $\mathcal{Z}_{rat}^*(X)_F$,

- l'équivalence homologique pour une cohomologie de Weil fixée H (à coefficients dans une extension K/F) : $\alpha \sim_{hom} 0$ si sa classe fondamentale en cohomologie H est nulle,

- l'équivalence numérique : $\alpha \sim_{num} 0$ dans $\mathcal{Z}^r(X)_F$ si pour tout cycle de dimension complémentaire β , le degré $\langle \alpha, \beta \rangle$ du 0-cycle $\alpha \cdot \beta$ est nul³.

Définissons à présent la *catégorie* $M_\sim(k)_F$ des *motifs purs* à coefficients dans F sur le corps de base k , pour l'équivalence \sim .

Les objets sont les triplets (X, e, r) , où $X \in \mathcal{P}(k)$, $r \in \mathbb{Z}$, et $e \in \mathcal{Z}_\sim^{\dim X+r}(X \times X)_F$ est une correspondance idempotente : $e^2 = e$. Il est suggestif de noter ce triplet $e\mathfrak{h}_\sim(X)(r)$ (ou simplement $e\mathfrak{h}(X)(r)$), en pensant au symbole \mathfrak{h} comme à une cohomologie universelle.

Un morphisme $e\mathfrak{h}(X)(r) \rightarrow f\mathfrak{h}(Y)(s)$ est un élément de

$$f \circ \mathcal{Z}_\sim^{\dim X-r+s}(X \times Y)_F \circ e$$

(correspondance de degré $s - r$). La composition des morphismes est celle des correspondances.

On a un foncteur contravariant $\mathfrak{h}_\sim : \mathcal{P}(k) \rightarrow M_\sim(k)_F$ qui associe à X le motif $\mathfrak{h}_\sim(X) = id \cdot \mathfrak{h}_\sim(X)(0)$ et à tout morphisme le transposé de son graphe.

²Voir [45] ou [1] pour plus de précisions.

³Une conjecture «standard» de Grothendieck prédit que l'équivalence homologique coïncide avec l'équivalence numérique.

Le produit cartésien des variétés (et des cycles) induit une structure monoïdale sur $M_{\sim}(k)_F$. On vérifie aisément que c'est une *catégorie F -tensorielle* (au sens de notre convention générale). L'unité est le motif du point $\mathbf{1} = \mathfrak{h}(\text{Spec } k)$.

Pour \sim_{rat} , on note plutôt $CHM(k)_F$ que $M_{rat}(k)_F$; c'est la catégorie des « motifs de Chow » à coefficients dans F . On a la formule

$$CH^r(X)_F = CHM(k)_F(\mathbf{1}, \mathfrak{h}(X)(r)),$$

ce qui suggère de poser

$$CH^r(M) := CHM(k)_F(\mathbf{1}, M(r))$$

pour un motif de Chow quelconque M à coefficients dans F .

Par ailleurs, si \sim est au moins aussi fine que l'équivalence homologique, H définit un \otimes -foncteur

$$M_{\sim}(k)_F \rightarrow sVec_K$$

vers la catégorie K -tensorielle des super-espaces vectoriels de dimension finie sur une extension K de F .

2.2. Exemples

Variétés abéliennes. Soit X une variété abélienne de dimension d sur k . D'après A. Sernesov [46] (et d'autres auteurs après lui), il existe une décomposition canonique

$$\mathfrak{h}(X) = \bigoplus_0^{2d} \mathfrak{h}^n(X),$$

avec la propriété que, pour tout $n \in \mathbb{N}$, la correspondance diagonale $\mathfrak{h}(X^n) = \mathfrak{h}(X)^{\otimes n} \rightarrow \mathfrak{h}^n(X)$ induit un isomorphisme

$$S^n(\mathfrak{h}^1(X)) \xrightarrow{\cong} \mathfrak{h}^n(X),$$

où S^n désigne la puissance symétrique n -ième⁴. Nous nous servirons du corollaire suivant :

PROPOSITION 2.1. — *Pour $n > 2d$, $S^n(\mathfrak{h}^1(X)) = 0$.*

Il existe plusieurs preuves voisines de ce fait bien connu. Voici celle de O'Sullivan, dans le cas où X est la jacobienne d'une courbe projective lisse géométriquement connexe C munie d'un k -point x . On a $\mathfrak{h}^1(X) \cong \mathfrak{h}^1(C) = \pi_C^1 \mathfrak{h}(C)$ où $\pi_C^1 = [\Delta_C - pr_1^*x - pr_2^*x]$, qui s'écrit aussi $\pi_C^1 = (\iota, \iota)^*[P]$, où ι est le plongement de C dans X envoyant x sur l'origine, et P est le diviseur de Poincaré sur $X \times X$. En posant $P_{i,j} = (pr_i, pr_j)^*[P]$, on est donc ramené à prouver que $\sum_{\sigma \in \mathfrak{S}_n} P_{1,\sigma(1)} \cdots P_{n,\sigma(n)} = 0$ dans $CH(X^n \times X^n)$ si $n > 2d$.

⁴Appliquant une cohomologie de Weil, on obtient la formule analogue $S^n H^1(X) = H^n(X)$ dans la catégorie tensorielle des super-espaces vectoriels de dimension finie (avec la contrainte de commutativité donnée par la règle des signes de Koszul). Dans la catégorie tensorielle des espaces vectoriels, compte tenu de ce que $H^1(X)$ est de degré impair 1, cela redonne l'isomorphisme canonique usuel pour la cohomologie de $X : \wedge^n H^1(X) = H^n(X)$.

Or on a $\sum_{\sigma \in \mathfrak{S}_n} P_{1,\sigma(1)} \cdots P_{n,\sigma(n)} = \sum_{I,J \subset [1,n]} (-1)^{|I|+|J|} (\sum_{i \in I, j \in J} P_{i,j})^n$, nul du fait que $[P]^n \in CH^n(X^2) = 0$ si $n > 2d$.

On note $CHM(k)_F^{ab}$ la plus petite sous-catégorie F -tensorielle strictement pleine, stable par facteurs directs, de $CHM(k)_F$ contenant les motifs des variétés abéliennes sur une extension finie séparable de k . Par exemple, le motif de toute variété projective lisse dominée par un produit de courbes projectives lisses est dans $CHM(k)_F^{ab}$.

Autre exemple : le motif de toute hypersurface de Fermat de degré N sur un corps k de caractéristique $\nmid N$ est isomorphe à un motif découpé sur une variété abélienne [23].

Surfaces Soient X une surface projective lisse géométriquement connexe sur k , Pic_X^0 sa variété de Picard, Alb_X sa variété d'Albanese. J. Murre [38] construit une décomposition

$$\mathfrak{h}(X) = \bigoplus_0^4 \mathfrak{h}^n(X)$$

avec $\mathfrak{h}^0(X) \cong \mathbf{1}$, $\mathfrak{h}^1(X) \cong \mathfrak{h}^3(X)(1) \cong \mathfrak{h}^1(Pic_X^0) \cong \mathfrak{h}^1(Alb_X)$. En outre, $\mathfrak{h}^2(X)(1)$ se décompose en $\mathbf{1}^{rg^{NS}(X)} \oplus \mathfrak{t}^2(X)(1)$ pour un $\mathfrak{t}^2(X)$ convenable (dont l'image par H est la « partie transcendante » de $H^2(X)$).

2.3. Équivalences adéquates et \otimes -idéaux

Un anneau n'est autre qu'une catégorie (pré-)additive avec un seul objet. Inversement, on peut considérer les catégories additives (*resp.* F -linéaires) comme des anneaux (*resp.* F -algèbres) à plusieurs objets, ce qui permet d'importer en théorie des catégories nombre de concepts, voire d'énoncés, de l'algèbre non commutative. Exemple : un idéal \mathcal{I} d'une catégorie F -linéaire \mathcal{A} est la donnée pour tout couple d'objets (M, N) d'un sous-espace $\mathcal{I}(M, N)$ de $\mathcal{A}(M, N)$, cette famille de sous-espaces (lorsque M et N varient) étant stable par composition à gauche et à droite avec un morphisme arbitraire. On peut alors former la catégorie F -linéaire quotient \mathcal{A}/\mathcal{I} . La somme $\mathcal{I} + \mathcal{J}$ et le produit $\mathcal{I}.\mathcal{J}$ de deux idéaux sont définis comme d'habitude.

Si \mathcal{T} est une catégorie F -tensorielle, on a aussi la notion de \otimes -idéal : idéal \mathcal{I} stable par produit tensoriel avec tout morphisme. Le quotient \mathcal{T}/\mathcal{I} hérite d'une structure monoïdale, et son enveloppe pseudo-abélienne $(\mathcal{T}/\mathcal{I})^\natural$ est une catégorie F -tensorielle.

EXEMPLES 2.2. — 1) Un exemple d'idéal est le *radical* (de Kelly) \mathcal{R} , défini par $\mathcal{R}(M, N) = \{f \in \mathcal{T}(M, N) \mid \forall g \in \mathcal{T}(N, M), 1_M - gf \text{ est inversible}\}$. Ce n'est pas en général un \otimes -idéal (voir ci-dessous 3.5).

2) \mathcal{T} admet un unique \otimes -idéal maximal (distinct de \mathcal{T}) noté \mathcal{N} [3, 7.1] : il est défini par

$$\mathcal{N}(M, N) = \{f \in \mathcal{T}(M, N) \mid \forall g \in \mathcal{T}(N, M), tr(gf) = 0\}.$$

Il coïncide avec \mathcal{R} si et seulement si \mathcal{R} est un \otimes -idéal.

3) Les morphismes dont une puissance tensorielle est nulle forment un \otimes -idéal $\sqrt[\otimes]{0}$, le \otimes -nilradical [3, 7.4]. C'est la réunion des \otimes -idéaux nilpotents⁵. Il est contenu dans $\mathcal{R} \cap \mathcal{N}$. Le \otimes -foncteur $\mathcal{T} \rightarrow \mathcal{T} / \sqrt[\otimes]{0}$ est conservatif, *i.e.* reflète les isomorphismes.

4) Le noyau de tout foncteur (*i.e.* l'ensemble des flèches qu'il annule) est un idéal. Le noyau de tout \otimes -foncteur est un \otimes -idéal.

Comme \sim_{rat} est l'équivalence adéquate la plus fine, on a pour toute équivalence adéquate un foncteur plein $CHM(k) \rightarrow M_{\sim}(k)$ dont on note \mathcal{I}_{\sim} le noyau⁶. On a (*cf.* [19, 1.7], [3, II.6.3].) :

LEMME 2.3. — *Tout \otimes -idéal \mathcal{I} de $CHM(k)$ et de la forme \mathcal{I}_{\sim} pour une équivalence adéquate \sim convenable. La catégorie $M_{\sim}(k)$ n'est alors autre que l'enveloppe pseudo-abélienne de $CHM(k)/\mathcal{I}$.*

EXEMPLES 2.4. — 1) Dans $CHM(k)$, on a $\mathcal{N} = \mathcal{I}_{num}$. En effet, $\mathcal{N}(\mathbf{1}, \mathfrak{h}(X)(r)) \subset CHM(k)(\mathbf{1}, \mathfrak{h}(X)(r)) = CH^r(X)$ est l'ensemble des cycles f tels que pour tout cycle $g \in CHM(k)(\mathfrak{h}(X)(r), \mathbf{1}) = CH^{\dim X - r}(X)$, $tr(fg) = 0$. Mais $tr(fg)$ n'est autre que $\langle f, g \rangle$.

2) Dans $CHM(k)$, l'équivalence adéquate associée à $\sqrt[\otimes]{0}$ est l'équivalence de \otimes -nilpotence : $\alpha \sim_{\otimes nil} 0$ si $\alpha^{\times n} \sim_{rat} 0$ pour $n \gg 0$. Du fait que les $\sqrt[\otimes]{0}(M, M)$ sont des nil-idéaux, les idempotents se relèvent, et il suit que $M_{\otimes nil}(k) = CHM(k) / \sqrt[\otimes]{0}$.

2.4. Le théorème de semi-simplicité de Jannsen

La théorie des motifs purs n'est vraiment sortie des limbes qu'avec le théorème de semi-simplicité de U. Jannsen [17], qui résout une conjecture de Grothendieck, appartenant au cercle des conjectures « standard » (l'état essentiellement conjectural de la théorie avant le théorème de Jannsen est résumé à la fin de [43]).

THÉORÈME 2.5 (Jannsen). — $M_{num}(k)$ est abélienne semi-simple.

Ce résultat n'utilise que l'existence d'un \otimes -foncteur H de $CHM(k) \rightarrow sVec_K$ (un tel \otimes -foncteur est fourni par toute cohomologie de Weil à coefficients dans K). Modulo les remarques précédentes, il découle immédiatement du résultat général suivant appliqué à $\mathcal{T} = CHM(k)$:

PROPOSITION 2.6. — *Soient \mathcal{T} une catégorie F -tensorielle et $H : \mathcal{T} \rightarrow \mathcal{T}'$ un \otimes -foncteur à valeur dans une catégorie K -tensorielle (pour une extension K/F convenable) dans laquelle les Hom sont de dimension finie et les endomorphismes nilpotents sont de trace*

⁵La composition de n morphismes s'obtient à partir du produit tensoriel de ces morphismes en appliquant une permutation cyclique et une contraction. Il s'ensuit, comme l'ont remarqué plusieurs auteurs, que tout endomorphisme \otimes -nilpotent est nilpotent, et engendre même un idéal nilpotent, *cf. e.g.* [3, 7.4.2].

⁶On omet désormais l'indice F pour alléger

nulle. On pose $\bar{\mathcal{T}} = \mathcal{T}/\mathcal{N}$. Alors pour tout couple d'objets (M, N) , $\bar{\mathcal{T}}(M, N)$ est de dimension finie sur F , et $\bar{\mathcal{T}}$ est semi-simple.

Dans le cas de $CHM(k)$, la première assertion signifie que les espaces de cycles modulo équivalence numérique sont de dimension finie.

Prouvons 2.6. Comme le noyau de H est contenu dans \mathcal{N} , on peut remplacer \mathcal{T} par $\mathcal{T}/\text{Ker } H$, c'est-à-dire supposer H fidèle.

Commençons par le cas où $K = F$. Les espaces $\mathcal{T}(M, N)$ sont alors de dimension finie puisque par hypothèse il en est ainsi de leur image par H . Ceci implique que radical de la F -algèbre $\mathcal{T}(M, M)$ est nilpotent, et que le quotient par ce radical est semi-simple. On est donc ramené à prouver que tout idéal nilpotent I de $\mathcal{T}(M, M)$ est contenu dans $\mathcal{N}(M, M)$. Or pour tout $f \in I$ et tout $g \in \mathcal{T}(M, M)$, gf est nilpotent, donc $H(gf)$ aussi, ce qui implique $\text{tr}H(gf) = 0$ par hypothèse, d'où aussi $\text{tr}(gf) = 0$, et finalement $f \in \mathcal{N}(M, M)$. Ainsi $\bar{\mathcal{T}}(M, M)$ est semi-simple pour tout M , donc $\bar{\mathcal{T}}$ est semi-simple (cf. e.g. [3, A.2.10]).

On ramène le cas général au cas où $F = K$ de la manière suivante. Soit \mathcal{T}_K la catégorie K -tensorielle obtenue en étendant les scalaires de \mathcal{T} de F à K et en passant au quotient par le noyau de H puis à l'enveloppe pseudo-abélienne. Alors H s'étend en un \otimes -foncteur fidèle sur la catégorie K -tensorielle \mathcal{T}_K . Il reste à voir que $\bar{\mathcal{T}}(M, N) \otimes_F K$ s'identifie à $\bar{\mathcal{T}}_K(M, N)$ ⁷. Soit V le sous- K -espace de $\text{Hom}_K(H(M^\vee \otimes N), H(\mathbf{1}))$ engendré par $H(\mathcal{T}(M^\vee \otimes N, \mathbf{1}))$. On peut donc trouver des éléments b_1, \dots, b_n de $\mathcal{T}(M^\vee \otimes N, \mathbf{1})$, dont les images par H forment une base de V . Soit $a \in \mathcal{T}(M, N) \cong \mathcal{T}(\mathbf{1}, M^\vee \otimes N)$. Dire que $a \in \mathcal{N}(M, N) \cong \mathcal{N}(\mathbf{1}, M^\vee \otimes N)$, c'est dire que $b \circ a = 0$ dans $F = \text{End } \mathbf{1}$ pour tout $b \in \mathcal{T}(M^\vee \otimes N, \mathbf{1})$, ou encore que $V \circ H(a) = 0 \in K$. On voit donc que $a \mapsto (b_i \circ a)_i$ définit un isomorphisme F -linéaire de $\bar{\mathcal{T}}(M, N) \cong \bar{\mathcal{T}}(\mathbf{1}, M^\vee \otimes N)$ sur F^n . On conclut en appliquant le même argument à \mathcal{T}_K , en remplaçant F par K .

Le théorème de Jannsen explicite la structure de la catégorie des motifs pour la plus grossière des équivalences adéquates. Au-delà, on aimerait comprendre la structure de la catégorie des motifs pour des équivalences plus fines, notamment les motifs de Chow (à coefficients dans $F = \mathbb{Q}$ pour fixer les idées). Nous allons énoncer trois conjectures fondamentales allant dans ce sens, en commençant par celle de Kimura-O'Sullivan.

2.5. Motifs de dimension finie. La conjecture de Kimura-O'Sullivan

CONJECTURE 2.7 (Kimura, O'Sullivan). — *Tout motif de Chow M est somme d'un motif M_+ dont une puissance extérieure est nulle et d'un motif M_- dont une puissance symétrique est nulle.*

⁷Dans le cas des motifs, il s'agit du fait que les cycles modulo équivalence numérique commutent à l'extension des coefficients, ce qui était connu depuis longtemps.

Appliquant ceci au motif $M = \mathfrak{h}(X)$, en posant $CH(X)_+ = CH(M_+)$, $CH(X)_- = CH(M_-)$, on obtient la décomposition $CH(X)_{\mathbb{Q}} = CH(X)_+ \oplus CH(X)_-$ dont il était question dans l'introduction.

THÉORÈME 2.8 (Shermenev, Kimura, O'Sullivan). — *La conjecture de Kimura-O'Sullivan est vraie pour les motifs de $CHM(k)^{ab}$ (cf. 2.2).*

Compte tenu de ce que $CHM(k)^{ab}$ est la plus petite sous-catégorie F -tensorielle strictement pleine, stable par facteurs directs, de $CHM(k)_F$ contenant les \mathfrak{h}^1 des variétés abéliennes sur une extension finie séparable de k , cela découle de 2.1 (via 3.7 ci-dessous).

Cela fournit un réservoir assez abondant d'exemples illustrant la théorie⁸.

2.6. La conjecture de nilpotence de Voevodsky

Dans une direction un peu différente, V. Voevodsky a proposé dans [49] une conjecture de nilpotence très forte pour les groupes de Chow. Notons $CH^*(X)_0$ le sous-groupe de $CH^*(X)_{\mathbb{Q}}$ formé des cycles numériquement équivalents à 0.

CONJECTURE 2.9 (Voevodsky). — *Pour tout $\alpha \in CH^*(X)_0$, sa puissance « cartésienne » n -ième $\alpha^{\times n} \in CH^*(X^n)_{\mathbb{Q}}$ est nulle pour n assez grand.*

Traduisant en termes d'idéaux tensoriels (cf. 2.4), cela donne : $\sqrt[n]{0} \stackrel{?}{=} \mathcal{N}$.

Un petit pas dans cette direction est le suivant, qui montre du moins que négliger la \otimes -nilpotence simplifie énormément la structure des groupes de cycles :

PROPOSITION 2.10. — *Les groupes de cycles modulo \otimes -nilpotence sont dénombrables. Ce sont des invariants géométriques : si k est algébriquement clos, leur formation commute à toute extension algébriquement close de k .*

En effet, Voevodsky montre dans [49] que les cycles algébriquement équivalents⁹ à 0 sont \otimes -nilpotents, et la théorie des variétés de Chow montre que les espaces de cycles modulo équivalence algébrique ont les propriétés énoncés en 2.10 [27]. Il en est donc de même pour toute équivalence adéquate moins fine.

PROPOSITION 2.11. — *La conjecture de Voevodsky implique celle de Kimura-O'Sullivan.*

En effet, il est clair que 2.9 implique la conjecture standard selon laquelle $\sim_{hom} = \sim_{num}$ (M_{hom} est via H une sous-catégorie non-pleine de $sVec_K$, et toute application linéaire dont une puissance tensorielle est nulle est nulle). Par le formalisme des conjectures standard (cf. [26], [1]), cela implique l'existence d'un idempotent dans $M_{hom} = CHM(k) / \sqrt[n]{0}$ qui

⁸Je ne connais en revanche aucun motif hors de $CHM(k)^{ab}$ pour lequel on sache démontrer la conjecture de Kimura-O'Sullivan.

⁹L'équivalence algébrique est définie comme l'équivalence rationnelle en remplaçant \mathbf{P}^1 par une courbe projective lisse quelconque.

découpe la partie paire de la cohomologie $H(M)$. Comme $\sqrt[0]{0}(M, M)$ est un nil-idéal, il se relève en un idempotent de $CHM(k)(M, M)$. Notons $M_+ \oplus M_-$ la décomposition correspondante de M . Que $\wedge^n M_+ = S^n M_- = 0$ pour $n \gg 0$ se voit en appliquant H , compte tenu de ce que le \otimes -foncteur $CHM(k) \rightarrow M_{hom} = CHM(k)/\sqrt[0]{0}$ est conservatif.

La question de la réciproque sera considérée en 3.33.

2.7. Filtration sur les groupes de Chow. La conjecture de Bloch-Beilinson

L'existence de filtrations remarquables sur les groupes de Chow a été conjecturée par S. Bloch et A. Beilinson [6], [5] d'une part, et par J. Murre d'autre part [39] sous une forme différente (mais équivalente d'après Jannsen [18]). Bien que concernant les motifs purs, sa véritable justification se situe dans le cadre de la philosophie des motifs mixtes, cf. [5], [18], [19], [1, ch. 21].

CONJECTURE 2.12 (Bloch-Beilinson). — *On suppose que les projecteurs de Künneth π_X^i (relatifs à une cohomologie H fixée) sont algébriques. Il existe une filtration $F^*CH^*(X)_{\mathbb{Q}}$ d'anneau, décroissante, finie, exhaustive et séparée, respectée par l'action des correspondances algébriques, telle que F^1 soit l'idéal des cycles $\sim_{hom} 0$ et¹⁰ que π_X^i agisse sur le gradué $Gr^{\nu}CH^r(X)_{\mathbb{Q}}$ par l'identité si $i = 2r - \nu$, par 0 sinon.*

On renvoie aux références ci-dessus ainsi qu'à [1, ch. 10] pour la discussion de cette conjecture et de ses implications, notamment l'existence d'une *filtration croissante par le poids* canonique dans $CHM(k)_{\mathbb{Q}}$. Pour deux motifs de Chow M et N purement de poids respectifs m et n , la conjecture implique que $CHM(k)(M, N) = M_{hom}(M, N)$ si $m = n$, et $CHM(k)(M, N) = 0$ si $m < n$.

Si la filtration de Bloch-Beilinson existe, chaque cran F^i de la filtration définit un \otimes -idéal de $CHM(k)$ (donc une relation d'équivalence adéquate), encore noté F^i . On a $F^1 = \mathcal{I}_{hom}$.

THÉORÈME 2.13 (Jannsen [19]). — *Si la conjecture de Bloch-Beilinson est vraie, de même que la conjecture standard de type Lefschetz¹¹, alors $F^i = (\mathcal{I}_{hom})^i$.*

THÉORÈME 2.14 (O'Sullivan [40]). — *La conjecture de Bloch-Beilinson, jointe à la conjecture standard $\sim_{hom} = \sim_{num}$, implique la conjecture de nilpotence de Voevodsky.*

Voir aussi [1, 10.5].

Les liens entre ces conjectures, et ce qu'elles disent sur la structure de $CHM(k)$, deviendront nettement plus transparents dans le formalisme de O'Sullivan, voir 4.5 ci-dessous.

¹⁰Ces conditions impliquent que les correspondances algébriques modulo l'équivalence homologique agissent sur les gradués $Gr^{\nu}CH^*(X)_{\mathbb{Q}}$.

¹¹La conjecture standard de type Lefschetz pour $X \in \mathcal{P}(k)$ de dimension d prédit que l'inverse de l'isomorphisme de Lefschetz $H^i(X) \cong H^{2d-i}(X)(d-i)$ (induit par le $d-i$ -ème cup-produit itéré avec une polarisation de X) est induit par une correspondance algébrique.

Auparavant, il nous faut toutefois effectuer un assez long périple au pays des catégories tensorielles et aborder l'étude générale de la notion d'objet pair et d'objet impair.

3. CATÉGORIES DE KIMURA-O'SULLIVAN

Commençons par quelques préliminaires sur les représentations du groupe symétrique (voir [33] et [10] pour plus de détails).

3.1. Foncteurs de Schur

Soit \mathcal{T} une catégorie F -tensorielle. Pour tout objet M de \mathcal{T} et tout $n \geq 1$, le groupe symétrique \mathfrak{S}_n agit canoniquement sur $M^{\otimes n}$.

Par ailleurs les classes d'isomorphie V_λ de \mathbb{Q} -représentations irréductibles de \mathfrak{S}_n sont en bijection canonique avec les partitions λ de n . On a donc $\mathbb{Q}[\mathfrak{S}_n] = \prod_{\lambda, |\lambda|=n} \text{End}_{\mathbb{Q}} V_\lambda$. Notons c_λ l'idempotent de $\mathbb{Q}[\mathfrak{S}_n]$ définissant V_λ . Le *foncteur de Schur* $S_\lambda : \mathcal{T} \rightarrow \mathcal{T}$ est défini par

$$S_\lambda(M) = c_\lambda(M^{\otimes n}).$$

EXEMPLE 3.1. — $S_{(n)} = S^n$ (puissance symétrique), $S_{(1,1,\dots,1)} = \wedge^n$ (puissance extérieure).

Le diagramme de Young d'une partition λ de $n = |\lambda|$ est l'ensemble $[\lambda]$ des couples (i, j) d'entiers ≥ 1 tels que $j \leq \lambda_i$.

Soient n_1 et n_2 deux entiers de somme n , et soient μ et ν des partitions de n_1 et n_2 respectivement. On définit le coefficient de Littlewood-Richardson $N_{\mu\nu\lambda} = [V_\lambda : V_\mu \otimes V_\nu]$ comme étant la multiplicité de $V_\mu \otimes V_\nu$ dans V_λ (vues comme représentations de $\mathfrak{S}_{n_1} \times \mathfrak{S}_{n_2} \subset \mathfrak{S}_n$). Ces nombres se calculent par une règle combinatoire bien connue [33, I9]. Le formulaire suivant, classique dans le cas où $\mathcal{T} = \text{Vec}_F$ (ou $s\text{Vec}_F$), s'étend formellement à toute catégorie F -tensorielle \mathcal{T} ([10]) :

- FORMULAIRE 3.2. — 1) $S_\mu(M) \otimes S_\nu(M) \cong \bigoplus_{\lambda, |\lambda|=|\mu|+|\nu|} S_\lambda(M)^{N_{\mu\nu\lambda}}$,
 2) $S_\lambda(M \oplus N) \cong \bigoplus_{\mu, \nu, |\lambda|=|\mu|+|\nu|} (S_\mu(M) \otimes S_\nu(N))^{N_{\mu\nu\lambda}}$,
 3) $S_\lambda(M \otimes N) \cong \bigoplus_{\mu, \nu, |\lambda|=|\mu|+|\nu|} (S_\mu(M) \otimes S_\nu(N))^{[V_\mu \otimes V_\nu : V_\lambda]}$,
 4) $S_\lambda(M^\vee) = S_\lambda(M)^\vee$,
 5) si $[\lambda] \subset [\mu]$, $S_\lambda(M) = 0 \Rightarrow S_\mu(M) = 0$.

3.2. Objets pairs et objets impairs

Soit de nouveau \mathcal{T} une catégorie F -tensorielle.

DÉFINITIONS 3.3. — ¹² 1) Un objet M de \mathcal{T} est dit *pair* (*resp. impair*) s'il existe n tel que $\wedge^n M = 0$ (*resp.* $S^n M = 0$).

2) Un objet M de \mathcal{T} est dit *de dimension finie* au sens de Kimura-O'Sullivan s'il est somme d'un objet pair M_+ et d'un objet impair M_- .

3) \mathcal{T} est dite *de Kimura-O'Sullivan* si tout objet est de dimension finie.

Le modèle qui inspire la définition 2) est évidemment celui des super-espaces vectoriels de dimension finie sur F .

EXEMPLES 3.4. — Outre cet exemple, et celui de $CHM(k)^{ab}$ que nous avons déjà rencontré, voici quelques autres exemples de catégories de Kimura-O'Sullivan (nous en verrons d'autres) :

1) la catégorie $Rep_F G$ des représentations (de dimension finie) d'un schéma en groupes affine sur F ; plus généralement, toute catégorie tannakienne. En fait, d'après [9], une catégorie F -tensorielle abélienne est tannakienne si et seulement si tout objet est pair,

2) la catégorie $sRep_F G$ des super-représentations d'un F -schéma en groupes affine G ; plus généralement, si ε est un élément central de $G(k)$ d'ordre divisant 2, la catégorie $Rep_F(G, \varepsilon)$ des super-représentations pour lesquelles l'automorphisme de parité est induit par l'action de ε (*cf.* [10]),

3) la catégorie F -tensorielle des fibrés vectoriels sur une base X projective géométriquement connexe sur le corps F ,

4) si G est un F -schéma en groupes affine, la catégorie F -tensorielle des (super-)fibrés vectoriels G -équivariants sur un (super-) G -schéma affine $X = \text{Spec } A$ avec $A^G = F$.

REMARQUE 3.5. — Dans les exemples 2) et 4), il est essentiel que G soit un « vrai » schéma en groupes affine, et non un super-schéma en groupes affine. La catégorie des représentations du super-schéma en groupes affine $GL(1, 1)$ (formé des matrices 2-2 de la forme $\begin{pmatrix} a & \epsilon b \\ \epsilon c & d \end{pmatrix}$ avec $\epsilon^2 = 0$) n'est pas de Kimura-O'Sullivan [3, 10.1] : la représentation irréductible standard $V = F[\epsilon] \oplus \overline{F[\epsilon]}$ ¹³ est annulée par le foncteur de Schur $S_{(2,2)}$ mais n'est ni paire ni impaire¹⁴.

¹²Nous nous écartons ici un peu tant de la terminologie de Kimura que de celle de O'Sullivan. Au lieu d'objet pair, Kimura parle d'objet pairement de dimension finie, tandis que O'Sullivan parle d'objet positif. O'Sullivan dit semi-positif plutôt que de dimension finie, et appelle catégorie semi-positive une catégorie dans laquelle tout objet est semi-positif. Par ailleurs, Kimura ne considère que le cas où \mathcal{T} est la catégorie des motifs de Chow, mais ses arguments s'étendent au cas général, ainsi qu'il est montré dans [3, §7, §9].

¹³La barre indique le changement de parité.

¹⁴Dans cet exemple le radical de Kelly n'est pas un \otimes -idéal (il est distinct de \mathcal{N}) puisque le passage au quotient par \mathcal{N} n'est pas conservatif.

C'est un exemple de catégorie super-tannakienne. D'après [10], une catégorie F -tensorielle abélienne est super-tannakienne si et seulement si tout objet est annulé par un foncteur de Schur.

Il résulte du point 5) du formulaire 3.2 que pour M pair (*resp.* impair), il existe un plus petit entier naturel $\text{kim } M$ tel que pour tout $n > \text{kim } M$, $\wedge^n M = 0$ (*resp.* $S^n M = 0$). Pour tout objet de dimension finie, notons $\text{kim } M$ le minimum des $\text{kim } M_+ + \text{kim } M_-$ parmi les décompositions de M en objet pair et objet impair¹⁵. Il sera calculé en 3.20.

LEMME 3.6. — *Le rang $\text{rg } M$ de tout objet de dimension finie est un entier, et $|\text{rg } M| \leq \text{kim } M$.*

Il suffit de traiter séparément les cas pair et impair. Supposons M pair, de sorte que $\wedge^n M = 0$ si et seulement si $n > \text{kim } M$. On a $\text{rg}(\wedge^n M) = \binom{\text{rg } M}{n}$, qui ne s'annule que si $\text{rg } M$ est un entier naturel $> n$. D'où $\text{rg } M \leq \text{kim } M$. Le cas d'un objet impair est analogue, en utilisant le fait que $\text{rg}(S^n M) = \binom{\text{rg } M + n - 1}{n}$.

LEMME 3.7. — 1) *La notion d'objet pair (*resp.* impair) est stable par somme, facteur direct, et passage au dual.*

2) *Le produit tensoriel de deux objets M, N de même parité (*resp.* parité différente) est pair (*resp.* impair), et en outre $\text{kim}(M \otimes N) \leq \text{kim } M \cdot \text{kim } N$.*

3) *La notion d'objet de dimension finie est stable par \oplus, \otimes , dualité [et facteur direct].*

Les points 1) et 2) se déduisent du formulaire 3.2, en s'aidant de la règle de Littlewood-Richardson (*cf. e.g.* [25, 5]). Le point 3) en découle immédiatement [excepté en ce qui concerne le passage aux facteurs directs (il n'est pas immédiat qu'une décomposition en pair et impair en induit une sur tout facteur direct), qui suivra du corollaire 3.11 ci-dessous].

On peut aussi se demander si ces notions sont stables par extension¹⁶. C'est le cas pour la notion d'objet pair (*resp.* impair) [35], [13]. Par là, C. Mazza et V. Guletskii prouvent indépendamment (et différemment) que le motif de toute courbe non nécessairement lisse ni projective, vu comme objet de la catégorie triangulée des motifs mixtes¹⁷ à coefficients rationnels, est de dimension finie.

¹⁵Nous verrons bientôt qu'en fait cette somme ne dépend pas de la décomposition, 3.10.

¹⁶En un sens convenable; par exemple si \mathcal{T} est abélienne, ou triangulée.

¹⁷Par exemple celle construite par V. Voevodsky.

3.3. Propriétés de \otimes -nilpotence

PROPOSITION 3.8 (Kimura, O’Sullivan). — *Tout morphisme entre objets M, N de parités différentes est \otimes -nilpotent (d’échelon $\leq \text{kim } M \cdot \text{kim } N + 1$).*

L’argument suivant est tiré de [3, 9.1.9]. La donnée d’un morphisme $f : M \rightarrow N$ équivaut à celle d’un morphisme $\mathbf{1} \rightarrow M^\vee \otimes N$, ce qui nous ramène au cas où $M = \mathbf{1}$ et N est impair. Le morphisme $f^{\otimes n} : \mathbf{1} \rightarrow N^{\otimes n}$ est alors invariant sous \mathfrak{S}_n , donc se factorise à travers $S^n N$. Pour $n = \text{kim } N + 1$, on trouve 0.

COROLLAIRE 3.9. — *Pour tout objet \tilde{M} de dimension finie dans $\mathcal{T} / \sqrt[0]{}$, la décomposition $\tilde{M}_+ \oplus \tilde{M}_-$ en pair et impair est unique.*

C’est immédiat.

COROLLAIRE 3.10. — *Tout objet M dont l’image \tilde{M} dans $\mathcal{T} / \sqrt[0]{}$ est de dimension finie dans $\mathcal{T} / \sqrt[0]{}$ est lui-même de dimension finie. La décomposition $M_+ \oplus M_-$ en pair et impair n’est pas unique en général, mais les classes d’isomorphie de M_+ et de M_- respectivement sont uniques. En particulier, un objet à la fois pair et impair est nul.*

L’argument est le même que celui vu en 2.11. Cela découle du fait que $\sqrt[0]{}(M, M)$ est un nil-idéal. L’idempotent définissant la partie paire de \tilde{M} se relève en un idempotent de M . Que cet idempotent définisse «la» partie paire de M et que cette partie paire soit unique à isomorphisme près découle de ce que le \otimes -foncteur plein $\mathcal{T}_{\text{kim}} \rightarrow (\mathcal{T} / \sqrt[0]{})_{\text{kim}}$ est conservatif.

COROLLAIRE 3.11 ([3], 9.2.1). — 1) *Soit \mathcal{T}_{kim} la sous-catégorie pleine de \mathcal{T} formée des objets de dimension finie. C’est une sous-catégorie F -tensorielle, de Kimura-O’Sullivan.*

2) *Son image dans $\mathcal{T} / \sqrt[0]{}$ n’est autre que $(\mathcal{T} / \sqrt[0]{})_{\text{kim}}$. Elle admet une $\mathbb{Z}/2$ -graduation canonique, compatible à la structure monoïdale, qui induit la décomposition de chaque objet en pair et impair.*

3) *En changeant le signe de la contrainte de commutativité (tressage) de $(\mathcal{T} / \sqrt[0]{})_{\text{kim}}$ suivant la règle de Koszul, on obtient une nouvelle catégorie F -tensorielle dont tout objet est pair.*

Le lemme 3.7 montre que \mathcal{T}_{kim} et $(\mathcal{T} / \sqrt[0]{})_{\text{kim}}$ ont toutes les propriétés d’une catégorie F -tensorielle, sauf peut-être la pseudo-abélianité non encore démontrée. Mais cette dernière ainsi que les autres assertions découlent aisément du corollaire précédent.

EXEMPLE 3.12. — Pour $\mathcal{T} = M_{\text{hom}}(k)$, \mathcal{T}_{kim} est formée des motifs M pour lesquels l’idempotent (de Künneth) π_M^+ de $H(M)$ d’image la partie paire de la cohomologie provient d’un endomorphisme de M . On note $M_{\text{hom}}^\pm(k)$ cette catégorie de Kimura-O’Sullivan.

L’une des conjectures standard prédit que $M_{\text{hom}}^\pm(k) = M_{\text{hom}}(k)$. Par [24], on sait du moins que c’est le cas si H est la cohomologie ℓ -adique (ou cristalline) et si k est algébrique sur \mathbf{F}_p , $p \neq \ell$.

3.4. Propriétés de nilpotence

Soient \mathcal{T} une catégorie F -tensorielle, \mathcal{N} son \otimes -idéal maximal, et $\bar{\mathcal{T}}$ le quotient \mathcal{T}/\mathcal{N} . Pour tout objet M de \mathcal{T} , notons \bar{M} son image dans $\bar{\mathcal{T}}$. Notons ι_M l'isomorphisme $\mathcal{T}(\mathbf{1}, M^\vee \otimes M) \cong \mathcal{T}(M, M)$ et ϵ_M l'évaluation $M \otimes M^\vee \rightarrow \mathbf{1}$.

LEMME 3.13. — *Soit M un objet pair ou impair. Alors pour tout $f \in \mathcal{N}(M, M)$, $f^{\text{kim } M+1} = 0$ ¹⁸.*

La preuve repose sur des calculs formels de traces dans \mathcal{T} . On a une formule du type suivant¹⁹ :

$$\iota_M((id_{M^\vee} \otimes \epsilon_{M^{\otimes n-1}} \otimes id_M) \circ (\iota_{M^{\otimes n}}^{-1}(S_\lambda f))) = \sum_{\sigma \in \mathfrak{S}_n} t_{\lambda, \sigma} f^{\ell(\sigma)},$$

où $\ell(\sigma)$ est la longueur du cycle de σ ayant 1 dans son support, et où $t_{\lambda, \sigma}$ est un nombre rationnel (explicite) si $\ell(\sigma) = n$, et un nombre rationnel multiplié par un produit de traces de puissances non-nulles de f si $\ell(\sigma) < n$. Si $f \in \mathcal{N}(M, M)$, les traces en question sont nulles, et le second membre se réduit à $(\sum_{\sigma \text{ cyclique}} t_{\lambda, \sigma}) f^n$. Pour $\lambda = (1, 1, \dots, 1)$ ou $\lambda = (n)$, il s'avère que le coefficient $\sum_{\sigma \text{ cyclique}} t_{\lambda, \sigma}$ est non nul (il vaut $\frac{(-1)^{n-1}}{n}$ et $\frac{1}{n}$ respectivement).

Prenons $\lambda = (1, 1, \dots, 1)$ si M est pair, et $\lambda = (n)$ si M est impair, avec $n = \text{kim } M + 1$. Alors $S_\lambda f = 0$, donc le membre de gauche de la formule s'annule, et l'on obtient $f^{\text{kim } M+1} = 0$.

THÉORÈME 3.14 ([3]). — ²⁰ *Si M est de dimension finie, $\mathcal{N}(M, M)$ est un idéal nilpotent, d'échelon borné en fonction de $\text{kim } M$. En outre, $\bar{\mathcal{T}}(\bar{M}, \bar{M})$ est une F -algèbre semi-simple de dimension finie.*

Soit $M = M_+ \oplus M_-$ une décomposition en pair et impair. On décompose alors tout endomorphisme en $f = f_+ + f_- + f_\pm$, où f_+ est le terme préservant M_+ , f_- le terme préservant M_- , $f_+ f_- = f_- f_+ = 0$. Un monôme typique intervenant dans le développement de f^n est de la forme

$$m = f_{\epsilon_1}^{k_1} \circ f_\pm \circ f_{\epsilon_2}^{k_2} \circ f_\pm \circ \dots \circ f_\pm \circ f_{\epsilon_r}^{k_r}$$

avec $\epsilon_i \in \{+, -\}$.

Supposons $f \in \mathcal{N}(M, M)$. D'après le lemme précédent, on a alors $m = 0$ dès que l'un des k_i dépasse $\text{kim } M$. D'autre part, d'après la proposition 3.8, on a $f_\pm^{\otimes N} = 0$ dès que N

¹⁸On peut en fait remplacer $\text{kim } M + 1$ par $\text{kim } M$, comme l'ont remarqué Jannsen (*cf.* §10 de la version récente de [25]) et O'Sullivan (*cf.* 3.22 ci-dessous).

¹⁹Dans le cas des motifs, cette formule apparaît dans [25], dans le langage des correspondances ; elle est prouvée en général dans [3, 7.2.6].

²⁰Des énoncés similaires, un peu plus faibles, ont aussi été obtenus par Kimura (dans le cas des motifs) et par O'Sullivan.

dépasse $(\text{kim } M)^2$, ce qui implique que $m = 0$ dès que le nombre de fois où f_{\pm} apparaît est $> (\text{kim } M)^2$.

Ceci montre que $\mathcal{N}(M, M)$ est un nil-idéal d'échelon borné en fonction de $\text{kim } M$. Grâce à un théorème de Nagata-Higman ([16], voir aussi [3, 7.2.8]), il résulte de là que $\mathcal{N}(M, M)$ est un idéal nilpotent d'échelon borné en fonction de $\text{kim } M$.

Pour voir que $\bar{\mathcal{T}}(\bar{M}, \bar{M})$ est semi-simple de dimension finie sur F , il est loisible de remplacer \mathcal{T} par $\bar{\mathcal{T}}_{\text{kim}}$, puis de changer le tressage comme en 3.11.3), ce qui nous ramène au cas pair. Il reste à établir, sous ces hypothèses, que $\dim_F \mathcal{T}(M, M) < \infty$. Soient f_1, \dots, f_n des endomorphismes de M linéairement indépendants sur F . Compte tenu de l'hypothèse $\mathcal{N} = 0$, ils définissent un facteur direct $\mathbf{1}^n$ de $M^\vee \otimes M$. On en déduit que $n \leq \text{rg } M^2$, d'où $\dim_F \mathcal{T}(M, M) \leq (\text{rg } M)^2$.

COROLLAIRE 3.15. — *Tout idempotent (resp. tout système orthogonal d'idempotents) de \bar{M} se relève en un idempotent (resp. en un système orthogonal d'idempotents) de M .*

COROLLAIRE 3.16. — 1) $\bar{\mathcal{T}}_{\text{kim}}$ est abélienne semi-simple et le foncteur plein $\mathcal{T}_{\text{kim}} \rightarrow \bar{\mathcal{T}}_{\text{kim}}$ est conservatif.

2) Deux objets de \mathcal{T}_{kim} sont isomorphes si leurs images dans $\bar{\mathcal{T}}_{\text{kim}}$ le sont. En particulier, \mathcal{T}_{kim} n'a pas d'objet fantôme, c'est-à-dire d'objet non nul qui devient nul dans $\mathcal{T}_{\text{kim}} \rightarrow \bar{\mathcal{T}}_{\text{kim}}$.

1) En effet, $\bar{\mathcal{T}}_{\text{kim}}$ est semi-simple en vertu de 3.14 et pseudo-abélienne en vertu de 3.15, donc abélienne semi-simple. Pour la conservativité du foncteur plein $\mathcal{T}_{\text{kim}} \rightarrow \bar{\mathcal{T}}_{\text{kim}}$, il suffit de vérifier que tout endomorphisme de M qui induit l'identité sur \bar{M} est un automorphisme, ce qui est clair puisque $\mathcal{N}(M, M)$ est nilpotent.

2) résulte de ce que $\mathcal{T}_{\text{kim}} \rightarrow \bar{\mathcal{T}}_{\text{kim}}$ est plein et conservatif.

COROLLAIRE 3.17. — *Supposons M de dimension finie. Alors M est pair (resp. impair) si et seulement si \bar{M} l'est dans $\bar{\mathcal{T}}$.*

Cela découle de ce que $M \in \mathcal{T}_{\text{kim}}$ et de la conservativité de $\mathcal{T}_{\text{kim}} \rightarrow \bar{\mathcal{T}}_{\text{kim}}$.

COROLLAIRE 3.18. — *Tout objet M de \mathcal{T}_{kim} est somme d'objets indécomposables; M est indécomposable si et seulement si son image \bar{M} dans $\bar{\mathcal{T}}_{\text{kim}}$ est irréductible.*

COROLLAIRE 3.19. — *Supposons $M \in \mathcal{T}$ pair ou bien impair. Si $\text{rg } M = 0$, $M = 0$. Si $\text{rg } M = 1$ ou -1 , M est inversible (eu égard à \otimes).*

Modulo \mathcal{N} , c'est clair. On remonte à M par nilpotence de $\mathcal{N}(M, M)$.

COROLLAIRE 3.20. — *Pour tout objet $M = M_+ \oplus M_-$ de \mathcal{T} , $\text{kim } M = \text{rg } M_+ - \text{rg } M_-$.*

L'inégalité \geq résulte de 3.6 et 3.10. L'inégalité inverse dit que $\wedge^{\text{rg } M_+} M_+ \otimes S^{\text{rg } M_-} M_- \neq 0$. Or $\wedge^{\text{rg } M_+} M_+$ (resp. $S^{\text{rg } M_-} M_-$) est pair (resp. impair) de rang 1 (resp. -1), donc inversible d'après 3.19.

COROLLAIRE 3.21 ([41]). — *Supposons M pair de dimension m . Alors il existe un \otimes -foncteur $Rep_F GL(m) \rightarrow \mathcal{T}$ qui envoie la représentation standard de $GL(m)$ sur M .*

(Il y a un énoncé correspondant pour les objets impairs, en remplaçant $Rep_F GL(m)$ par $Rep_F (GL(-m), -id)$).

Cela suit de 3.19. Le point est que $Rep_F GL(m)$ est la catégorie F -tensorielle obtenue à partir de la catégorie monoïdale symétrique rigide librement engendrée par un objet V en passant à l'enveloppe pseudo-abélienne, en tuant $\wedge^{m+1}V$ et en inversant $\wedge^m V$ (la démonstration de ce fait se ramène au résultat classique selon lequel l'homomorphisme canonique $F[\mathfrak{S}_n] \rightarrow End_{GL(m)}(V^{\otimes n})$ est bijectif pour $n \leq m$, et surjectif pour $n > m$ de noyau engendré par l'antisymétriseur $c_{(1, \dots, 1)}$ (qui induit \wedge^{m+1})).

Pour tout n et tout endomorphisme f d'un objet M de \mathcal{T} (non nécessairement pair), on a la formule $tr(\wedge^n(1_M + f)) = \sum_{i=0}^n tr(\wedge^i f)$, d'où, en étendant les scalaires de F à $F(t)$:

$$tr(\wedge^n(1_M - tf)) = \sum_{i=0}^n (-1)^i tr(\wedge^i f) t^i,$$

cf. [3, 7.2.5]. Pour M pair de rang m , $\wedge^m M$ est inversible d'après ce qui précède, donc $\wedge^m(1_M - tf)$ s'identifie à un polynôme $P_f(t)$, égal à $\sum_{i=0}^{\infty} (-1)^i tr(\wedge^i f) t^i \in F[t]$, et appelé *polynôme caractéristique* de f .

COROLLAIRE 3.22 (Cayley-Hamilton-O'Sullivan). — *Pour tout endomorphisme f d'un objet pair, on a $P_f(f) = 0$.*

Par le corollaire précédent, cela résulte du théorème de Cayley-Hamilton usuel.

Le résultat suivant est un complément à 3.14 dans le « cas pair » :

THÉORÈME 3.23 ([3], 8.2.4). — *Soit \mathcal{T} une catégorie F -tensorielle. On suppose que la dimension de tout objet est un entier naturel. Alors les conditions suivantes sont équivalentes :*

- i) *tout objet est pair,*
- ii) *\mathcal{T}/\mathcal{N} est tannakienne semi-simple sur F , et pour tout $M \in \mathcal{T}$, $\mathcal{N}(M, M)$ est un idéal nilpotent,*
- iii) *$\mathcal{R} = \mathcal{N}$,*
- iv) *il n'y a pas d'objet fantôme dans \mathcal{T} .*

REMARQUE 3.24. — Dans l'exemple 3.5, aucune de ces conditions n'est vérifiée, bien que $\bar{\mathcal{T}} \cong Vec_F$. Que M soit annulé par un foncteur de Schur ne suffit donc pas à entraîner la nilpotence de $\mathcal{N}(M, M)$. Cette propriété met à part les puissances extérieures et symétriques parmi tous les foncteurs de Schur.

3.5. Le théorème de scindage

THÉORÈME 3.25 (André-Kahn). — *Supposons que pour tout objet M , $\mathcal{N}(M, M)$ soit un idéal nilpotent (ce qui est le cas notamment si \mathcal{T} est de Kimura-O’Sullivan par 3.14), et que $\bar{\mathcal{T}}$ soit semi-simple. Alors le \otimes -foncteur de projection $\pi : \mathcal{T} \rightarrow \bar{\mathcal{T}}$ admet une section σ (comme \otimes -foncteur), qui est unique à isomorphisme près.*

Ce résultat est un analogue \otimes -catégorique du théorème classique de Wedderburn-Malcev affirmant que pour toute F -algèbre associative A dont le radical est nilpotent et dont le quotient \bar{A} par le radical est semi-simple (ce qui est le cas notamment si A est de dimension finie sur F), l’homomorphisme $A \rightarrow \bar{A}$ admet une section, unique à conjugaison près.

La preuve de [3] procède par approximations successives, partant d’une section « ensembliste », et la corrigeant petit à petit pour en faire 1) une section fonctorielle, puis 2) une section monoïdale; 3) on prouve enfin que toute section monoïdale est compatible à la contrainte de commutativité (tressage).

Elle est très longue et nous ne pouvons en dire ici que quelques mots.

Le pas 1) s’inspire de la preuve de Hochschild du théorème de Wedderburn, basée sur le fait que le H^1 (de Hochschild) d’une F -algèbre semi-simple est nul (la cohomologie de Hochschild a été étendue aux catégories F -linéaires par B. Mitchell [36]). Ceci fournit une section fonctorielle σ . Pour le pas 2), on montre que dans des gradués convenables associés aux puissances du radical, $\sigma(f \otimes g) - \sigma(f) \otimes \sigma(g)$ a la propriété de 2-cocycle, puis que c’est un 2-cobord en utilisant le fait que le H^2 du monoïde libre sur les classes d’isomorphisme d’objets de \mathcal{T} est nul. Ceci permet de corriger σ par approximations successives, et d’en faire une section monoïdale. Le pas 3) repose sur un calcul de cocycles similaire.

Ce théorème s’applique notamment à $CHM(k)^{ab}$, et fournit une section (fonctorielle monoïdale symétrique) de la projection $CHM(k)^{ab} \rightarrow M_{num}^{ab}$, unique à isomorphisme près. D’après O’Sullivan, elle est même unique à isomorphisme unique près²¹.

3.6. La stratégie de O’Sullivan

Dans le cas où \mathcal{T} est de Kimura-O’Sullivan, O’Sullivan a proposé indépendamment une démonstration complètement différente du théorème de scindage [40], [41]. L’intérêt majeur de cette preuve est qu’elle conduit à un théorème de structure pour les catégories de Kimura-O’Sullivan, comme nous le verrons au paragraphe suivant.

En voici les grandes lignes, en supposant pour simplifier que tout objet est pair (cf. 3.11), de sorte que $\bar{\mathcal{T}}$ est tannakienne semi-simple, et qu’il existe un foncteur fibre à valeurs dans Vec_F , de sorte $\bar{\mathcal{T}} \cong Rep_F G_{\mathcal{T}}$ pour un certain groupe pro-réductif $G_{\mathcal{T}}$ sur F (c’est par exemple le cas si F est algébriquement clos).

²¹Ceci est dû à la structure d’algèbre de Hopf du motif de Chow d’une variété abélienne.

Soit G un groupe pro-réductif sur F . Si A est une G -algèbre commutative (c'est-à-dire une algèbre commutative dans la \otimes -catégorie $REP_F G$ des Ind-objets de $Rep_F G$), on note $X = \text{Spec } A$ le G -schéma affine sur F correspondant, et $Vec(G, X)$ la catégorie F -tensorielle des fibrés vectoriels équivariants sur X . On a un \otimes -foncteur évident $\mathcal{O}_X \otimes - : Rep_F G \rightarrow Vec(G, X)$. On peut démontrer que tout objet de $Vec(G, X)$ est facteur direct d'un objet dans l'image de ce foncteur.

EXEMPLE 3.26. — Prenons $G = \mathbf{G}_m$, agissant sur $A = F[X, Y]$ de manière standard (X, Y en poids 1). Alors $Vec(\mathbf{G}_m, \mathbf{A}_F^2)$ est équivalente à la catégorie F -tensorielle $Vec(\mathbf{P}_F^1)$ des fibrés vectoriels sur \mathbf{P}_F^1 . La représentation de dimension 1 et de poids i de \mathbf{G}_m correspond à $\mathcal{O}_{\mathbf{P}^1}(i)$, et tout objet de $Vec(\mathbf{P}_F^1)$ est somme directe d'objets du type $\mathcal{O}_{\mathbf{P}^1}(i)$, $i \in \mathbb{Z}$ (Grothendieck).

La preuve de O'Sullivan repose sur les trois lemmes suivants.

LEMME 3.27. — Soit $\Sigma : Rep_F G \rightarrow \mathcal{T}$ un \otimes -foncteur vers une catégorie F -tensorielle \mathcal{T} . Alors il existe un G -schéma affine $X = \text{Spec } A$ sur F , et un \otimes -foncteur pleinement fidèle $S : Vec(G, X) \rightarrow \mathcal{T}$ tel que $\Sigma = S \circ (\mathcal{O}_X \otimes -)$. Le couple (X, S) est universel parmi ceux réalisant $\Sigma = S \circ (\mathcal{O}_X \otimes -)$.

Par semi-simplicité de $Rep_F G$, Σ admet un Ind-adjoint à droite $\Theta : \mathcal{T} \rightarrow REP_F G$, et $A = \Theta(\mathbf{1})$ est automatiquement une G -algèbre, dont on note X le spectre. On définit alors S sur les objets en envoyant $\mathcal{O}_X \otimes M$ sur $\Sigma(M)$. Pour définir S sur les morphismes, on utilise les isomorphismes canoniques :

$$Vec(G, X)(\mathcal{O}_X \otimes M, \mathcal{O}_X \otimes N) \cong REP_F G(M \otimes N^\vee, A) \cong \mathcal{T}(\Sigma(M \otimes N^\vee), \mathbf{1}) \cong \mathcal{T}(\Sigma(M), \Sigma(N)).$$

L'universalité du couple (X, S) se vérifie sans difficulté.

On notera que $\mathcal{T}(\mathbf{1}, \mathbf{1}) = F \Leftrightarrow Vec(G, X)(\mathbf{1}, \mathbf{1}) = F \Leftrightarrow A^G = F$.

LEMME 3.28. — On suppose $A^G = F$. Les conditions suivantes sont équivalentes :

- i) $X = \text{Spec } A$ est un G -espace homogène,
- ii) X n'a pas de sous-schéma fermé $\neq X$ stable sous G ,
- iii) $Vec(G, X)$ est semi-simple.

L'équivalence de i) et ii) est due à A. Magid [34]. Prouvons ii) \Leftrightarrow iii). L'hypothèse $A^G = F$ permet d'appliquer 3.23, qui montre que $\mathcal{R} = \mathcal{N}$ dans $Vec(G, X)$. Il suit que $Vec(G, X)$ est semi-simple si et seulement si tout morphisme non nul $M \rightarrow \mathbf{1} = \mathcal{O}_X$ dans $Vec(G, X)$ admet une section. Par ailleurs, ii) $\Leftrightarrow A$ est simple \Leftrightarrow tout morphisme non nul $M \rightarrow \mathcal{O}_X$ dans $Vec(G, X)$ est un épimorphisme. Or tout tel épimorphisme admet une section : le morphisme canonique $\mathbf{1} = F \rightarrow A$ dans la catégorie Ind-semi-simple $REP_F G$ se relève en un morphisme $F \rightarrow M$, donc l'automorphisme identique de $\mathbf{1} = \mathcal{O}_X$ dans $Vec(G, X)$ se relève en un morphisme $\mathcal{O}_X \rightarrow M$.

LEMME 3.29. — Soient $\iota : \bar{X} \hookrightarrow X$ une G -immersion fermée de G -schémas affines sur F , \bar{X} étant homogène, et $X = \text{Spec } A$ vérifiant $A^G = F$. Alors ι admet une G -rétraction.

Ce lemme est une variante du théorème du «slice étale» de D. Luna [32, Cor 2]. Une réduction assez délicate permet de se ramener au cas où G et l'algèbre A sont de type fini. On montre alors que A est complète vis-à-vis de l'idéal de \bar{X} dans X , ce qui permet de se ramener ensuite au cas où cet idéal est de carré nul. Comme $\bar{X} = \text{Spec } \bar{A}$ est homogène, donc lisse sur F , il existe des sections $\bar{A} \rightarrow A$ de la projection (non nécessairement G -équivariantes). De plus, il est clair que ces sections forment un espace affine sur lequel G agit. Comme G est réductif, son action a un point fixe, d'où l'existence d'une section équivariante.

Indiquons pour finir comment l'existence d'une section σ de $\pi : \mathcal{T} \rightarrow \bar{\mathcal{T}}$ découle de ces lemmes, sous l'hypothèse simplificatrice que tout objet de \mathcal{T} est pair et que $\bar{\mathcal{T}} \cong \text{Rep}_F G_{\mathcal{T}}$. D'après 3.21, il existe un \otimes -foncteur essentiellement surjectif $\text{Rep}_F G \rightarrow \mathcal{T}$, où G est un produit de groupes linéaires. En appliquant 3.27 à cette situation, on obtient $X = \text{Spec } A$ et une \otimes -équivalence $\text{Vec}(G, X) \cong \mathcal{T}$. Le quotient \bar{A} de A par un G -idéal maximal est simple, donc \bar{X} est homogène et $\text{Vec}(G, \bar{X})$ est semi-simple en vertu de 3.28 appliqué à $\bar{X} = \text{Spec } \bar{A}$. Il suit que $\text{Vec}(G, X) \rightarrow \text{Vec}(G, \bar{X})$ induit $\bar{\mathcal{T}} \cong \text{Vec}(G, X)/\mathcal{N} \cong \text{Vec}(G, \bar{X})$. Toute G -rétraction de $\bar{X} \hookrightarrow X$ comme dans 3.29 induit une section de $\text{Vec}(G, X) \rightarrow \text{Vec}(G, \bar{X})$, d'où un quasi-inverse à droite de π , qu'on modifie en une vraie section σ .

3.7. Objets de dimension finie et super-fibrés vectoriels équivariants

Une fois connue l'existence d'une section $\bar{\mathcal{T}} \cong \text{Rep}_F G_{\mathcal{T}} \rightarrow \mathcal{T}$ de la projection π , on peut lui appliquer le lemme 3.27. On en déduit :

THÉORÈME 3.30 (O'Sullivan [40]). — Soit \mathcal{T} une catégorie F -tensorielle dans laquelle tout objet est pair. Supposons en outre $\bar{\mathcal{T}}$ neutre, c'est-à-dire l'existence d'une équivalence $\bar{S} : \text{Rep}_F G \cong \bar{\mathcal{T}}$ pour un groupe pro-réductif G .

Alors il existe un G -schéma affine $X = \text{Spec } A$ muni d'un F -point fixe x , avec $A^G = F$, et une équivalence $S : \text{Vec}(G, X) \cong \mathcal{T}$ telle que $\bar{S} \circ x^* = \pi \circ S$. Le triplet (X, x, S) est unique à isomorphisme près.

EXEMPLE 3.31. — Même dans le cas classique où \mathcal{T} est tannakienne neutre, cela conduit à une description tout à fait nouvelle de telles catégories. Par exemple, si $\mathcal{T} = \text{Rep}_F \mathbf{G}_a$, il découle du théorème de Jacobson-Morosov que $G = SL(2)$, et on vérifie que $\text{Rep}_F \mathbf{G}_a \cong \text{Vec}(SL(2), \mathbf{A}_F^2)$. Le \otimes -foncteur π correspond à prendre la fibre en $0 \in \mathbf{A}^2$.

Plus généralement, le théorème de scindage permet d'associer à tout F -schéma en groupes affine H son enveloppe pro-réductive $G = {}^p\text{Red}(H)$, qui est un groupe pro-réductif bien défini à conjugaison près tel que $(\text{Rep}_F H)/\mathcal{N} \cong \text{Rep}_F G$ [3], [40]. Les classes d'isomorphismes de représentations indécomposables de H sont en bijection avec les classes

d'isomorphismes de représentations irréductibles de G . Le théorème précédent montre de plus l'existence d'un G -schéma affine X pointé tel que $Rep_F H \cong Vec(G, X)$.

Il n'est pas difficile de généraliser 3.30 au cas où \mathcal{T} est une catégorie de Kimura-O'Sullivan quelconque. Il suffit pour cela de remplacer A par une algèbre commutative dans $Ind \bar{\mathcal{T}}$, et $Vec(G, Spec A)$ par la catégorie $Proj_A$ des A -modules projectifs sur les objets de $\bar{\mathcal{T}}$ (c'est-à-dire l'enveloppe pseudo-abélienne de la catégorie dont les objets sont ceux de $\bar{\mathcal{T}}$ et les morphismes entre deux objets M et N définis par $Hom_A(A \otimes M, A \otimes N)$). C'est une catégorie F -tensorielle²², et il y a un \otimes -foncteur évident $A \otimes - : \bar{\mathcal{T}} \rightarrow Proj_A$ qui est l'identité sur les objets. Tout homomorphisme $a : A \rightarrow B$ détermine un \otimes -foncteur $a_* : Proj_A \rightarrow Proj_B$.

On a alors la variante suivante du théorème précédent :

VARIANTE 3.32. — Soit \mathcal{T} une catégorie de Kimura-O'Sullivan. Alors il existe une algèbre commutative A dans $Ind \bar{\mathcal{T}}$ munie d'une augmentation $a : A \rightarrow \mathbf{1}$, telle que $Ind \bar{\mathcal{T}}(\mathbf{1}, A) = F$, et une équivalence $S : Proj_A \cong \mathcal{T}$ telle que $a_* = \pi \circ S$. Le triplet (A, a, S) est unique à isomorphisme près.

Les propriétés de \mathcal{T} sont fidèlement décrites en termes de A . O'Sullivan en a dressé un dictionnaire dont voici un extrait :

\boxed{A}	$\boxed{\mathcal{T} \cong Proj_A}$
idéal de A	\otimes -idéal de \mathcal{T}
L'idéal maximal de A (le noyau de a)	\mathcal{N}
$\sqrt{0}$ (nilradical de A)	$\otimes \sqrt{0}$
A est intègre	\mathcal{T} admet un super-foncteur fibre <i>i.e.</i> un \otimes -foncteur fidèle $\mathcal{T} \rightarrow sVec_K$
$\sqrt{0} = \bigcap_{\mathfrak{p} \text{ premier}} \mathfrak{p}$	$\otimes \sqrt{0}$ est l'intersection des noyaux des \otimes -foncteurs $\mathcal{T} \rightarrow sVec_K$ (pour diverses extensions K/F)

Revenons enfin aux motifs, et ne les quittons plus. O'Sullivan déduit du dernier article de ce dictionnaire la forme renforcée suivante de 2.11 :

²²Si $\bar{\mathcal{T}} = Rep_F(G, \varepsilon)$, $Proj_A$ n'est autre que la catégorie F -tensorielle des super-fibrés vectoriels sur $Spec A$ équivariants sous G et tels que l'action de ε définisse la parité.

THÉORÈME 3.33 (O’Sullivan). — *La conjecture de Voevodsky équivaut à celle de Kimura-O’Sullivan jointe à la conjecture suivante : tout \otimes -foncteur $CHM(k)_{\mathbb{Q}} \rightarrow sVec_K$ se factorise à travers l’équivalence numérique.*

Spécialisant au cas de la sous-catégorie $CHM(k)_{\mathbb{Q}}^{ab}$, cela montre que la conjecture de Voevodsky pour les variétés abéliennes équivaut à une version forte de la conjecture standard $\sim_{hom} = \sim_{num}$ (qui est connue sur les variétés abéliennes en caractéristique 0, pour une cohomologie classique [31]) : le noyau de *tout* \otimes -foncteur $CHM(k)_{\mathbb{Q}}^{ab} \rightarrow sVec_K$ est \mathcal{I}_{num} , question ouverte.

4. APPLICATIONS

Il est temps de tirer les fruits de toutes ces considérations abstraites. Nous allons décliner divers résultats ou perspectives plus ou moins concrets qu’elles offrent en théorie des cycles algébriques.

4.1. Motifs de surfaces

Considérons à nouveau la décomposition de Murre du motif de Chow d’une surface projective lisse X (cf. 2.2) : $\mathfrak{h}(X) = \bigoplus_0^4 \mathfrak{h}^n(X)$, avec

$$\mathfrak{h}^0(X) \cong \mathbf{1}, \quad \mathfrak{h}^1(X) \cong \mathfrak{h}^3(X)(1) \cong \mathfrak{h}^1(Alb_X), \quad \mathfrak{h}^2(X) \cong (\mathbf{1}(-1))^{\text{rg}NS(X)} \oplus \mathfrak{t}^2(X).$$

On a $\text{Ker } AJ_X^2 \otimes \mathbb{Q} = CH^2(\mathfrak{h}^2(X)) = CH^2(\mathfrak{t}^2(X))$.

Comme $\mathfrak{h}_{hom}^2(X)$ est un objet pair de $M_{hom}(k)$, $\mathfrak{h}_{num}^2(X)$ est un objet pair de $CHM(k)/\mathcal{N}$. Si $\mathfrak{h}^2(X)$ est de dimension finie (ce qui revient à dire que $\mathfrak{h}(M)$ l’est, compte tenu de 2.8), il suit de 3.17 que $\mathfrak{h}^2(X)$ est pair et de 3.20 que $\wedge^{t_2+1} \mathfrak{h}^2(X) = 0$, en notant $t_2 = b_2 - \text{rg}NS$ la dimension de la partie transcendante de $H^2(X)$. On en déduit comme dans 1.2 que pour tout n -uplet de cycles $(\alpha_1, \dots, \alpha_n)$ dans le noyau de $AJ_X^2 \otimes \mathbb{Q}$, on a $\alpha_1 \wedge \dots \wedge \alpha_n = 0$ dès que $n > t_2$.

Si $X = C_1 \times C_2$ est un produit de courbes de genres respectifs g_1 et g_2 , on a $t_2 \leq 4g_1.g_2$, et on sait que $\mathfrak{t}^2(X) \subset \mathfrak{h}^1(C_1) \otimes \mathfrak{h}^1(C_2)$ est de dimension finie, d’où l’assertion de dimensionalité finie énoncée en 1.2²³.

Revenons à notre point de départ, le théorème de Mumford selon lequel si X est une surface sur un corps k assez gros (par exemple \mathbb{C}), l’injectivité de AJ_X^2 implique que $H^2(X)$ est engendré par les classes de diviseurs.

S. Bloch conjecture la réciproque, sur un corps k quelconque [7], [8]. En termes motiviques, cette conjecture revient à dire que $t_2 = 0 \Rightarrow CH^2(\mathfrak{t}^2(X)) = 0$.

Si $\mathfrak{h}(X)$ est de dimension finie, on a vu que $\mathfrak{t}^2(X)$ est alors pair, donc $t_2 = 0 \Rightarrow \mathfrak{t}^2(X) = 0$ (3.19), et a fortiori $CH^2(\mathfrak{t}^2(X)) = 0$. La réciproque est d’ailleurs vraie [15] :

²³Dans [25, 10.9], Kimura affine la borne de $4g_1g_2$ à g_1g_2 .

THÉORÈME 4.1. — Soit X une surface avec $b_2 = \text{rg } NS$. Alors X vérifie la conjecture de Bloch si et seulement si $\mathfrak{h}(X)$ est de dimension finie.

Par cette voie, V. Guletskii et C. Pedrini [14] donnent une preuve uniforme de tous les cas où la conjecture de Bloch est connue, c'est-à-dire lorsque X n'est pas de type général : dans tous ces cas, $\mathfrak{h}(X) \in CHM(k)^{ab}$. Par ailleurs, M. Saito [44] montre, du moins sur $k = \mathbb{C}$, que les conditions de 4.1 équivalent aussi à $(\mathcal{N}(\mathfrak{h}(X), \mathfrak{h}(X)))^3 = 0$.

On trouvera dans [25] d'autres applications concrètes de ces idées aux groupes de Chow, dans la même veine.

Un cas ouvert particulièrement intéressant de la conjecture de Bloch est celui des *faux plans projectifs* : surfaces de type général dont le motif homologique est isomorphe à celui de \mathbf{P}^2 . En est-il de même du motif de Chow ? Le fait qu'il s'agisse de surfaces de Shimura (cf. [22]), avec leur pléthore de correspondances de Hecke, pourrait s'avérer utile ici.

4.2. Motifs sur les corps finis

La catégorie \mathbb{Q} -tensorielle des motifs $M_{\sim}(\mathbf{F}_q)_{\mathbb{Q}}$ sur le *corps fini* \mathbf{F}_q (pour une équivalence adéquate \sim arbitraire) a la propriété remarquable de posséder un automorphisme non-trivial du \otimes -foncteur identique : l'automorphisme de Frobenius Fr .

Soit $X \in \mathcal{P}(\mathbf{F}_q)$ une variété projective lisse sur \mathbf{F}_q . La conjecture de Tate pour X s'énonce de deux façons équivalentes [48] :

- pour tout entier r , l'ordre du pôle en q^{-r} de la fonction zêta $Z(X, T) \in \mathbb{Q}(T)$ est le rang de $\mathcal{Z}_{num}^r(X)$,

- pour tout r , l'action de $Gal(\bar{\mathbf{F}}_q/\mathbf{F}_q) \cong \hat{\mathbb{Z}}$ sur $H_{\text{ét}}^r(X_{\bar{\mathbf{F}}_q}, \mathbb{Q}_{\ell}(r))$ est semi-simple pour la valeur propre 1^{24} , et l'application classe de cycle ℓ -adique ($\ell \nmid q$)

$$c_{X, \ell}^r : CH^r(X) \otimes \mathbb{Q}_{\ell} \rightarrow (H_{\text{ét}}^{2r}(X_{\bar{\mathbf{F}}_q}, \mathbb{Q}_{\ell}(r)))^{Gal(\bar{\mathbf{F}}_q/\mathbf{F}_q)}$$

est surjective,

D'autre part, une conjecture de Beilinson [5, 1.0] prédit que l'application classe de cycle est aussi injective, et qu'en fait $CH(X)_{\mathbb{Q}} = \mathcal{Z}_{num}(X)_{\mathbb{Q}}$.

THÉORÈME 4.2 (Kahn [20]). — Si le motif de Chow $\mathfrak{h}(X)$ est de dimension finie, la conjecture de Tate pour X implique la conjecture de Beilinson pour X .

La conjecture de Tate est connue pour les produits de courbes elliptiques [47] et pour certaines hypersurfaces de Fermat [48], [23], de même que la conjecture de Kimura-O'Sullivan par 2.8 ; on en déduit :

COROLLAIRE 4.3 ([20]). — Sur tout produit de courbes elliptiques sur \mathbf{F}_q , $\sim_{rat} = \sim_{num}$. De même sur toute hypersurface de Fermat sur \mathbf{F}_q dont le degré divise un nombre de la forme $q^m + 1$, $m \in \mathbb{N}$.

²⁴i.e. l'homomorphisme naturel des invariants vers les coinvariants est un isomorphisme.

Dans [20], B. Kahn va beaucoup plus loin et prouve la conjecture de Lichtenbaum [30] sur les valeurs spéciales de la fonction zêta de telles variétés.

Prouvons 4.2, c'est-à-dire que sous la conjecture de Tate pour X et en supposant $\mathfrak{h}(X) \in CHM(\mathbf{F}_q)_{kim}$, on a $CHM(\mathbf{F}_q)(\mathbf{1}, \mathfrak{h}(X)(r)) = M_{num}(\mathbf{F}_q)(\mathbf{1}, \mathfrak{h}(X)(r))$ pour tout $r \in \mathbb{N}$. Par 3.18, on peut remplacer $\mathfrak{h}(X)(r)$ par un facteur direct $M \neq 0$ indécomposable; le motif numérique \bar{M} correspondant est alors irréductible. Si $M \cong \mathbf{1}$, la question est réglée puisque $\text{End } \mathbf{1} = \mathbb{Q}$. Supposons donc $M \not\cong \mathbf{1}$ et montrons que $CHM(\mathbf{F}_q)(\mathbf{1}, M) = 0$. On a $\bar{M} \not\cong \mathbf{1}$ (conservativité du foncteur de passage aux motifs numériques 3.16), et a fortiori le motif homologique M_{hom} attaché à M n'est pas isomorphe à $\mathbf{1}$. Par la conjecture de Tate, ceci entraîne que l'action de $\text{Gal}(\bar{\mathbf{F}}_q/\mathbf{F}_q)$ sur $H_{\text{ét}}(M)$ est non triviale (et semi-simple pour la valeur propre 1). Il est bien connu que l'inverse du Frobenius galoisien $\in \text{Gal}(\bar{\mathbf{F}}_q/\mathbf{F}_q)$ agit par $Fr(M_{hom})$. Ainsi $Fr(M_{hom})$ est semi-simple pour la valeur propre 1 et distinct de l'identité. Il suit que $Fr(\bar{M})$ est distinct de l'identité, et comme \bar{M} est irréductible, $Fr(\bar{M}) - id_{\bar{M}}$ est un automorphisme de \bar{M} . Par conservativité du passage aux motifs numériques, $Fr(M) - id_M$ est donc un automorphisme de M . Pour tout $f \in CHM(\mathbf{F}_q)(\mathbf{1}, M)$, on a $Fr(M)f = fFr(\mathbf{1}) = f$, et on conclut que $f = 0$.

REMARQUE 4.4. — Dans [2], il est démontré que la conjecture standard de type Lefschetz pour les espaces X fibrés en variétés abéliennes sur une courbe projective lisse sur $\bar{\mathbf{F}}_q$ implique la conjecture de Tate pour les variétés abéliennes sur \mathbf{F}_q . Compte tenu de 4.2 et de 2.8, elle implique donc aussi la conjecture de Beilinson pour les variétés abéliennes sur \mathbf{F}_q .

Signalons aussi le résultat remarquable suivant, dont la preuve dépasse le cadre modeste de ces notes :

THÉORÈME 4.5 (Kahn [20]). — *Les énoncés suivants sont équivalents :*

- 1) *la conjecture de Tate vaut pour tout $X \in \mathcal{P}(\mathbf{F}_q)$, et la conjecture de Kimura-O'Sullivan vaut pour $CHM(\mathbf{F}_q)$,*
- 2) *la conjecture de Tate vaut pour toute variété abélienne sur \mathbf{F}_q , et $CHM(\mathbf{F}_q)^{ab} = CHM(\mathbf{F}_q)$,*
- 3) *les groupes de cohomologie de Lichtenbaum $H_W^i(X, \mathbb{Z}(n))$ sont de type fini pour tout $X \in \mathcal{P}(\mathbf{F}_q)$ et tout $(i, n) \in \mathbb{N} \times \mathbb{Z}$.*

(D'après le théorème précédent, ces énoncés impliquent aussi la conjecture de Beilinson.)

La cohomologie de Lichtenbaum (« Weil-étale cohomology ») se définit ainsi. Ses coefficients sont les faisceaux étales \mathbb{Z} -équivariants \mathcal{F} sur $X_{\bar{\mathbf{F}}_q}$, l'action de \mathbb{Z} étant donnée par le Frobenius galoisien (on n'exige pas la $\hat{\mathbb{Z}}$ -équivariance!). La cohomologie de Lichtenbaum est donnée par les foncteurs dérivés de $\mathcal{F} \rightarrow H_W^0(X, \mathcal{F}) := H_{\text{ét}}^0(X_{\bar{\mathbf{F}}_q}, \mathcal{F})^{\mathbb{Z}}$.

4.3. Rationalité de fonctions zêta motiviques

L'une des expressions de la fonction zêta $Z(X, T)$ est

$$Z(X, T) = 1 + |X(\mathbf{F}_q)| T + |S^2(X)(\mathbf{F}_q)| T^2 + \dots$$

(où S^n désigne la puissance symétrique n -ième de X).

Dans [21], M. Kapranov a eu l'idée de considérer l'expression

$$1 + X T + S^2(X) T^2 + \dots$$

pour une variété X quelconque (disons quasi projective pour que les puissances symétriques soient définies) sur un corps k non nécessairement fini. Pour donner un sens précis à cette expression, on se place dans le groupe $K_0(Var(k))$ quotient du groupe abélien libre sur les classes d'isomorphismes $[X]$ de k -variétés (non nécessairement projectives ni lisses) par le sous-groupe engendré par les éléments de la forme $[X] - [U] - [Z]$ si $X = U \amalg Z$ avec Z fermé dans X , et U l'ouvert complémentaire. Le produit des variétés en fait un anneau commutatif.

On vérifie alors que la formule précédente définit un homomorphisme $Z_{var}(-, T)$ de $K_0(Var(k))$ vers le groupe abélien (multiplicatif) $1 + TK_0(Var(k))[[T]]$. Kapranov montre que pour toute courbe X , $Z_{var}([X], T)$ est une fonction rationnelle; si X est projective lisse de genre g , $(1 - T)(1 - [\mathbf{A}^1]T)Z_{var}([X], T)$ est un polynôme de degré $2g$.

La question de la rationalité de $Z_{var}([X], T)$ en général, conjecturée par Kapranov, a été tranchée négativement par M. Larsen et V. Lunts [29]²⁵ : si X est un produit de deux courbes de genre non nul, $Z_{var}([X], T)$ n'est pas rationnelle.

Plutôt que $K_0(Var(k))$, on peut prendre comme groupe de coefficients $K_0(M_{\sim}(k))$, quotient du groupe abélien libre sur les classes d'isomorphisme $[M]$ d'objets de $M_{\sim}(k)_{\mathbb{Q}}$ par le sous-groupe engendré par les éléments de la forme $[M] - [M'] - [M'']$ avec $M \cong M' \oplus M''$. Le produit tensoriel des motifs en fait un anneau commutatif.

On obtient une variante motivique de la fonction zêta de Kapranov en montrant que l'expression

$$1 + [M] T + [S^2 M] T^2 + \dots$$

définit un homomorphisme $Z_{mot}(-, T)$ de $K_0(M_{\sim}(k))$ vers le groupe abélien (multiplicatif) $1 + TK_0(M_{\sim}(k))[[T]]$, [1, ch. 12].

PROPOSITION 4.6. — *Si le motif M est de dimension finie au sens de Kimura-O'Sullivan (par exemple si $M \in CHM(k)^{ab}$), alors $Z_{mot}([M], T)$ est rationnelle.*

Supposons que $M = M_+ \oplus M_-$ avec $\wedge^m M_+ = S^m M_- = 0$. Il est clair que $Z_{mot}([M_-], T)$ est de la forme $1 + TP_-(T)$ où P_- est un polynôme de degré $< m$. Pour $Z_{mot}([M_+], T)$,

²⁵La question de la rationalité reste ouverte si l'on localise $K_0(Var(k))$ en inversant $[\mathbf{A}^1]$ (qui est un diviseur de zéro!).

on utilise la formule suivante :

$$\sum [S^n M_+] T^n = \frac{1}{\sum [\wedge^n M_+] (-T)^n}$$

et le dénominateur est de la forme $1 - TP_+(T)$ où P_+ est un polynôme de degré $< m$.

COROLLAIRE 4.7. — *La conjecture de Kimura-O'Sullivan implique la rationalité de $Z_{mot}([M], T)$ pour tout motif de Chow M , et en outre l'égalité $K_0(CHM(k)) = K_0(M_{num}(k))$.*

(Pour la seconde assertion, utiliser la conservativité du passage aux motifs numériques.)

Pour plus de détails, notamment sur le lien entre $Z_{var}(-, T)$ et $Z_{mot}(-, T)$, voir [1, ch. 12].

Les deux derniers paragraphes sont consacrés à des applications du théorème de scindage 3.25, appliqué d'abord à des motifs homologiques, et ensuite à des motifs de Chow.

4.4. Construction inconditionnelle des groupes de Galois motiviques

Rappelons d'abord la construction de Grothendieck des groupes de Galois motiviques, en admettant la conjecture standard $\sim_{hom} = \sim_{num}$ (pour une cohomologie H fixée à coefficients dans une extension K/F). Cette conjecture entraîne l'algébricité des projecteurs de Künneth pairs π_X^\pm , ce qui permet de modifier le tressage selon la règle des signes de Koszul. Ceci fait, la catégorie \mathbb{Q} -tensorielle abélienne semi-simple $M_{hom}(k) = M_{num}(k)$ est munie d'un \otimes -foncteur fidèle exact vers Vec_K défini par la cohomologie (et encore noté H). Le groupe de Galois motivique $G_{mot,k}$ est le K -groupe pro-réductif $Aut H$. Remplaçant $M_{num}(k)$ par la sous-catégorie tannakienne engendré par un objet $\mathfrak{h}(X)$, on obtient un quotient $G_{mot}(X)$ de $G_{mot,k}$, qui est un sous-groupe réductif de $GL(H(X))$.

Un certain nombre de problèmes géométriques faisant intervenir les correspondances algébriques modulo équivalence homologiques se traduisent alors en des problèmes de représentations de groupes réductifs, traitables par la théorie classique des poids, cf. e.g. [1, ch. 5,6,8,9].

Pour définir les groupes de Galois motiviques sans avoir à admettre la conjecture standard, on peut utiliser le théorème de scindage monoïdal comme suit [4]. Considérons la catégorie \mathbb{Q} -tensorielle $M_{hom}^\pm(k)$ introduite en 3.12 (sous-catégorie pleine de $M_{hom}(k)$ formée des motifs dont le projecteur de Künneth pair est donné par un morphisme de $M_{hom}(k)$). C'est une catégorie de Kimura-O'Sullivan. On peut modifier le tressage selon la règle de Koszul pour faire en sorte que $M_{hom}^\pm(k)/\mathcal{N} = M_{num}^\pm(k)$ soit tannakienne semi-simple. Ceci fait, on dispose du \otimes -foncteur fidèle exact $H : M_{hom}^\pm(k) \rightarrow Vec_K$.

Le théorème de scindage fournit une \otimes -section σ de la projection $\pi : M_{hom}^\pm(k) \rightarrow M_{num}^\pm(k)$, unique à isomorphisme près. On définit alors $G_{mot,k}^\pm$ comme le K -groupe pro-réductif $Aut(H \circ \sigma)$. Remplaçant $M_{num}(k)$ par la sous-catégorie tannakienne engendré par

un objet $\mathfrak{h}(X)$, on obtient un quotient $G_{mot}(X)$ de $G_{mot,k}^\pm$, qui est un sous-groupe réductif de $GL(H(X))$ ²⁶.

REMARQUE 4.8. — Le composé $H \circ \sigma \circ \pi$ définit une nouvelle cohomologie²⁷, et l'équivalence homologique qui lui est associée n'est autre que l'équivalence numérique. Elle vérifie le théorème de Lefschetz fort si et seulement si la conjecture standard de type Lefschetz est vraie pour H .

4.5. Motifs de Chow vus comme super-fibrés vectoriels équivariants sous le groupe de Galois motivique. Hauteurs

Le théorème de structure 3.32 s'applique en particulier à toute catégorie F -tensorielle de motifs de Chow qui est Kimura-O'Sullivan. Plutôt que de nous limiter à des catégories telles $CHM(k)_F^{ab}$, nous prendrons ici le parti d'appliquer 3.32 à $CHM(k)_F$ tout entière, en supposant la conjecture de Kimura-O'Sullivan vraie, de même (pour simplifier) que la conjecture standard $\sim_{hom} = \sim_{num}$. D'après 3.32, il existe une algèbre commutative A dans la catégorie F -tensorielle des Ind-motifs numériques, vérifiant $Ind M_{num}(k)_F(\mathbf{1}, A) = F$ et munie d'une augmentation $a : A \rightarrow \mathbf{1}$, ainsi qu'une équivalence $S : Proj_A \cong CHM(k)_F$ telle que $a_* = \pi \circ S$. Le triplet (A, a, S) est unique à isomorphisme près. Les équivalences adéquates sont en bijection avec les idéaux de A .

REMARQUE 4.9. — S'il existe une cohomologie à coefficients dans F , on obtient une équivalence $\bar{S} : Rep_F(G_{mot,k}, -id) \cong M_{num}(k)_F$ (où $-id$ désigne l'image de -1 par le cocaractère des poids). On est alors dans le «super-analogue» de la situation de 3.30 : il existe un super- $G_{mot,k}$ -schéma affine X bien défini à isomorphisme près (sur lequel l'action de $-id$ définit la parité) tel que $CHM(k)_F$ s'identifie à la catégorie F -tensorielle des super-fibrés vectoriels sur X équivariants sous $G_{mot,k}$ (et sur lesquelles l'action de $-id$ définit la parité). Le super-schéma X a un unique F -point fixe sous $G_{mot,k}$ qui correspond à l'équivalence numérique. Prendre la fibre d'un super-fibré vectoriel équivariant en ce point correspond au passage du motif de Chow au motif numérique associé.

Comprendre la structure de $CHM(k)_F$ revient donc à comprendre celle de A . À cet égard, O'Sullivan [40] fait les observations suivantes :

- on a la graduation par le poids $A = \bigoplus_{-\infty}^{+\infty} A_i$, qui est une graduation d'algèbre,
- la conjecture de Voevodsky équivaut à dire que $\ker a$ est un nil-idéal,
- la conjecture de Bloch-Beilinson équivaut à dire que la partie de poids strictement négative de A est nulle. Elle équivaut aussi (via 2.13) à dire que $A_0 = \mathbf{1}$ et A est engendrée par A_1 ,

²⁶Changer σ change ce sous-groupe en un conjugué.

²⁷Mêmes espaces de cohomologie, mais nouvelle application classe de cycles.

- une version plus fine de la conjecture de Bloch-Beilinson implique que $\mathcal{I}_{hom}^i = 0$ dès que i dépasse la dimension de Kronecker $d(k)$ de k . Ceci équivaut à $(A_1)^i = 0$ pour $i > d(k)$,

Pour un corps fini, cela donne $A = \mathbf{1}$, et on retrouve la conjecture de Beilinson mentionnée en 4.2.

Pour un corps de nombres, cela donne $A = \mathbf{1} \oplus A_1$. L'échelon de nilpotence dans la conjecture de Voevodsky serait donc 2 dans ce cas.

Par ailleurs, toujours pour k un corps de nombres, et prenant $F = \mathbb{R}$, O'Sullivan traduit l'existence des accouplements de hauteur de Beilinson en l'existence d'un accouplement symétrique $S^2 A_1 \rightarrow \mathbf{1}(-1)$ induisant sur chaque sous-objet de rang fini de A_1 une forme de Weil relativement à la polarisation canonique de $M_{num}(k)_{\mathbb{R}}$.

RÉFÉRENCES

- [1] Y. André, *Une introduction aux motifs*, monographie soumise à publication.
- [2] Y. André, Cycles de Tate et cycles motivés sur les variétés abéliennes en caractéristique $p > 0$, soumis.
- [3] Y. André, B. Kahn, Nilpotence, radicaux et structures monoïdales (avec un appendice de P. O'Sullivan), *Rendiconti del Seminario Matematico dell'Università di Padova* **108** (2002), 107-291.
- [4] Y. André, B. Kahn, Construction inconditionnelle de groupes de Galois motiviques, *C. R. Acad. Sci. Paris, Ser I* **331** (2002), 989-994.
- [5] A. Beilinson, Height pairing between algebraic cycles, in : *K-theory, Arithmetic and Geometry, Lect. notes in Math.* **1289**, Springer (1987), 27-41.
- [6] S. Bloch, Some elementary theorems about algebraic cycles on Abelian varieties. *Invent. Math.* **37** (1976), no. 3, 215-228.
- [7] S. Bloch, *Lectures on algebraic cycles*, Duke Univ. Math. series IV (1980).
- [8] S. Bloch, A. Kas, D. Lieberman, Zero-cycles on surfaces with $p_g = 0$, *Compos. Math.* **33** (1976), 135-145.
- [9] P. Deligne, Catégories tannakiennes, in : *Grothendieck Festschrift*, vol. 2, Birkhäuser P.M. **87** (1990), 111-198.
- [10] P. Deligne, Catégories tensorielles, *Moscow Math. J.* **2** (2002), 227-248.
- [11] M. Demazure, Motifs des variétés algébriques, exp. 365 (1969), *Sém. Bourbaki* (1969/70), Springer L. N. **180** (1971), 11-38.
- [12] M. Green, P. Griffiths, An interesting 0-cycle, *Duke Math. J.* **119** (2003), 261-313.

- [13] V. Guletskii, Finite dimensional objects in distinguished triangles, prépublication <http://www.math.uiuc.edu/K-theory/0637>.
- [14] V. Guletskii, C. Pedrini, The Chow motive of the Godeaux surface, in : Algebraic Geom. (in memory of P. Francia), de Gruyter (2002), 179-195.
- [15] V. Guletskii, C. Pedrini, Finite dimensional motives and the conjecture of Murre and Beilinson, K-theory **550** (2003), 1-21.
- [16] G. Higman, On a conjecture of Nagata, Proc. Camb. Philos. Soc. **52** (1956), 1-4.
- [17] U. Jannsen, Motives, numerical equivalence and semi-simplicity, Invent. Math. **107** (1992), 447-452.
- [18] U. Jannsen, Motivic sheaves and filtrations on Chow groups, in : *Motives* (Seattle, WA, 1991), Proc. Sympos. Pure Math., **55** (I), AMS, Providence (1994), 245-302.
- [19] U. Jannsen, Equivalence relations on algebraic cycles, in : *The arithmetic and Geometry of algebraic cycles*, proc. Nato conference, Banff 98, Nato series **548** (2000), 225-260.
- [20] B. Kahn, Équivalences rationnelle et numérique sur certaines variétés de type abélien sur un corps fini, Ann. Sci. Éc. Norm. Sup. **36** (2003), 977-1002.
- [21] M. Kapranov, The elliptic curve in the S-duality theory and Eisenstein series for Kac-Moody groups, ArXiv AG/0001005.
- [22] F. Kato, An overview of p -adic uniformization, appendice 2 à : Y. André, *Period mappings and differential equations; from \mathbb{C} to \mathbb{C}_p* , Memoir **12** of the Jap. Math. Soc.
- [23] T. Katsura, T. Shioda, On Fermat varieties, Tôhoku Math. J. **31** (1979), 97-115.
- [24] N. Katz, W. Messing, Some consequences of the Riemann hypothesis for varieties over finite fields, Invent. Math. **23** (1974), 73-77.
- [25] S.-I. Kimura, Chow motives can be finite-dimensional, in some sense, soumis.
- [26] S. Kleiman, Algebraic cycles and the Weil conjectures, in : dix exposés sur la cohomologie des schémas, North Holland (1968), 359-386.
- [27] S. Kleiman, Finiteness theorems for algebraic cycles, *Actes, Congrès intern. math. Nice 1970*, tome 1, 445-449.
- [28] K. Künneman, On the Chow motive of an abelian scheme, in : *Motives*, Proc. Symposia pure Math. **55** (I), AMS (1994), 189-205.
- [29] M. Larsen, V. Lunts, Rationality criteria for motivic zeta-functions, ArXiv AG/0212158.

- [30] S. Lichtenbaum, Values of zeta functions at non-negative integers, in : Springer L. N. **1068** (1984), 127-138.
- [31] D. Lieberman, Numerical and homological equivalence of algebraic cycles on Hodge manifolds, Amer. J. Math. **90** (1968), 366-374.
- [32] D. Luna, Théorème du slice étale, Mém. Soc. Math. Fr. **33** (1973), 81-105.
- [33] I. Macdonald, Symmetric functions and Hall polynomials, Clarendon Press, Oxford (1979).
- [34] A. Magid, Equivariant completions of rings with reductive group action, J. Pure Appl. Algebra **49** (1987), 173-185.
- [35] C. Mazza, Schur functors and motives, Prépublication <http://www.math.uiuc.edu/K-theory/0641>.
- [36] B. Mitchell, Rings with several objects, Adv. Math. **8** (1972), 1-161.
- [37] D. Mumford, Rational equivalence of zero-cycles on surfaces, J. Math. Kyoto Univ. **9** (1969), 195-204.
- [38] J. Murre, On the motive of an algebraic surface, J. reine angew. Math. **409** (1990), 190–204.
- [39] J. Murre, On a conjectural filtration on the Chow groups of an algebraic variety, parts I and II, Indag. Math. **4** (1993), 177–201.
- [40] P. O’Sullivan, Deux lettres à Y. André et B. Kahn, 29/4/02, 12/5/02. + Papiers secrets.
- [41] P. O’Sullivan, The structure of certain rigid tensor categories, soumis.
- [42] A. Roitman, Rational equivalence of zero-cycles, Math. USSR Sbornik **18** (1972), 571-588.
- [43] N. Saavedra Rivano, *Catégories tannakiennes*, Lect. Notes in Math. **265**, Springer, 1972.
- [44] M. Saito, Bloch’s conjecture and Chow motives, preprint RIMS (2000).
- [45] P. Samuel, Relations d’équivalence en géométrie algébrique. in : *Proc. Internat. Congress Math.* 1958, 470–487, Cambridge Univ. Press (1960).
- [46] A. Shermenev, The motive of an abelian variety, Funct. Anal. **8** (1974), 47-53.
- [47] M. Spieß, Proof of the Tate conjecture for products of elliptic curves over finite fields, Math. Ann. **314** (1999), 285–290.
- [48] J. Tate, Conjectures on algebraic cycles in ℓ -adic cohomology, in : *Motives* (Seattle, WA, 1991), Proc. Sympos. Pure Math., **55** (I), AMS, Providence (1994), 71–83.

- [49] V. Voevodsky, A nilpotence theorem for cycles algebraically equivalent to zero, International Mathematics Research Notices **4** (1995), 1–12.

Yves ANDRÉ

D.M.A.

École Normale Supérieure

UMR 8553 du CNRS

45 rue d'Ulm

F-75230 PARIS Cedex 05

E-mail : `yves.andre@ens.fr`

**ÉQUATIONS DE CHAMP MOYEN
POUR LA DYNAMIQUE QUANTIQUE
D'UN GRAND NOMBRE DE PARTICULES**
[d'après Bardos, Erdős, Golse, Gottlieb, Mauser, Yau]

par **Patrick GÉRARD**

INTRODUCTION

Considérons N particules quantiques dans l'espace \mathbf{R}^3 , interagissant deux à deux selon un potentiel V fonction de la distance r entre les deux particules. L'exemple le plus courant est le potentiel coulombien

$$(1) \quad V(r) = \frac{C}{r},$$

où la constante C ci-dessus est positive dans le cas d'une interaction répulsive (par exemple entre charges de même signe), et négative dans le cas d'une interaction attractive (par exemple gravitationnelle). Après adimensionnement des constantes, le hamiltonien quantique associé est l'opérateur différentiel

$$(2) \quad H_N = - \sum_{j=1}^N \Delta_j + \sum_{1 \leq j < k \leq N} V(|x_j - x_k|)$$

où Δ_j désigne l'opérateur de Laplace agissant sur la j -ième position x_j . Le système est alors décrit à l'instant t par sa fonction d'onde $\Psi_N(t) \in L^2(\mathbf{R}^{3N})$ selon l'équation de Schrödinger

$$(3) \quad i \frac{\partial \Psi_N}{\partial t} = H_N \Psi_N, \quad \Psi_N(0) = \Psi_{N,0}.$$

Le fait que ce problème de Cauchy soit bien posé pour tout élément $\Psi_{N,0}$ de $L^2(\mathbf{R}^{3N})$ fait l'objet d'un théorème démontré par Kato en 1951. Pour l'énoncer, rappelons quelques notations. Pour tout entier naturel d , on désigne par $C_0^\infty(\mathbf{R}^d)$ l'espace des fonctions de classe C^∞ à support compact sur \mathbf{R}^d . Si m est un entier naturel, on introduit l'espace de Sobolev

$$H^m(\mathbf{R}^d) = \{u \in L^2(\mathbf{R}^d) \mid \forall \alpha \in \mathbf{N}^d, |\alpha| \leq m, \partial^\alpha u \in L^2(\mathbf{R}^d)\},$$

les dérivées étant définies au sens des distributions.

THÉORÈME 0.1 (Kato [19], 1951). — On suppose que la fonction $x \mapsto V(|x|)$ est à valeurs réelles et appartient à $L^2(\mathbf{R}^3) + L^\infty(\mathbf{R}^3)$. L'opérateur $H_N : C_0^\infty(\mathbf{R}^{3N}) \rightarrow L^2(\mathbf{R}^{3N})$ admet une unique extension autoadjointe. Son domaine est l'espace de Sobolev $H^2(\mathbf{R}^{3N})$.

On note encore H_N l'opérateur autoadjoint ainsi défini. En considérant le groupe à un paramètre unitaire $\exp(-itH_N)$ engendré par H_N grâce au théorème de Stone (cf. par exemple [25]), on en déduit :

COROLLAIRE 0.2. — Sous les hypothèses du théorème 0.1, soit $\Psi_{N,0} \in L^2(\mathbf{R}^{3N})$. Il existe une unique solution $\Psi_N \in C(\mathbf{R}, L^2(\mathbf{R}^{3N}))$ au problème de Cauchy (3), l'équation aux dérivées partielles étant satisfaite au sens des distributions dans $\mathbf{R} \times \mathbf{R}^{3N}$.

Dans le résultat ci-dessus, le caractère unitaire de l'évolution équivaut à la loi de conservation

$$(4) \quad \|\Psi_N(t)\|_{L^2(\mathbf{R}^{3N})} = \|\Psi_{N,0}\|_{L^2(\mathbf{R}^{3N})}$$

tandis que la conservation de l'énergie est traduite par la continuité de l'opérateur $\exp(-itH_N)$ sur $H^1(\mathbf{R}^{3N})$ et par le fait que la quantité

$$(5) \quad E_N = \int_{\mathbf{R}^{3N}} |\nabla_X \Psi_N(t, X)|^2 dX + \sum_{1 \leq j < k \leq N} \int_{\mathbf{R}^{3N}} V(|x_j - x_k|) |\Psi_N(t, X)|^2 dX$$

reste constante au cours du temps. Dans la formule ci-dessus, on a noté $X = (x_1, \dots, x_N)$ le point courant de \mathbf{R}^{3N} .

Un problème fondamental en physique mathématique consiste à décrire cette évolution lorsque N tend vers l'infini. Précisons ce que l'on entend par là. Tout d'abord, afin d'assurer que la force exercée sur chaque particule reste bornée, il convient de normaliser le potentiel V en imposant

$$(6) \quad V(r) = \frac{1}{N} V_1(r)$$

où V_1 est un potentiel indépendant de N . Il faut ensuite préciser la notion de convergence utilisée pour étudier, à chaque instant t , une suite $(\Psi_N(t))$ de fonctions dont le nombre de variables tend vers l'infini avec N . Dans ce but, rappelons que, selon les principes de la mécanique quantique (cf. par exemple [29]), la fonction d'onde Ψ_N est de norme 1 dans L^2 , et les quantités physiques sont évaluées dans l'état Ψ_N par des expressions du type

$$\langle A \rangle_{\Psi_N} = \langle \Psi_N | A \Psi_N \rangle_{L^2},$$

où A est un opérateur (éventuellement non borné) sur L^2 (le produit scalaire est ici supposé linéaire par rapport au deuxième vecteur). Par exemple, l'énergie (5) du système n'est autre que la quantité $E_N = \langle H_N \rangle_{\Psi_N}$. On constate que toutes ces quantités ne dépendent de Ψ_N qu'à travers l'opérateur de projection orthogonale sur la droite engendrée par Ψ_N ,

$$\rho_N = |\Psi_N\rangle \langle \Psi_N|$$

ou encore l'opérateur de noyau

$$\rho_N(X, Y) = \Psi_N(X) \overline{\Psi_N(Y)}.$$

Nous allons nous restreindre à des opérateurs A bornés n'agissant que sur un nombre fixe k de variables, de sorte que, pour tout $N \geq k$,

$$\langle A \rangle_{\Psi_N} = \text{Tr}(A \rho_{N:k})$$

où $\rho_{N:k}$ est l'opérateur sur $L^2(\mathbf{R}^{3k})$ de noyau

$$(7) \quad \rho_{N:k}(x_1, \dots, x_k; y_1, \dots, y_k) = \int_{\mathbf{R}^{3(N-k)}} \Psi_N(x_1, \dots, x_k, Z) \overline{\Psi_N(y_1, \dots, y_k, Z)} dZ .$$

Compte tenu des hypothèses sur Ψ_N , $\rho_{N:k}$ est un opérateur positif à trace sur $L^2(\mathbf{R}^{3k})$, de trace égale à 1. La suite $(\rho_{N:k})_{N \geq 1}$ admet donc une valeur d'adhérence $\rho^{(k)}$ pour la topologie faible $\sigma(\mathcal{L}^1(L^2(\mathbf{R}^{3k})), \mathcal{K}(L^2(\mathbf{R}^{3k})))$ issue de la dualité entre opérateurs à trace et opérateurs compacts. En d'autres termes, il existe une sous-suite $(\rho_{N_n:k})$ et un opérateur $\rho^{(k)} \in \mathcal{L}^1(L^2(\mathbf{R}^{3k}))$ tels que, pour tout opérateur compact A sur $L^2(\mathbf{R}^{3k})$,

$$\text{Tr}(A \rho_{N_n:k}) \rightarrow \text{Tr}(A \rho^{(k)}) .$$

Un argument d'extraction diagonale assure ainsi l'existence d'une suite $(\rho^{(k)})_{k \geq 1}$. En supposant qu'un tel procédé puisse être réalisé pour tout temps t (ce qui est le cas grâce à un peu d'équicontinuité), une question naturelle est bien sûr de décrire l'évolution d'une telle suite $(\rho^{(k)}(t))$.

La réponse à cette question dépend beaucoup de la donnée initiale $\Psi_{N,0}$. On distingue à ce sujet deux types de particules, correspondant à des propriétés différentes des fonctions d'onde.

a) *Les bosons* : leurs fonctions d'onde sont symétriques en les variables (x_1, \dots, x_N) : pour toute permutation σ de $\{1, \dots, N\}$,

$$\Psi_N(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \Psi_N(x_1, \dots, x_N) .$$

Cette propriété est conservée par le flot $\exp(-itH_N)$. Un exemple typique de donnée initiale qui la vérifie est

$$(8) \quad \Psi_{N,0}(x_1, \dots, x_N) = \psi_0(x_1) \dots \psi_0(x_N) = \psi_0^{\otimes N}(x_1, \dots, x_N) ,$$

où ψ_0 est un élément de $L^2(\mathbf{R}^3)$, de norme 1. La présence du potentiel d'interaction V_1 s'oppose à ce que la structure de produit tensoriel (8) soit conservée par l'évolution. L'objet des travaux présentés ici est de montrer que, sous des hypothèses raisonnables sur ψ_0 et sur V_1 , cette structure réapparaît après le passage à la limite décrit ci-dessus, au sens où il existe, pour tout t , un élément $\psi(t)$ de $L^2(\mathbf{R}^3)$ tel que, pour tout $k \geq 1$,

$$\rho^{(k)}(t, x_1, \dots, x_k; y_1, \dots, y_k) = \psi(t, x_1) \dots \psi(t, x_k) \overline{\psi(t, y_1)} \dots \overline{\psi(t, y_k)} .$$

De plus, l'évolution de $\psi(t)$ est décrite par une équation aux dérivées partielles non linéaire, appelée équation de Hartree.

b) *Les fermions* : leurs fonctions d'onde sont antisymétriques en les variables (x_1, \dots, x_N) : pour toute permutation σ de $\{1, \dots, N\}$,

$$\Psi_N(x_{\sigma(1)}, \dots, x_{\sigma(N)}) = \varepsilon(\sigma) \Psi_N(x_1, \dots, x_N) .$$

Là encore, cette propriété est conservée par le flot $\exp(-itH_N)$. Un exemple typique de donnée initiale qui la vérifie est un « déterminant de Slater » (*cf.* [26])

$$(9) \quad \Psi_{N,0}(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \det(\psi_{j,0}(x_l))_{1 \leq j, l \leq N} ,$$

où $(\psi_{j,0})_{1 \leq j \leq N}$ est un système orthonormé de $L^2(\mathbf{R}^3)$. Là encore, une telle structure n'est pas conservée par l'évolution. De plus, si l'on se donne un système orthonormé $(\psi_{j,0})_{j \geq 1}$ de $L^2(\mathbf{R}^3)$, on montre que, pour tout $k \geq 1$, $\rho_{N:k}(t)$ tend vers 0 pour la topologie faible $\sigma(\mathcal{L}^1, \mathcal{K})$. Néanmoins, sous des hypothèses supplémentaires sur ce système, il existe, pour tout t et pour tout N , un système orthonormé $(\psi_j(t))_{1 \leq j \leq N}$ de $L^2(\mathbf{R}^3)$ dont le déterminant de Slater $\Psi_N^S(t)$ approche bien la dynamique des N corps au sens suivant : si $\rho_N^S(t)$ désigne le projecteur orthogonal sur $\Psi_N^S(t)$, alors, pour tout k , pour tout t ,

$$\mathrm{Tr}(|\rho_{N:k}(t) - \rho_{N:k}^S(t)|) \rightarrow 0$$

quand N tend vers l'infini. Enfin, pour chaque N , le repère mobile $(\psi_j(t))_{1 \leq j \leq N}$ évolue dans $L^2(\mathbf{R}^3)$ selon un système d'équations aux dérivées partielles non linéaires, appelé système de Hartree–Fock.

Dans les deux cas décrits ci-dessus, on a donc réduit la dynamique d'un grand nombre de particules à des équations non linéaires « modèles » sur des fonctions de trois variables. La suite de cet exposé est consacrée à la description de ces équations, ainsi qu'aux résultats de convergence correspondants.

REMARQUE 0.3. — Un cas particulier important de la dynamique à N corps est bien sûr l'étude des fonctions propres (états liés) et des valeurs propres de l'opérateur H_N , en particulier son état fondamental. Nous n'aborderons pas ici l'abondante littérature consacrée à cette question, qui justifierait largement un autre exposé, et renvoyons à l'article de revue de Lieb [20], à l'ouvrage de Catto–Le Bris–Lions [8], ou au cours donné cette année par P.–L. Lions au Collège de France.

Je remercie C. Bardos et F. Golse pour les discussions que nous avons eues sur les travaux présentés dans cet exposé, et pour m'avoir transmis le manuscrit [6]. Je suis reconnaissant à C. Gérard de m'avoir aidé à comprendre les résultats de [14].

1. LE CAS DES BOSONS : ÉNONCÉ DU RÉSULTAT

1.1. Hypothèses sur le potentiel

Dans ce paragraphe et le suivant, on désigne par V_1 une fonction de classe C^∞ de $]0, +\infty[$ dans \mathbf{R} , vérifiant les estimations suivantes : pour tout $m \in \mathbf{N}$, il existe une

constante $C_m > 0$ telle que

$$(10) \quad \forall r \in]0, r_1] , |V_1^{(m)}(r)| \leq \frac{C_m}{r^{m+1}} .$$

Il est clair que $V = V_1/N$ satisfait aux hypothèses du théorème 0.1.

1.2. L'équation de Hartree

Soit ψ_0 une fonction de $H^1(\mathbf{R}^3)$, de norme L^2 égale à 1, et soit $\Psi_{N,0}$ la fonction d'ondes $\psi_0^{\otimes N}$ qui lui est associée par (8). L'énergie totale E_N du système, donnée par (5), vérifie, compte tenu de la normalisation (6), lorsque N tend vers l'infini :

$$\frac{E_N}{N} \rightarrow \int_{\mathbf{R}^3} |\nabla \psi_0(x)|^2 dx + \frac{1}{2} \int_{\mathbf{R}^3 \times \mathbf{R}^3} V_1(|x-y|) |\psi_0(x)|^2 |\psi_0(y)|^2 dx dy .$$

Le second membre de l'identité ci-dessus est une fonction régulière de $\psi_0 \in H^1(\mathbf{R}^3)$, et, en utilisant la forme symplectique

$$\sigma(f, g) = 2 \operatorname{Im} \langle f | g \rangle_{L^2}$$

sur $L^2(\mathbf{R}^3)$, on lui associe un champ hamiltonien dont les courbes intégrales sont les solutions de

$$(11) \quad i \frac{\partial \psi}{\partial t} = -\Delta \psi + W \psi \quad , \quad W(t, x) = \int_{\mathbf{R}^3} V_1(|x-y|) |\psi(t, y)|^2 dy .$$

L'équation (11) est l'équation de Hartree (cf. [17], où le contexte est plutôt celui des fonctions propres). Dans le cas où $V_1(r) = C/r$, le potentiel W est donné par l'équation de Poisson $-\Delta W = 4\pi C |\psi|^2$, de sorte que (11) est également connue sous le nom de système de Schrödinger–Poisson. Le problème de Cauchy pour (11) a été étudié en détail par Ginibre–Velo dans [15], dont nous citons le résultat suivant.

THÉORÈME 1.1 ([15], 1980). — *Pour tout entier $m \geq 1$, pour toute donnée initiale ψ_0 appartenant à $H^m(\mathbf{R}^3)$, il existe une unique solution $\psi \in C(\mathbf{R}, H^m(\mathbf{R}^3))$ de l'équation (11) vérifiant $\psi(0) = \psi_0$.*

Dans le cadre du théorème ci-dessus, on a de plus les lois de conservation

$$\begin{aligned} \|\psi(t)\|_{L^2} &= \text{cste} , \\ \int_{\mathbf{R}^3} |\nabla \psi(t, x)|^2 dx + \frac{1}{2} \int_{\mathbf{R}^3 \times \mathbf{R}^3} V_1(|x-y|) |\psi(t, x)|^2 |\psi(t, y)|^2 dx dy &= \text{cste} . \end{aligned}$$

1.3. Le théorème de convergence

Nous pouvons maintenant énoncer le résultat qui fait l'objet de cet exposé.

THÉORÈME 1.2 ([2], [3], [11]). — *Soit $\psi_0 \in H^2(\mathbf{R}^3)$, et soit $\psi \in C(\mathbf{R}, H^2(\mathbf{R}^3))$ la solution du problème de Cauchy pour l'équation de Hartree (11) avec donnée initiale ψ_0 . Pour tout $N \geq 1$, soit Ψ_N la solution du problème à N corps (3), (6) telle que $\Psi_N(0) = \Psi_{N,0} = \psi_0^{\otimes N}$.*

Pour tout entier $k \geq 1$, pour tout nombre réel t , la suite $(\rho_{N:k}(t))_{N \geq k}$ définie par (7) vérifie, pour tout opérateur compact A sur $L^2(\mathbf{R}^{3k})$,

$$\mathrm{Tr}(A\rho_{N:k}(t)) \rightarrow \langle \psi(t)^{\otimes k} | A \psi(t)^{\otimes k} \rangle_{L^2}$$

lorsque N tend vers l'infini, localement uniformément par rapport à t .

L'équation de Hartree (11) apparaît donc comme l'évolution d'une particule typique de \mathbf{R}^3 , soumise à l'interaction moyenne résultant d'un grand nombre d'autres particules du même type ; c'est un exemple d'équation de champ moyen.

Les premiers résultats proches du théorème 1.2 semblent avoir été obtenus dans le cadre de la théorie quantique des champs, c'est-à-dire avec des données dépendant effectivement d'une infinité de variables : cf. Hepp [18] pour des potentiels réguliers, puis Ginibre–Velo [14]. Dans ce contexte, l'asymptotique peut être comprise comme une limite classique associée à l'hamiltonien de Hartree, la valeur moyenne de l'opérateur de nombre tendant vers l'infini. Néanmoins, le type de données utilisées ne permet pas de couvrir le théorème ci-dessus.

En 1980, une première démonstration est proposée par Spohn [27], dans le cas d'un potentiel borné, avec une donnée initiale $\psi_0 \in L^2(\mathbf{R}^3)$, en passant directement à la limite dans l'expression de ρ_N en série de perturbations. De plus, la convergence de $\rho_{N:k}(t)$ vers $|\psi(t)^{\otimes k}\rangle\langle\psi(t)^{\otimes k}|$ a alors lieu en norme trace (cf. l'appendice de [5] pour une démonstration plus détaillée).

Afin de couvrir des potentiels plus singuliers, Bardos, Golse et Mauser ont récemment proposé dans [3] une autre approche, consistant d'abord à passer à la limite dans le système d'équations vérifié par $(\rho_{N:k})_{1 \leq k \leq N}$. On obtient un système infini d'équations pour toute valeur d'adhérence $(\rho^{(k)})_{k \geq 1}$ dont la suite des projecteurs $(|\psi^{\otimes k}\rangle\langle\psi^{\otimes k}|)_{k \geq 1}$ est solution. Il reste alors à établir un théorème d'unicité pour ce système. Si la première partie de cette stratégie est menée à bien dans [3] pour des potentiels coulombiens (avec d'ailleurs une plus grande généralité sur les données initiales, cf. *infra*), la seconde partie n'y est traitée que sous l'hypothèse V_1 borné. Plus récemment, Erdős et Yau [11] ont complété ce programme en démontrant l'unicité de ce même système dans une classe de suites d'opérateurs à trace qui permet de contrôler les singularités coulombiennes du potentiel.

2. LE CAS DES BOSONS : ESQUISSE DE LA DÉMONSTRATION

2.1. La hiérarchie BBGKY quantique

Écrivons l'équation sur $\rho_N(t) = |\Psi_N(t)\rangle\langle\Psi_N(t)|$ déduite de (3), dite équation de von Neumann,

$$(12) \quad i\partial_t \rho_N = [H_N, \rho_N].$$

On en déduit le système suivant sur les traces partielles $(\rho_{N:k})_{1 \leq k \leq N}$, que nous écrivons pour les noyaux :

$$(13) \quad i\partial_t \rho_{N:k}(t, X, Y) = \\ - (\Delta_X - \Delta_Y) \rho_{N:k}(t, X, Y) + \frac{1}{N} \sum_{1 \leq j < l \leq k} (V_1(|x_j - x_l|) - V_1(|y_j - y_l|)) \rho_{N:k}(t, X, Y) \\ + \frac{N-k}{N} \sum_{1 \leq j \leq k} \int_{\mathbf{R}^3} (V_1(|x_j - z|) - V_1(|y_j - z|)) \rho_{N:k+1}(t, X, z, Y, z) dz, \quad 1 \leq k \leq N-1,$$

où l'on a regroupé les termes en utilisant la symétrie par rapport aux variables. En passant à la limite formellement dans (13) lorsque N tend vers l'infini, on obtient un système infini sur une suite d'inconnues $(\rho^{(k)})_{k \geq 1}$,

$$(14) \quad i\partial_t \rho^{(k)} = -(\Delta_X - \Delta_Y) \rho^{(k)}(t, X, Y) \\ + \sum_{1 \leq j \leq k} \int_{\mathbf{R}^3} (V_1(|x_j - z|) - V_1(|y_j - z|)) \rho^{(k+1)}(t, X, z, Y, z) dz, \quad k \geq 1.$$

Les systèmes (13) et (14) constituent la version quantique de la « hiérarchie BBGKY » (Bogoliubov, Born et Green, Kirkwood, Yvon) introduite à partir des années 1930 en mécanique classique. Nous renvoyons à [16] pour une introduction unifiée aux théories de champ moyen dans les cadres classique et quantique.

Passons à la justification du passage à la limite ci-dessus. Montrons tout d'abord l'existence d'une valeur d'adhérence $(\rho^{(k)})$. Compte tenu du fait que les opérateurs $\rho_{N:k}$ sont positifs de trace 1, le seul point restant à élucider est l'équicontinuité en temps des quantités $\text{Tr}(A\rho_{N:k})$ lorsque N tend vers l'infini, et A décrit les opérateurs compacts. Par un argument facile de densité, on peut supposer que le noyau de A est C^∞ à support compact dans $\mathbf{R}^{3k} \times \mathbf{R}^{3k}$. On utilise alors l'équation (13) pour estimer la dérivée par rapport à t de $\text{Tr}(A\rho_{N:k})$. Seule la contribution du troisième terme au second membre de (13) pose une difficulté, que l'on résout en utilisant l'inégalité de Cauchy-Schwarz

$$|\rho_{N:k+1}(t, X, z, Y, z)| \leq \rho_{N:k+1}(t, X, z, X, z)^{1/2} \rho_{N:k+1}(t, Y, z, Y, z)^{1/2}$$

et le fait que le potentiel d'interaction appartient à $L^2(\mathbf{R}^3) + L^\infty(\mathbf{R}^3)$.

THÉORÈME 2.1 ([3]). — *Supposons de plus qu'il existe une constante B telle que l'énergie totale E_N du système vérifie, pour tout $N \geq 1$,*

$$(15) \quad E_N \leq B N.$$

Alors toute valeur d'adhérence $(\rho^{(k)})$ de $(\rho_{N:k})$ vérifie le système (14).

Remarquons que le résultat ci-dessus concerne des données beaucoup plus générales que le théorème 1.2. Dans le cas d'une donnée du type $\Psi_{N,0} = \psi_0^{\otimes N}$ et d'un potentiel V_1 vérifiant (10), on vérifie aisément que la condition (15) équivaut à $\psi_0 \in H^1(\mathbf{R}^3)$. En

outre, les hypothèses sur le potentiel V_1 peuvent être affaiblies : la démonstration du théorème 2.1 utilise seulement que la fonction V_1 est continue sur $]0, +\infty[$, tend vers 0 à l'infini et vérifie

$$\int_0^1 r^2 (V_1(r))^2 dr < +\infty.$$

PREUVE (esquisse). Compte tenu de la symétrie de Ψ_N par rapport aux N variables x_j , la condition (15) assure que, pour tout j , la norme $L^2(\mathbf{R}^{3N})$ de $\nabla_{x_j} \Psi_N$ est bornée lorsque N tend vers l'infini. Cette estimation induit la propriété de continuité suivante : pour toute fonction $a \in C_0^\infty(\mathbf{R}^{3k} \times \mathbf{R}^{3k})$, le noyau

$$K_N(t, z, w) = \int_{\mathbf{R}^{3k} \times \mathbf{R}^{3k}} a(X, Y) \rho_{N:k+1}(t, X, z, Y, w) dX dY$$

vérifie

$$\sup_{N \geq 1, t \in \mathbf{R}} \int_{\mathbf{R}^3} |K_N(t, z+h, z) - K_N(t, z, z)| dz \leq C|h|.$$

Cette propriété permet de passer à la limite faible dans le troisième terme du second membre de (13), les autres termes ne posant pas de problème.

Enfin, on vérifie facilement qu'une suite $(\rho^{(k)})$ du type $(|\psi^{\otimes k}\rangle\langle\psi^{\otimes k}|)$ est solution de la hiérarchie (14) dès que ψ est solution de l'équation de Hartree (11). Le théorème 1.2 sera donc démontré dès que l'on aura établi un résultat d'unicité du problème de Cauchy pour le système (14).

Remarquons par ailleurs que le système (13) fournit une (autre) introduction naturelle de l'équation de Hartree (11) : si l'on cherche à «fermer» le système (13) dès la première équation en déterminant une courbe $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ dans l'espace des projecteurs de $L^2(\mathbf{R}^3)$ telle que $\rho^{(1)} = \rho$ et $\rho^{(2)} = \rho \otimes \rho$ vérifient la première équation de (13), on obtient l'équation de Hartree pour $\psi(t)$ (à un facteur de phase $e^{i\theta(t)}$ près).

2.2. Un théorème d'unicité dans le cas d'un potentiel borné

Introduisons les opérateurs suivants agissant sur les espaces d'opérateurs à trace :

$$(16) \quad A_k \gamma^{(k)}(X, Y) = -(\Delta_X - \Delta_Y) \gamma^{(k)}(X, Y),$$

$$C_{k,k+1} \gamma^{(k+1)}(X, Y) = \sum_{1 \leq j \leq k} \int_{\mathbf{R}^3} (V_1(|x_j - z|) - V_1(|y_j - z|)) \gamma^{(k+1)}(X, z, Y, z) dz.$$

L'opérateur A_k est non borné sur l'espace de Banach $\mathcal{L}^1(L^2(\mathbf{R}^{3k}))$ et engendre le groupe d'isométries défini par

$$\exp(-itA_k) \gamma^{(k)} = \exp(it\Delta) \gamma^{(k)} \exp(-it\Delta).$$

En général, $C_{k,k+1}$ est un opérateur non borné de $\mathcal{L}^1(L^2(\mathbf{R}^{3k+3}))$ dans $\mathcal{L}^1(L^2(\mathbf{R}^{3k}))$. Néanmoins, si V_1 est une fonction bornée, alors on vérifie sans mal que $C_{k,k+1}$ est borné, avec l'estimation

$$\text{Tr}(|C_{k,k+1} \gamma^{(k+1)}|) \leq 2 \|V_1\|_{L^\infty} k \text{Tr}(|\gamma^{(k)}|).$$

Le résultat d'unicité souhaité est alors un cas particulier du lemme suivant (avec $\nu = 1$) :

LEMME 2.2. — Soit $(E_k)_{k \geq 1}$ une chaîne d'espaces de Banach. Pour tout k , soit A_k un opérateur non borné sur E_k engendrant un groupe à un paramètre d'isométries $\exp(-itA_k)$ sur E_k , et soit $C_{k,k+1}$ un opérateur borné de E_{k+1} dans E_k . On suppose qu'il existe $C > 0$ tel que, pour tout $k \geq 1$,

$$\|C_{k,k+1}\|_{E_{k+1} \rightarrow E_k} \leq Ck .$$

Pour tout k , soit $\gamma^{(k)} \in C(\mathbf{R}, E_k)$ vérifiant $\gamma^{(k)}(0) = 0$; on suppose que la suite $(\gamma^{(k)})$ vérifie le système infini

$$i\partial_t \gamma^{(k)} = A_k \gamma^{(k)} + C_{k,k+1} \gamma^{(k+1)} , \quad k \geq 1$$

et qu'il existe $\nu > 0$ tel que

$$(17) \quad \forall k \geq 1 , \quad \forall t \in \mathbf{R} , \quad \|\gamma^{(k)}(t)\|_{E_k} \leq \nu^k .$$

Alors $\gamma^{(k)}(t) = 0$ pour tout k et pour tout t .

PREUVE. La formule de variation de la constante donne

$$\gamma^{(k)}(t) = -i \int_0^t \exp(-i(t-s)A_k) C_{k,k+1} \gamma^{(k+1)}(s) ds ,$$

et les estimées sur les opérateurs entraînent donc

$$\|\gamma^{(k)}(t)\|_{E_k} \leq Ck \left| \int_0^t \|\gamma^{(k+1)}(s)\|_{E_{k+1}} ds \right| .$$

Pour tout $T > 0$, pour tout $\sigma > 0$, posons

$$M_\sigma(T) = \sup_{|t| \leq T} \sup_{k \geq 1} \frac{\|\gamma^{(k)}(t)\|_{E_k}}{\nu^k (1 + \sigma|t|)^k}$$

qui est fini grâce à l'hypothèse (17). Alors l'estimation ci-dessus conduit à

$$M_\sigma(T) \leq \sup_{|t| \leq T} \sup_{k \geq 1} \frac{C\nu M_\sigma(T)}{(1 + \sigma|t|)^k} \int_0^{|t|} k(1 + \sigma s)^{k+1} ds \leq \frac{C\nu}{\sigma} (1 + \sigma T)^2 M_\sigma(T) ,$$

ce qui, si l'on choisit $\sigma > 4C\nu$ et $T = 1/\sigma$, impose $M_\sigma(T) = 0$, c'est-à-dire $\gamma^{(k)}(t) = 0$ pour tout $k \geq 1$ et pour $|t| < (4C\nu)^{-1}$. Le lemme résulte alors de l'invariance de l'équation par translation en temps.

Le lemme ci-dessus est relié aux versions abstraites du théorème de Cauchy-Kowalevsky développées dans les années 1970 par Ovcyannikov [24], Nirenberg [22], Nishida [23] et Baouendi-Goulaouic [1].

2.3. Le cas d'un potentiel coulombien : choix de la chaîne d'espaces

Si le potentiel V_1 présente une singularité coulombienne, l'opérateur $C_{k,k+1}$ n'est plus borné de $\mathcal{L}^1(L^2(\mathbf{R}^{3k+3}))$ dans $\mathcal{L}^1(L^2(\mathbf{R}^{3k}))$ et la démonstration ci-dessus ne s'applique plus. L'idée d'Erdős et Yau est d'introduire une nouvelle chaîne d'espaces d'opérateurs qui permette de se ramener au lemme 2.2. Introduisons, pour tout j ,

$$S_j = \sqrt{I - \Delta_j}$$

et les «espaces de Sobolev d'opérateurs à trace» suivants,

$$\mathcal{H}^{1,(k)} = \{\gamma^{(k)} \in \mathcal{L}^1(L^2(\mathbf{R}^{3k})), |S_1 \cdots S_k \gamma^{(k)} S_k \cdots S_1| \in \mathcal{L}^1(L^2(\mathbf{R}^{3k}))\}.$$

On vérifie que $\mathcal{H}^{1,(k)}$ est un espace de Banach pour la norme

$$\|\gamma^{(k)}\|_{\mathcal{H}^{1,(k)}} = \text{Tr}(|S_1 \cdots S_k \gamma^{(k)} S_k \cdots S_1|).$$

Il est par ailleurs trivial que A_k engendre un groupe d'isométries de $\mathcal{H}^{1,(k)}$. De plus, Erdős et Yau montrent le lemme suivant :

LEMME 2.3 ([11]). — *Il existe une constante $C > 0$ telle que, pour tout $k \geq 1$, l'opérateur $C_{k,k+1}$ est borné de $\mathcal{H}^{1,(k+1)}$ dans $\mathcal{H}^{1,(k)}$ avec une norme au plus égale à Ck .*

La preuve du lemme 2.3 est un ensemble assez technique de manipulations sur les opérateurs à traces, combinant des inégalités abstraites avec la version suivante, bien connue, de l'inégalité de Hardy :

$$(18) \quad \int_{\mathbf{R}^3} \frac{|\psi(x)|^2}{|x|^2} dx \leq 4 \int_{\mathbf{R}^3} |\nabla \psi(x)|^2 dx.$$

2.4. Le cas d'un potentiel coulombien : régularisation et propagation des estimées

Pour achever la démonstration du théorème 1.2 en appliquant le lemme 2.2, il convient de vérifier que les deux suites $(\gamma^{(k)}) = (\rho^{(k)})$ et $\gamma^{(k)} = (|\psi^{\otimes k}\rangle\langle\psi^{\otimes k}|)$ satisfont à l'hypothèse (17) de croissance géométrique des normes. Dans le cas de la seconde suite, cela ne pose pas de problème dès que $\psi_0 \in H^1(\mathbf{R}^3)$. Le cas de la première suite est beaucoup plus délicat, car il faut propager des estimations dans $\mathcal{H}^{1,(k)}$ par le flot du système (13). La démarche la plus naturelle consiste bien sûr à remplacer l'opérateur $S_1 \cdots S_k$ par une fonction de H_N , et un premier pas dans cette direction est l'inégalité suivante, que l'on déduit sans mal de la symétrie de ρ_N par rapport aux variables :

$$(19) \quad \|\rho_{N:k}(t)\|_{\mathcal{H}^{1,(k)}} \leq 2^k \text{Tr}\left(\left(I - \frac{\Delta}{N}\right)^k \rho_N(t)\right), \quad 2k \leq N.$$

D'un autre côté, à $t = 0$, le second membre de (19) est plus délicat à estimer, car il fait apparaître des termes du type $\Delta^k \psi_0$, dont l'estimation nécessite une trop grande régularité pour ψ_0 . Pour pallier cet inconvénient, Erdős et Yau régularisent la donnée ψ_0 par le noyau de la chaleur $\exp(\kappa\Delta)$, où $\kappa = \kappa(N)$ doit être choisi assez grand pour

permettre d'estimer le second membre de (19), et assez petit pour que l'erreur commise sur les données disparaisse par passage à la limite. C'est cette double contrainte qui impose finalement la régularité additionnelle $\psi_0 \in H^2(\mathbf{R}^3)$. En introduisant, pour tout $\delta > 0$, la solution Ψ_N^δ de (3) associée à la donnée initiale régularisée

$$\Psi_{N,0}^\delta = (e^{\delta\Delta/N} \psi_0)^{\otimes N},$$

on vérifie les estimées suivantes :

LEMME 2.4 ([11]). — *Il existe $C > 0$ et, pour tous $\delta > 0, k \geq 1$, il existe $N(\delta, k)$ tel que, pour tout $N \geq N(\delta, k)$,*

$$\left\langle \left(I - \frac{\Delta}{N} \right)^k \Psi_{N,0}^\delta \middle| \Psi_{N,0}^\delta \right\rangle_{L^2(\mathbf{R}^{3N})} \leq (C \|\psi_0\|_{H^2})^k.$$

De plus,

$$(20) \quad \sup_{t \in \mathbf{R}} \|\Psi_N^\delta(t) - \Psi_N(t)\|_{L^2(\mathbf{R}^{3N})} \leq \delta \|\psi_0\|_{H^2}.$$

Il faut maintenant propager les estimées sur Ψ_N^δ . On est alors conduit à comparer les puissances de $(I - \Delta/N)$ à celles de $(I + H_N/N)$, ce qui nécessite d'estimer des dérivées d'ordre arbitraire de V_1 . C'est à ce stade qu'intervient pleinement l'hypothèse (10). La singularité à l'origine impose alors de tronquer le potentiel V_1 en le remplaçant par

$$V_1^\varepsilon(r) = \theta\left(\frac{\sqrt{N}r}{\varepsilon}\right) V_1(r),$$

où θ est une fonction de classe C^∞ valant 0 près de $r = 0$, et valant 1 pour $r \geq 1$. On note H_N^ε l'hamiltonien correspondant, et $\Psi_N^{\delta,\varepsilon}$ la solution correspondante, avec donnée initiale $\Psi_{N,0}^\delta$. On démontre alors le lemme de perturbation suivant :

LEMME 2.5 ([11]). — *Il existe $\beta > 0, \tilde{C} > 0$ et, pour tous $\varepsilon > 0, k \geq 1$, il existe un entier $\tilde{N}(\varepsilon, k)$ tel que, pour tout $N \geq \tilde{N}(\varepsilon, k)$,*

$$\tilde{C}^{-k} \left(\beta I - \frac{\Delta}{N} \right)^k \leq \left(\beta I + \frac{H_N^\varepsilon}{N} \right)^k \leq \tilde{C}^k \left(\beta I - \frac{\Delta}{N} \right)^k.$$

De plus, il existe $B > 0$ tel que, pour tout $t \in \mathbf{R}$,

$$(21) \quad \sup_{\delta > 0} \limsup_{N \rightarrow \infty} \|\Psi_N^{\delta,\varepsilon}(t) - \Psi_N^\delta(t)\|_{L^2}^2 \leq B(1 + \|\psi_0\|_{H^2}^2) \varepsilon |t|.$$

La démonstration du théorème 1.2 peut alors se conclure ainsi : à δ, ε fixés, on passe à la limite dans le système (13) tronqué, selon le théorème 2.1, ou plutôt sa variante avec le potentiel tronqué V_1^ε . On constate que la troncature et la régularisation de la donnée, qui dépendent de N , disparaissent par passage à la limite quand N tend vers l'infini, de sorte que la valeur d'adhérence $(\rho^{\delta,\varepsilon,(k)})$ ainsi obtenue est solution de la hiérarchie BBGKY (14) avec la condition initiale $(|\psi_0^{\otimes k}\rangle\langle\psi_0^{\otimes k}|)$, tout en vérifiant les estimations de croissance géométrique (17) dans la chaîne $(\mathcal{H}^{1,(k)})$, grâce aux lemmes 2.4 et 2.5. Le lemme 2.2 assure donc que $\rho^{\delta,\varepsilon,(k)}(t) = |\psi(t)^{\otimes k}\rangle\langle\psi(t)^{\otimes k}|$ pour tout t , et les estimations d'erreur (20) et (21) permettent de passer à la limite quand ε et δ tendent vers 0.

3. LE CAS DES FERMIONS

Les résultats dans ce cas sont plus récents et, dans le cas coulombien, n'ont probablement pas atteint leur forme définitive. Aussi nous contentons-nous d'en donner un bref aperçu.

3.1. Les déterminants de Slater

L'espace des phases fermionique est le sous-espace fermé $L_a^2(\mathbf{R}^{3N})$ de $L^2(\mathbf{R}^{3N})$ constitué des fonctions d'ondes antisymétriques en les N variables x_1, \dots, x_N de \mathbf{R}^3 ; il s'identifie à $\bigwedge^N(L^2(\mathbf{R}^3))$. Pour toute permutation σ de $\{1, \dots, N\}$, on désigne par U_σ l'opérateur unitaire de $L^2(\mathbf{R}^{3N})$ défini par

$$U_\sigma \Psi(x_1, \dots, x_N) = \Psi(x_{\sigma(1)}, \dots, x_{\sigma(N)}),$$

de sorte que le projecteur orthogonal sur $L_a^2(\mathbf{R}^{3N})$ est $P_N = (N!)^{-1} \Sigma_N$, avec

$$\Sigma_N = \sum_{\sigma} \varepsilon(\sigma) U_\sigma.$$

Si (ψ_1, \dots, ψ_N) est un système orthonormé de $L^2(\mathbf{R}^3)$, on normalise l'élément $\psi_1 \wedge \dots \wedge \psi_N$ de $L_a^2(\mathbf{R}^{3N})$ en posant

$$(22) \quad S(\psi_1, \dots, \psi_N)(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \det(\psi_j(x_k))_{1 \leq j, k \leq N}.$$

Les références historiques pour l'introduction de ces quantités comme *Ansatz* de l'équation de Schrödinger stationnaire sont Fock [13] et Slater [26]. Les traces partielles du projecteur orthogonal associé

$$\rho_N^S = |S(\psi_1, \dots, \psi_N)\rangle \langle S(\psi_1, \dots, \psi_N)|$$

sont données par les expressions suivantes :

$$(23) \quad \rho_{N:1}^S = \frac{1}{N} \sum_{j=1}^N |\psi_j\rangle \langle \psi_j| \quad , \quad \rho_{N:k}^S = \frac{N^k (N-k)!}{N!} (\rho_{N:1}^S)^{\otimes k} \Sigma_k .$$

Ces formules inspirent la définition suivante, introduite dans [4].

DÉFINITION 3.1. — *Pour tout N , soit ρ_N un opérateur positif sur $L^2(\mathbf{R}^{3N})$ de trace 1, et qui commute aux opérateurs U_σ . On dit que la suite (ρ_N) vérifie la propriété de fermeture de Slater si, pour tout entier $k \geq 1$, quand N tend vers l'infini,*

$$\mathrm{Tr}(|\rho_{N:k} - (\rho_{N:1})^{\otimes k} \Sigma_k|) \rightarrow 0 .$$

3.2. Le système de Hartree-Fock

Revenons à la hiérarchie (13) et cherchons cette fois un système orthonormé mobile de N vecteurs $(\psi_1(t), \dots, \psi_N(t))$ tel que

$$\rho^{(1)}(t) = (|S(\psi(t), \dots, \psi_N(t))\rangle\langle S(\psi(t), \dots, \psi_N(t))|)_{:1} \quad , \quad \rho^{(2)}(t) = (\rho^{(1)}(t) \otimes \rho^{(1)}(t))\Sigma_2$$

vérifie la première équation de (13),

$$i\partial_t \rho^{(1)}(t, x, y) = -(\Delta_x - \Delta_y)\rho^{(1)}(t, x, y) + \int_{\mathbf{R}^3} (V_1(|x-z|) - V_1(|y-z|))\rho^{(2)}(t, x, z, y, z) dz .$$

Après une renormalisation convenable, on aboutit au système d'évolution de Hartree-Fock, introduit par Dirac [12],

$$(24) \quad i\partial_t \psi_j = -\Delta \psi_j + W \psi_j - \frac{1}{N} \sum_{l=1}^N W_{jl} \psi_l \quad , \quad 1 \leq j \leq N,$$

$$W(t, x) = \int_{\mathbf{R}^3} V_1(|x-z|) \frac{1}{N} \sum_{l=1}^N |\psi_l(t, z)|^2 dz \quad , \quad W_{jl}(t, x) = \int_{\mathbf{R}^3} V_1(|x-z|) \psi_j(t, z) \overline{\psi_l(t, z)} dz .$$

Notons qu'au potentiel d'interaction moyenne W déjà présent dans l'équation de Hartree (11) s'est ajouté un «potentiel d'échange» W_{jl} . On peut également obtenir le système (24) comme une évolution hamiltonienne sur $L^2(\mathbf{R}^3)^N$ associée à l'énergie

$$(25) \quad E_N^{HF} = \sum_{j=1}^N \int_{\mathbf{R}^3} |\nabla \psi_j|^2 dx + \frac{1}{2N} \int_{\mathbf{R}^3 \times \mathbf{R}^3} V_1(|x-z|) \left(\left(\sum_j |\psi_j(x)|^2 \right) \left(\sum_j |\psi_j(z)|^2 \right) - \left| \sum_j \psi_j(x) \overline{\psi_j(z)} \right|^2 \right) dx dz .$$

Le problème de Cauchy pour (24) ne pose pas de difficultés si V_1 est borné. Dans le cas d'un potentiel coulombien, il a été étudié par Chadam et Glassey [10] puis par Bove, Da Prato et Fano [7] dans le formalisme des opérateurs.

THÉORÈME 3.2 ([10], [7]). — *On suppose V_1 coulombien. Pour tout entier $m \geq 1$, pour tout système orthonormé de $L^2(\mathbf{R}^3)$ $(\psi_{1,0}, \dots, \psi_{N,0})$ constitué de fonctions de $H^m(\mathbf{R}^3)$, il existe une unique solution $(\psi_1, \dots, \psi_N) \in C(\mathbf{R}, H^m(\mathbf{R}^3)^N)$ au système (3.2) avec la donnée initiale $(\psi_{1,0}, \dots, \psi_{N,0})$ en $t = 0$. De plus, le système $(\psi_1(t), \dots, \psi_N(t))$ est orthonormé pour tout t , et l'énergie (25) est constante au cours du temps.*

3.3. Le théorème d'approximation

Le résultat suivant est dû à Bardos, Golse, Gottlieb et Mauser.

THÉORÈME 3.3 ([4]). — *Soit $(\psi_{j,0})_{j \geq 1}$ un système orthonormé de $L^2(\mathbf{R}^3)$. Pour tout $N \geq 1$, on désigne par Ψ_N la solution de l'équation de Schrödinger (3) ayant pour*

donnée initiale $\Psi_{N,0} = S(\psi_{1,0}, \dots, \psi_{N,0})$. On note par ailleurs $(\psi_1^{(N)}, \dots, \psi_N^{(N)})$ la solution du système de Hartree–Fock (24) avec la donnée initiale $(\psi_{1,0}, \dots, \psi_{N,0})$, et $\Psi_N^S(t) = S(\psi_1^{(N)}(t), \dots, \psi_N^{(N)}(t))$ le déterminant de Slater correspondant. Alors les projecteurs $\rho_N(t) = |\Psi_N(t)\rangle\langle\Psi_N(t)|$ et $\rho_N^S(t) = |\Psi_N^S(t)\rangle\langle\Psi_N^S(t)|$ vérifient, pour tout $k \geq 1$, lorsque N tend vers l’infini,

$$\text{Tr}(|\rho_{N:k}(t) - \rho_{N:k}^S(t)|) \rightarrow 0.$$

La démonstration de ce théorème est basée sur une analyse du second membre du système (13) pour la différence $\rho_{N:k} - \rho_{N:k}^S$, combinée avec une adaptation de la démonstration du lemme d’unicité (2.2). Pour $|t|$ au plus de l’ordre de l’inverse de $\|V_1\|_{L^\infty}$, on peut estimer la norme trace de la différence $\rho_{N:k}(t) - \rho_{N:k}^S(t)$ par ν^k/N (cf [5]), mais ces estimations se dégradent vite au cours du temps.

La même approche montre, plus généralement, que la propriété de fermeture de Slater est conservée par la hiérarchie (13).

Dans le cas d’un potentiel coulombien répulsif, un résultat tout récent des mêmes auteurs [6] établit, en adaptant les techniques d’Erdős et Yau, un résultat analogue, mais sous des hypothèses beaucoup plus fortes sur les données initiales. Pour tout $m \geq 1$, on demande que

$$\sup_{N \geq 1} \frac{1}{N} \sum_{j=1}^N \|\psi_{j,0}\|_{H^m(\mathbf{R}^3)}^2 < +\infty.$$

4. REMARQUES ET QUESTIONS OUVERTES

4.1. Le problème de la régularité des données

Le théorème de Kato 0.1 assure l’existence d’un groupe à un paramètre unitaire sur $L^2(\mathbf{R}^{3N})$. Par ailleurs, on peut aussi vérifier (bien que cela ne soit pas explicitement écrit dans [15]) que l’équation de Hartree (11) est bien posée dans $L^2(\mathbf{R}^3)$. L’idée est d’utiliser les propriétés dispersives du flot de Schrödinger sur \mathbf{R}^3 à travers les inégalités suivantes, dites de Strichartz (voir par exemple [9]),

$$\left(\int_{\mathbf{R}} \|e^{it\Delta} f\|_{L^q(\mathbf{R}^3)}^p dt \right)^{1/p} \leq C \|f\|_{L^2(\mathbf{R}^3)}$$

pour tout couple (p, q) de nombres réels vérifiant

$$\frac{2}{p} + \frac{3}{q} = \frac{3}{2}, \quad p \geq 2,$$

en s’inspirant d’une démonstration analogue donnée par Tsutsumi [28] pour une équation de Schrödinger avec une perturbation non linéaire locale.

Il est donc naturel de se demander si le théorème 1.2 reste vrai pour des données $\psi_0 \in L^2(\mathbf{R}^3)$ (ce qui est déjà connu, comme on l’a dit, si le potentiel est borné). Dans

le cas coulombien, la réponse à cette question nécessiterait sans doute de comprendre ce que deviennent les inégalités de Strichartz sur la solution Ψ_N et sur les opérateurs $\rho_{N:k}$, et en quoi elles peuvent éviter le recours à la chaîne d'espaces $(\mathcal{H}^{1,(k)})$, qui est une grosse consommatrice de dérivées... En outre, cette approche pourrait se révéler utile dans le cas des fermions, les inégalités de Strichartz pouvant également s'appliquer au système de Hartree-Fock.

4.2. Principe d'exclusion de Pauli et énergie d'interaction

Dans le cas d'un potentiel coulombien répulsif, la pertinence de l'approximation par le système de Hartree-Fock se heurte à la remarque suivante, qui nous a été communiquée par C. Bardos et F. Golse. Si (ψ_1, \dots, ψ_N) est un système orthonormé de $L^2(\mathbf{R}^3)$ un résultat de Lieb–Thirring [21] assure que la densité moyenne

$$\tilde{\rho}(x) = \frac{1}{N} \sum_{j=1}^N |\psi_j(x)|^2$$

est contrôlée à l'aide de l'énergie cinétique moyenne

$$T = \frac{1}{N} \sum_{j=1}^N \int_{\mathbf{R}^3} |\nabla \psi_j|^2 dx$$

selon l'inégalité

$$\int_{\mathbf{R}^3} \tilde{\rho}(x)^{5/3} dx \leq A N^{-2/3} T$$

où A est une constante universelle. En combinant cette information avec l'inégalité de Hardy–Littlewood–Sobolev, on constate que

$$\int_{\mathbf{R}^3 \times \mathbf{R}^3} \frac{\tilde{\rho}(x) \tilde{\rho}(y)}{|x-y|} dx dy \leq B \|\tilde{\rho}\|_{L^{6/5}}^2 \leq B N^{-1/3} T^{1/2}.$$

Il en résulte que, si l'énergie moyenne de Hartree–Fock E_N^{HF}/N est bornée, l'énergie d'interaction tend uniformément vers 0 comme $N^{-1/3}$. En reportant cette information dans le système (3.2), on en déduit que la solution $(\psi_1^{(N)}, \dots, \psi_N^{(N)})$ de ce système est approchée pour tout temps par la solution de l'équation de Schrödinger *libre* selon

$$\frac{1}{N} \sum_{j=1}^N \|\psi_j^{(N)}(t) - e^{it\Delta} \psi_{j,0}\|_{L^2}^2 \leq C N^{-1/3}.$$

Dans ce cas, la pertinence du système de Hartree–Fock par rapport au système de Schrödinger libre ne serait confirmée que si l'approximation établie dans la version coulombienne du théorème 3.3 était d'ordre supérieur. Ce fait ne semble pas être une conséquence des résultats de [6].

Une autre interprétation de ce phénomène pourrait être que la normalisation (6), si elle est justifiée dans le cas des bosons, devient trop brutale dans le cadre du principe

d'exclusion de Pauli, et qu'il conviendrait de la remplacer par une normalisation mieux adaptée aux interactions coulombiennes des fermions.

BIBLIOGRAPHIE

- [1] M.S. BAOUENDI, C. GOULAOUIC – *Remarks on the abstract form of nonlinear Cauchy–Kovalevsky theorems*. Comm. Partial Differential Equations **2** (1977), 1151–1162.
- [2] C. BARDOS, L. ERDÖS, F. GOLSE, N. MAUSER, H.T. YAU – *Derivation of the Schrödinger–Poisson equation from the quantum N -body problem*. C. R. Acad. Sci. Paris, Ser. I **334** (2002), 515–520.
- [3] C. BARDOS, F. GOLSE, N. MAUSER – *Weak coupling limit of the N -particle Schrödinger equation*. Methods and Appl. of Analysis **7** (2000), 275–293.
- [4] C. BARDOS, F. GOLSE, A. GOTTLIEB, N. MAUSER – *Mean field dynamics of fermions and the time-dependent Hartree-Fock equation*. J. Math. Pures Appl. **82** (2003), 665–683.
- [5] C. BARDOS, F. GOLSE, A. GOTTLIEB, N. MAUSER – *Accuracy of the time-dependent Hartree-Fock approximation for uncorrelated initial states*, J. Stat. Phys. **115** (2004), à paraître.
- [6] C. BARDOS, F. GOLSE, A. GOTTLIEB, N. MAUSER – *Derivation of the time-dependent Hartree–Fock equation with Coulomb potential*, manuscrit, 2004.
- [7] A. BOVE, G. DA PRATO, G. FANO – *On the Hartree-Fock time-dependent problem*. Comm. Math. Phys. **49** (1976), 25–33.
- [8] I. CATTO, C. LE BRIS, P.–L. LIONS – *Mathematical theory of thermodynamic limits: Thomas–Fermi type models*. Oxford Mathematical Monographs, Clarendon Press, 1998.
- [9] T. CAZENAVE – *Semilinear Schrödinger Equations*. Courant Lecture Notes in Mathematics, **10**, American Mathematical Society, 2003.
- [10] G. CHADAM, R. GLASSEY – *Global existence of solutions to the Cauchy problem for time-dependent Hartree–Fock equations*. J. Math. Phys. **16** (1975), 1122–1130.
- [11] L. ERDÖS, H.T. YAU – *Derivation of the nonlinear Schrödinger equation from a many body Coulomb system*. Adv. Theor. Math. Phys. **5** (2001), 1169–1205.
- [12] P.A.M. DIRAC – *Note on exchange phenomena in the Thomas atom*. Proc. Cambridge Philos. Soc. **26** (1930), 376–385.
- [13] V. FOCK – *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*. Z. Phys. **61** (1930), 126–148.

- [14] J. GINIBRE, G. VELO – *The classical field limit of scattering theory for non-relativistic many-bosons systems*, I–II. Commun. Math. Phys. **66** (1979), 37–76, et **68** (1979), 45–68.
- [15] J. GINIBRE, G. VELO – *On a class of nonlinear Schrödinger equations with nonlocal interactions*. Math. Z. **170** (1980), 109–145.
- [16] F. GOLSE – *The mean-field limit for the dynamics of large particle systems*. Actes des Journées *Équations aux dérivées partielles*, Forges-les-Eaux, 2–6 juin 2003, GDR 2434 du CNRS. Version électronique disponible sur le site : <http://www.math.sciences.univ-nantes.fr/edpa>.
- [17] D. HARTREE – *The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods*. Proc. Cambridge Philos. Soc. **24** (1928), 89–132.
- [18] K. HEPP – *The classical limit for quantum mechanical correlation functions*, Commun. Math. Physics. **35** (1974), 265–277.
- [19] T. KATO – *Fundamental properties of Hamiltonian operators of Schrödinger type*. Trans. Amer. Math. Soc. **70** (1951), 195–201.
- [20] E.H. LIEB – *The stability of matter : from atoms to stars*. Bull. Amer. Math. Soc. **22** (1990), 1–49.
- [21] E.H. LIEB, W.E. THIRRING – *Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities*. Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann, E.H. Lieb, B. Simon, W.E. Thirring editors, Princeton University Press, 1976.
- [22] L. NIRENBERG – *On an abstract form of the nonlinear Cauchy-Kowalewski theorem*. J. Differential Geometry **6** (1972), 561–576.
- [23] T. NISHIDA – *A note on a theorem of Nirenberg*. J. Differential Geometry **12** (1977), 629–633.
- [24] L.V. OVCYANNIKOV – *A nonlinear Cauchy problem in a scale of Banach spaces*. Dokl. Akad. Nauk SSSR **200** (1971) n° 4. Traduction anglaise dans Sov. Math. Dokl. **12** (1971), 1497–1502.
- [25] M. REED, B. SIMON – *Methods of Modern Mathematical Physics*. Vol. I–IV. Academic Press, 1978.
- [26] J.C. SLATER – *A note on Hartree’s method*. Phys. Rev. **35** (1930), 210–211.
- [27] H. SPOHN – *Kinetic equations from Hamiltonian dynamics : Markovian limits*. Reviews of Modern Physics **53** (1980), 569–615.
- [28] Y. TSUTSUMI – *L^2 -solutions for nonlinear Schrödinger equations and nonlinear groups*. Funkcial. Ekvac. **30** (1987), 115–125.

- [29] H. WEYL – *Gruppentheorie und Quantenmechanik*, 1928. Traduction et deuxième édition : *The Theory of Groups and Quantum Mechanics*. Dover, 1950.

Patrick GÉRARD

Université Paris XI

Département de Mathématiques

UMR 8628 du CNRS

Bâtiment 425

F-91405 ORSAY Cédex

E-mail : `patrick.gerard@math.u-psud.fr`

OBSTRUCTIONS AU PRINCIPE DE HASSE ET À L'APPROXIMATION FAIBLE

par Emmanuel PEYRE

Introduction

Face à un système d'équations polynomiales à coefficients entiers, les questions de l'existence d'une solution à coordonnées entières ou rationnelles sont les premières qui viennent à l'esprit. Bien que les travaux de Davis, Putnam, Robinson, Matijacevič et Čudnovskiĭ (cf. [Az] et [Ma2]) sur le dixième problème de Hilbert aient montré que l'existence d'une solution entière ne peut être déterminée de façon algorithmique, le problème analogue sur \mathbf{Q} est toujours ouvert et il est raisonnable de chercher des critères d'existence de solutions rationnelles pour certaines classes de variétés. Une condition nécessaire évidente pour l'existence d'une telle solution est l'existence d'une solution sur le corps des réels et sur tout complété p -adique de \mathbf{Q} . Hasse fut le premier à étudier de manière systématique la réciproque de cette condition nécessaire. En s'appuyant sur les progrès du corps de classes, il put démontrer cette réciproque pour certaines classes de variétés, dont les quadriques [Hasse1], qui avaient été antérieurement étudiées par Legendre et Minkowski [Mi]. Mais ce passage du local au global ne s'étend pas à tout système d'équations. Hasse lui-même donna des contre-exemples à cette réciproque [Hasse2]. Par la suite, ceux-ci se multiplièrent : Reichardt [Re] et Lind [Li] vers 1940 produisirent de manière indépendante une courbe de genre un, intersection de deux quadriques qui n'admet pas de points sur \mathbf{Q} mais en admet sur tous ses complétés (cet exemple est également décrit dans [Ca]). La construction d'autres contre-exemples parmi les variétés géométriquement rationnelles demanda plus de temps, mais Swinnerton-Dyer en exhiba en 1962 au sein des surfaces cubiques non singulières [SD1]. En 1970, dans son exposé au congrès international de Nice [Ma1], Manin décrivit un critère général basé sur la formule de réciprocité de la théorie du corps de classes qui permit d'expliquer tous les contre-exemples antérieurs (cf. [CTKS]). De manière plus précise, supposons que V soit une variété projective, lisse et géométriquement intègre sur \mathbf{Q} et notons $V(\mathbf{A}_{\mathbf{Q}})$ l'espace adélique associé à V , c'est-à-dire l'espace topologique produit $V(\mathbf{R}) \times \prod_p V(\mathbf{Q}_p)$ où p décrit l'ensemble des nombres premiers. Le

principe du critère de Manin est de construire une partie $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}}$ de $V(\mathbf{A}_{\mathbf{Q}})$ qui contient l'adhérence de l'ensemble des points rationnels dans l'espace adélique. Si $V(\mathbf{A}_{\mathbf{Q}}) \neq \emptyset$ mais que $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}}$ est vide, alors V n'admet pas de point rationnel sur \mathbf{Q} , alors qu'elle en admet sur tous ses complétés. Cette construction fournit également des contre-exemples à l'approximation faible : si $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}} \neq V(\mathbf{A}_{\mathbf{Q}})$, alors les points rationnels de V ne sont pas denses dans l'espace adélique (cf. [CTS3]).

Muni de cette construction, on peut alors affiner les questions précédentes de la façon suivante : étant donnée une variété projective, lisse et géométriquement intègre telle que $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}}$ soit non vide,

- la variété V admet-elle un point rationnel ?
- L'ensemble de ces points rationnels est-il dense dans $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}}$?

Ces questions ont été explorées par de nombreux auteurs parmi lesquels on peut citer Borovoi, Colliot-Thélène, Harari, Salberger, Sansuc, Skorobogatov et Swinnerton-Dyer et une réponse positive à la seconde question fut obtenue pour diverses classes de variétés géométriquement rationnelles. Deux types de méthodes se sont révélées particulièrement efficaces. La première dite de descente consiste à construire des morphismes

$$f_i : W_i \rightarrow V$$

de sorte que W_i soit plus simple du point de vue arithmétique et de déduire le résultat pour V des propriétés des W_i . Cette méthode fut systématisée par Colliot-Thélène et Sansuc ([CTS1], [CTS2], [CTS3], [CTS4] et [CTS5]) à l'aide des toseurs universels. La seconde dite de fibration s'applique dans le cas où l'on dispose d'un morphisme

$$p : V \rightarrow B.$$

On cherche alors à déduire le résultat pour V des résultats connus pour les fibres de p .

Tous les experts se doutaient que la non-vacuité de $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}}$ n'entraîne pas toujours celle de $V(\mathbf{Q})$, mais l'obtention d'un tel contre-exemple se révéla difficile. En 1999, dans [Sk3], Skorobogatov fut le premier à exhiber un exemple explicite de variété V vérifiant $V(\mathbf{A}_{\mathbf{Q}})^{\text{Br}} \neq \emptyset$ et ne possédant pas de points rationnels. Ce contre-exemple peut s'expliquer à l'aide d'un analogue non commutatif de l'obstruction de Brauer-Manin dû à Harari et Skorobogatov [HS]. Par ailleurs, Sarnak et Wang en 1995 dans [SW], puis Poonen en 2001 dans [Po] montrent que des conjectures de Lang impliquent l'existence de variétés sans point rationnel qui échappent au critère de Manin et à ses généralisations non abéliennes.

Dans la première partie de ces notes, nous revenons sur les notions de principe de Hasse et d'approximation faible, la seconde partie est consacrée à la description du critère de Brauer-Manin, la troisième aux méthodes utilisées pour montrer la densité des points rationnels dans l'espace $V(\mathbf{A}_K)^{\text{Br}}$ et la quatrième aux contre-exemples à cette densité. Nous terminons par des extensions de cette démarche à d'autres cadres.

1. LE PRINCIPE DE HASSE ET L'APPROXIMATION FAIBLE

1.1. Terminologie

Fixons quelques notations pour la suite de cet exposé.

NOTATIONS 1.1. — Désormais K désigne un corps de nombres, \mathcal{O}_K son anneau des entiers et M_K l'ensemble des places de K . Pour toute place v de K , on note K_v le complété de K pour la topologie définie par v . Si la place v est non archimédienne, on note \mathcal{O}_v l'anneau des entiers de K_v . Si \mathcal{X} est un schéma sur le spectre d'un anneau A et B une A -algèbre commutative, on note $\mathcal{X}(B)$ l'ensemble $\text{Hom}_{\text{Spec}(A)}(\text{Spec}(B), \mathcal{X})$ et \mathcal{X}_B le produit $\mathcal{X} \times_{\text{Spec}(A)} \text{Spec}(B)$. En particulier, si V est une variété sur un corps F , \overline{F} une clôture algébrique de F et F^s la clôture séparable de F dans \overline{F} , $V(F)$ est l'ensemble des points rationnels de V et \overline{V} (resp. V^s) désigne la variété $V_{\overline{F}}$ (resp. V_{F^s}) sur \overline{F} (resp. F^s).

Nous dirons qu'une variété V sur un corps F est une *bonne variété* si elle est projective, lisse et géométriquement intègre. Deux variétés intègres V et W sur F sont *F -birationnellement équivalentes* si un ouvert non vide de V est isomorphe à un ouvert de W . Une variété intègre V est dite *F -rationnelle* si elle est F -birationnellement équivalente à un espace projectif. Une variété géométriquement intègre V est dite *géométriquement rationnelle* si \overline{V} est \overline{F} -rationnelle.

Une bonne variété V sur un corps F est dite *F -rationnellement connexe* s'il existe une variété M , un ouvert non vide U du produit $\mathbf{P}^1 \times M$ et un morphisme $e : U \rightarrow V$ de sorte que l'application induite $U \times_M U \rightarrow V \times_F V$ soit dominante. On dit que V est *géométriquement rationnellement connexe* si \overline{V} est \overline{F} -rationnellement connexe ; autrement dit, il existe une famille de courbes rationnelles sur \overline{V} de sorte que deux points généraux de $V(\overline{F})$ puissent être reliés par une courbe de cette famille.

Si V est une variété irréductible et \mathcal{V} un modèle de V sur $\text{Spec}(\mathcal{O}_K)$, c'est-à-dire que \mathcal{V}_K est isomorphe à V , alors pour toute place non archimédienne v de K , on a une application naturelle $j_v : \mathcal{V}(\mathcal{O}_v) \rightarrow V(K_v)$. L'espace des adèles associé à V est défini comme l'ensemble des $(x_v)_{v \in M_K}$ du produit $\prod_{v \in M_K} V(K_v)$ tels que x_v appartienne à l'image de j_v pour tout v en dehors d'une partie finie de M_K . Cet espace topologique est indépendant du modèle choisi et si V est projective, il coïncide avec le produit.

1.2. Passage du local au global

Si V est une variété sur le corps de nombres K , on a une implication évidente :

$$(1) \quad V(K) \neq \emptyset \quad \implies \quad \forall v \in M_K, \quad V(K_v) \neq \emptyset.$$

L'intérêt de ce critère provient du fait qu'il est algorithmiquement possible de déterminer si la condition

$$\forall v \in M_K, \quad V(K_v) \neq \emptyset$$

est vérifiée ou pas. En effet, si K_v est isomorphe au corps des réels, la vacuité de $V(K_v)$ est une question décidable grâce aux travaux de Tarski et Seidenberg (cf. [Ta] et [Sei]) ; le lemme de Hensel et ses généralisations montrent que pour toute place non archimédienne v de K correspondant à un idéal premier \mathfrak{p} de \mathcal{O}_K , déterminer si $V(K_v)$ est non vide se réduit à étudier des solutions d'équations polynomiales sur $\mathcal{O}_K/\mathfrak{p}^n$ pour un n convenable et, par conséquent, est décidable. Enfin les estimations de Lang-Weil [LW] permettent de montrer que si V n'est pas vide, alors $V(K_v) \neq \emptyset$ pour toute place v en dehors d'une partie finie de M_K que l'on peut déterminer explicitement.

Hasse fut le premier à étudier de façon systématique la réciproque de l'implication (1). Il énonça en particulier le résultat suivant :

THÉORÈME 1.2 (Hasse [Hasse1], Minkowski [Mi]). — *Une forme quadratique $\sum_{i=1}^n a_i x_i^2$ à coefficients dans \mathbf{Q} admet un zéro non trivial dans \mathbf{Q}^n si et seulement si elle en a sur tout complété de \mathbf{Q} .*

Le cas $n = 2$ est élémentaire, le cas $n = 3$ remonte à Legendre. La difficulté de la démonstration réside dans le passage de trois à quatre variables qui utilise le théorème de la progression arithmétique de Dirichlet et la formule de réciprocité de Gauss. Nous suggérons à l'auditeur de cet exposé qui n'aurait jamais lu la démonstration de ce théorème de se reporter à l'article de Hasse [Hasse1] ou au cours d'arithmétique de Serre [Se2]. Ce résultat se généralise à tout corps de nombres.

Une autre famille d'équations considérée par Hasse est celle associée aux normes d'extensions galoisiennes cycliques de corps.

THÉORÈME 1.3 (Hasse [Hasse2]). — *Soit L/K une extension galoisienne cyclique de corps de nombres. Notons $N_{L/K} : L^\times \rightarrow K^\times$ le morphisme de norme. Un élément a de K^\times appartient à son image si et seulement si, pour toute place v de K et toute extension w de v à L , a appartient à l'image de la norme N_{L_w/K_v} .*

Dans ce cas, si (e_1, \dots, e_n) est une base du K -espace vectoriel L , la variété considérée est la variété affine définie par l'équation

$$N_{L/K} \left(\sum_{i=1}^n X_i e_i \right) = a.$$

1.3. Le principe de Hasse et l'approximation faible

Pour une variété V arbitraire sur un corps de nombres K , il est aisé de donner des contre-exemples à la réciproque de (1). Considérons par exemple le polynôme

$$P(X) = (X^2 - 3)(X^2 + 3)(X^2 + 1)(X^2 + 23).$$

Pour tout nombre premier p impair, le groupe $\mathbf{F}_p^\times / \mathbf{F}_p^{\times 2}$ est cyclique d'ordre 2. Par conséquent si $p \geq 5$, au moins un des entiers -1 , 3 ou -3 est un carré modulo p et donc

dans \mathbf{Q}_p . D'autre part -23 est un carré dans \mathbf{Q}_2 et \mathbf{Q}_3 . Enfin 3 est un carré sur \mathbf{R} . Le polynôme P a donc une racine dans chacun des complétés de \mathbf{Q} . Toutefois, il n'a pas de racine sur \mathbf{Q} .

Plus généralement, si $(V_i)_{i \in I}$ est une famille finie de sous-variétés d'une variété V sur le corps de nombres K telle que

$$\forall v \in M_K, \quad \exists i \in I, \quad V_i(K_v) \neq \emptyset$$

et

$$\forall i \in I, \quad \exists v \in M_K, \quad V_i(K_v) = \emptyset,$$

alors $\bigcup_{i \in I} V_i$ est un contre-exemple à la réciproque de (1).

D'autre part, si on considère la surface affine S d'équation

$$Y^2 + Z^2 = -(X^2 - 3)^2(X^2 + 3)(X^2 + 1)(X^2 + 23),$$

cette surface est irréductible, les remarques faites sur les racines de P montrent qu'elle admet un point sur chaque complété de \mathbf{Q} , mais ses seuls points réels sont $(\sqrt{3}, 0, 0)$ et $(-\sqrt{3}, 0, 0)$ qui ne sont pas définis sur \mathbf{Q} . On obtient à nouveau un contre-exemple à la réciproque de (1). Là encore, cet exemple se généralise aisément en considérant des variétés dont les seuls points sur un des complétés sont des points singuliers.

Par contre, si V est une variété géométriquement irréductible, v une place de K et x_0 un point lisse de $V(K_v)$, le théorème des fonctions implicites [Bki, VAR, §1, n°5] assure qu'il existe un homéomorphisme d'un voisinage ouvert de x_0 pour la topologie v -adique sur un ouvert de $K_v^{\dim(V)}$. En particulier, $V(K_v)$ est dense dans V_{K_v} pour la topologie de Zariski. Les arguments qui précèdent ne permettent donc plus de montrer la vacuité de $V(K)$.

Ces considérations amènent aux définitions qui suivent :

DÉFINITION 1.4. — *On dira par la suite qu'une bonne variété V sur le corps de nombres K vérifie le principe de Hasse si et seulement si elle vérifie l'implication suivante*

$$(\forall v \in M_K, \quad V(K_v) \neq \emptyset) \quad \implies \quad V(K) \neq \emptyset.$$

On dira qu'une bonne variété V sur le corps de nombres K vérifie l'approximation faible si et seulement si l'ensemble des points rationnels $V(K)$ est dense dans l'espace topologique produit $\prod_{v \in M_K} V(K_v)$.

REMARQUES 1.5. — (i) Si le produit $\prod_{v \in M_K} V(K_v)$ n'est pas vide, la variété V vérifie l'approximation faible si et seulement si $V(K)$ est dense dans $\prod_{v \in S} V(K_v)$ pour toute partie finie S de M_K .

(ii) Notons qu'avec ces définitions, toute variété qui vérifie l'approximation faible vérifie également le principe de Hasse.

Ces deux propriétés sont des invariants birationnels des bonnes variétés :

PROPOSITION 1.6. — *Soient V et V' deux bonnes variétés sur le corps de nombres K qui sont K -birationnellement équivalentes. Alors V vérifie le principe de Hasse (resp. l'approximation faible) si et seulement s'il en est de même pour V' .*

Cette proposition découle du lemme de Nishimura [Ni], qui assure que l'existence d'un point rationnel est un invariant birationnel des bonnes variétés, et du théorème des fonctions implicites.

1.4. Classes de variétés vérifiant le principe de Hasse et l'approximation faible

L'approximation faible et, par conséquent, le principe de Hasse ont été démontrés pour plusieurs familles de variétés. Nous en donnerons ici deux exemples qui nous semblent particulièrement marquants :

THÉORÈME 1.7. — *Si V est une variété de drapeaux généralisée sur K , c'est-à-dire une variété projective homogène sous l'action d'un groupe algébrique linéaire connexe, alors V vérifie l'approximation faible et le principe de Hasse.*

Ce résultat englobe le résultat de Legendre, Hasse et Minkowski sur les quadriques et celui de Châtelet sur les variétés de Severi-Brauer, c'est-à-dire les K -formes des espaces projectifs. Eichler [Ei], Landherr [Lan], Kneser [Kn], Harder [Harder1] [Harder2] et Chernousov [Che] ont montré le principe de Hasse pour les espaces principaux homogènes sous un groupe algébrique semi-simple simplement connexe. Platonov et Rapinchuk ont prouvé l'approximation faible pour ces groupes [PR, theorem 7.8]. Les travaux de Harder [Harder3, Satz 4.3.3] et Borovoi [Bo1] permettent d'en déduire le théorème qui précède.

Le second résultat que je souhaite mentionner est une conséquence de la méthode du cercle et s'applique aux intersections complètes lisses dans l'espace projectif lorsque la dimension est grande relativement aux degrés des équations.

THÉORÈME 1.8 (Hardy, Littlewood, Davenport, Birch [Bir]). — *Si $V \subset \mathbf{P}_{\mathbf{Q}}^N$ est une intersection complète non singulière définie par des équations f_1, \dots, f_m de même degré d avec*

$$N > 2^{d-1}m(m+1)(d-1),$$

alors V vérifie l'approximation faible et le principe de Hasse.

Pour les cubiques, ce résultat vaut en fait dès que $N \geq 8$ (Heath-Brown [HB] et Hooley [Ho1], [Ho2], cf. l'exposé de Deshouillers [De]).

La méthode du cercle s'applique aussi à des équations de degrés différents (cf. les travaux de Schmidt [Sch]) et fournit également des résultats sur un corps de nombres (Skinner [Skinner]).

2. L'OBSTRUCTION DE BRAUER-MANIN

2.1. Un premier contre-exemple au principe de Hasse

Hasse lui-même savait que le théorème 1.3 qu'il avait démontré pour les extensions de corps cycliques L/K ne se généralisait pas : il connaissait des exemples d'extension galoisienne L/K de groupe de Galois $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ telle que l'équation

$$N_{L/K}(x) = a$$

avec a un élément de K a une solution sur tous les complétés de K mais n'en possède pas sur K .

Le principe de Hasse n'est donc pas toujours vérifié. Je voudrais en donner maintenant un exemple dû à Iskovskih [Is] qui fut revisité par Colliot-Thélène, Coray et Sansuc [CTCS]. On considère la surface S fibrée en coniques définie sur \mathbf{Q} par l'équation

$$(2) \quad Y^2 + Z^2 = (3 - X^2)(X^2 - 2).$$

Rappelons (cf. [Se2, chap. III]) que si v est une place de \mathbf{Q} et si a, b sont deux éléments inversibles de \mathbf{Q}_v , le symbole de Hilbert $(a, b)_v$ est défini par

$$(a, b)_v = \begin{cases} 1 & \text{si } Z^2 - aX^2 - bY^2 = 0 \text{ a une solution non nulle dans } \mathbf{Q}_v^3, \\ -1 & \text{sinon,} \end{cases}$$

que ce symbole est bilinéaire et qu'il vérifie la formule du produit :

$$\forall a, b \in \mathbf{Q}^\times, \quad \prod_{v \in M_{\mathbf{Q}}} (a, b)_v = 1$$

qui découle de la formule de réciprocité quadratique de Gauss. En outre, on dispose des formules explicites suivantes :

- si p est un nombre premier impair et $a \in \mathbf{Q}_p^\times$,

$$(-1, a)_p = (-1)^{v_p(a)\epsilon(p)}$$

où $\epsilon(p)$ est la classe de $(p-1)/2$ modulo 2,

- si $a = 2^k u$ avec $u \in \mathbf{Z}_2^\times$, on a

$$(-1, a)_2 = (-1)^{\epsilon(u)}$$

- et si $a \in \mathbf{R}^\times$,

$$(-1, a)_\infty = \frac{a}{|a|}.$$

Si v est une place de \mathbf{Q} et x un élément de \mathbf{Q}_v tel que $(3-x^2)(x^2-2)$ soit inversible, alors la conique définie par $Y^2 + Z^2 = (3-x^2)(x^2-2)$ admet un point défini sur \mathbf{Q}_v si et seulement si

$$(3) \quad (-1, 3-x^2)_v = (-1, x^2-2)_v.$$

Des expressions qui précèdent on déduit que cette condition est vérifiée pour tout x de \mathbf{Q}_p tel que $(3 - x^2)(x^2 - 2)$ soit inversible si $4 \mid p - 1$. Si $4 \mid p - 3$, cette égalité vaut dès que $v_p(x) < 0$. Pour $p = 2$, elle est vérifiée pour $x = 0$ et sur \mathbf{R} on l'obtient pour $x \in]-\sqrt{3}, -\sqrt{2}[\cup]\sqrt{2}, \sqrt{3}[$. Il en résulte que la surface définie par (2) admet un point lisse sur tout complété de \mathbf{Q} . Supposons que $(x, y, z) \in \mathbf{Q}^3$ soit solution de (2). Notons tout d'abord que $(3 - x^2)(x^2 - 2) \neq 0$. D'autre part, les formules qui précèdent donnent que

- si p est un nombre premier tel que $4 \mid p - 1$, alors $(-1, 3 - x^2) = 1$,
- si p est un nombre premier congru à 3 modulo 4, alors soit $v_p(x) < 0$ auquel cas on obtient $(-1, 3 - x^2) = 1$; soit $v_p(x) \geq 0$, mais la relation $x^2 - 2 = 1 - (3 - x^2)$ fournit $v_p(x^2 - 2) = 0$ ou $v_p(3 - x^2) = 0$ et en utilisant la relation (3), on en déduit que $(-1, 3 - x^2) = 1$.
- Pour $p = 2$, si $v_2(x) > 0$, alors $3 - x^2 \equiv 3 \pmod{4}$ et donc $(-1, 3 - x^2) = -1$; de même si $v_2(x) < 0$, alors $3/x^2 - 1 \equiv 3 \pmod{4}$ et donc $(-1, 3 - x^2)_2 = -1$. Enfin si $v_2(x) = 0$, alors $x^2 \equiv 1 \pmod{8}$ d'où $(3 - x^2)/2 \equiv 1 \pmod{4}$ et $x^2 - 2 \equiv 3 \pmod{4}$, ce qui contredit la relation (3).
- Pour la place archimédienne, la relation $(-1, 3 - x^2)_\infty = (-1, x^2 - 2)_\infty$ entraîne $x \in]-\sqrt{3}, \sqrt{3}[$ et donc $(-1, 3 - x^2)_\infty = 1$.

En définitive, on a obtenu

$$(-1, 3 - x^2)_v = \begin{cases} 1 & \text{si } v \neq 2, \\ -1 & \text{si } v = 2. \end{cases}$$

Mais ceci implique la relation

$$\prod_{v \in M_{\mathbf{Q}}} (-1, 3 - x^2)_v = -1,$$

ce qui contredit la formule du produit. L'équation (2) n'a donc pas de solution sur \mathbf{Q} .

Un modèle projectif et lisse \tilde{S} de S est obtenu en recollant les deux sous-variétés lisses de $\mathbf{A}_{\mathbf{Q}}^1 \times \mathbf{P}_{\mathbf{Q}}^2$ d'équations respectives

$$Y^2 + Z^2 = (3 - X^2)(X^2 - 2)T^2$$

et

$$Y^2 + Z^2 = (3X^2 - 1)(1 - 2X^2)T^2.$$

Cette dernière n'ayant pas de point rationnel pour $X = 0$, on obtient que \tilde{S} ne vérifie pas le principe de Hasse.

Le critère de Brauer-Manin est une généralisation de cet exemple, la formule de réciprocité de la théorie du corps de classes prenant le rôle de la formule du produit pour les symboles de Hilbert.

2.2. Rappels sur le groupe de Brauer cohomologique des variétés

Avant de passer au critère lui-même, nous rappelons ici quelques propriétés du groupe de Brauer cohomologique qui nous seront utiles par la suite (cf. [Ma3, chap. VI]).

NOTATIONS 2.1. — Si V est une variété lisse et géométriquement intègre sur un corps F , on note $\text{Pic}(V)$ le groupe de Picard de V et $\text{Br}(V)$ le groupe de Brauer cohomologique de V , c'est-à-dire le groupe

$$\text{Br}(V) = H_{\text{ét}}^2(V, \mathbf{G}_m),$$

où \mathbf{G}_m désigne le faisceau associé au groupe multiplicatif, qui à une variété U associe $\Gamma(U, \mathcal{O}_U)^\times$. Si V est le spectre d'un corps F , son groupe de Brauer peut être décrit comme un groupe de cohomologie galoisienne

$$\text{Br}(\text{Spec}(F)) = H^2(\text{Gal}(F^s/F), F^{s\times}),$$

où $\text{Gal}(F^s/F)$ désigne le groupe de Galois de l'extension F^s/F . Dans ce cas, ce groupe coïncide avec le groupe des classes d'équivalence d'algèbres simples centrales sur F pour la relation d'équivalence \sim définie par $A \sim B$ si et seulement s'il existe deux entiers strictement positifs m, n et un isomorphisme d'algèbres $M_m(A) \xrightarrow{\sim} M_n(B)$, la loi de groupe étant induite par le produit tensoriel des algèbres simples centrales sur le corps F (cf. [Se1, X, §5]). Si v est une valuation discrète de rang un sur F , de corps résiduel parfait $\kappa(v)$, on dispose d'un morphisme résidu

$$\partial_v : \text{Br}(F) \rightarrow \text{Hom}_{\text{cont}}(\text{Gal}(\overline{\kappa(v)}/\kappa(v)), \mathbf{Q}/\mathbf{Z}).$$

Si F est un corps local non archimédien ce morphisme induit un isomorphisme

$$\text{inv}_v : \text{Br}(F) \xrightarrow{\sim} \mathbf{Q}/\mathbf{Z}$$

(cf [Se1, XII §3]). D'autre part, le groupe $\text{Br}(\mathbf{R})$ est isomorphe à $\mathbf{Z}/2\mathbf{Z}$, engendré par la classe de l'algèbre de quaternions. Pour toute place v du corps de nombres K , on dispose donc d'un morphisme injectif canonique

$$\text{inv}_v : \text{Br}(K_v) \rightarrow \mathbf{Q}/\mathbf{Z}.$$

La théorie du corps de classes global (cf. [NSW, theorem 8.1.17]) fournit alors une suite exacte naturelle :

$$(4) \quad 0 \rightarrow \text{Br}(K) \rightarrow \bigoplus_{v \in M_K} \text{Br}(K_v) \xrightarrow{\sum_{v \in M_K} \text{inv}_v} \mathbf{Q}/\mathbf{Z} \rightarrow 0.$$

Notons au passage que l'injectivité du morphisme de gauche peut se réinterpréter comme le principe de Hasse pour les variétés de Severi-Brauer ; d'autre part, la formule de réciprocité du corps de classes, c'est-à-dire le fait que cette suite forme un complexe, implique la formule du produit pour les symboles de Hilbert.

Décrivons deux méthodes pour construire des éléments dans le groupe de Brauer d'une variété.

Si V est une variété intègre sur un corps F , de corps de fonctions $F(V)$, la contravariance du groupe de Brauer fournit un morphisme

$$\mathrm{Br}(V) \rightarrow \mathrm{Br}(F(V)).$$

Si V est une bonne variété sur un corps F de caractéristique 0, il découle des théorèmes de pureté de Grothendieck [Gr, §6.7] que ce morphisme induit un isomorphisme de $\mathrm{Br}(V)$ sur

$$\bigcap_{v \in \mathcal{P}(F(V)/F)} \ker(\partial_v)$$

où $\mathcal{P}(F(V)/F)$ désigne l'ensemble des valuations discrètes de rang un sur $F(V)$ dont la restriction à F est triviale. On retrouve en particulier que le groupe de Brauer est un invariant birationnel des bonnes variétés (cf. [Gr, corollaire 7.5]).

Si V est une bonne variété sur un corps F de caractéristique nulle, la suite spectrale de Hochschild-Serre pour la cohomologie étale à coefficients dans \mathbf{G}_m donne une suite exacte

$$(5) \quad 0 \rightarrow \mathrm{Pic}(V) \rightarrow \mathrm{Pic}(\overline{V})^{\mathcal{G}} \rightarrow \mathrm{Br}(F) \rightarrow \\ \rightarrow \ker(\mathrm{Br}(V) \rightarrow \mathrm{Br}(\overline{V})) \rightarrow H^1(F, \mathrm{Pic}(\overline{V})) \rightarrow H^3(F, \mathbf{G}_m),$$

où \mathcal{G} désigne le groupe de Galois $\mathrm{Gal}(\overline{F}/F)$. Si V est une bonne variété sur le corps de nombres K , alors $H^3(K, \mathbf{G}_m)$ est nul et on en déduit une suite exacte

$$\mathrm{Br}(K) \rightarrow \ker(\mathrm{Br}(V) \rightarrow \mathrm{Br}(\overline{V})) \rightarrow H^1(K, \mathrm{Pic}(\overline{V})) \rightarrow 0.$$

NOTATIONS 2.2. — Si V est une bonne variété sur un corps F de caractéristique nulle, on note

$$\mathrm{Br}_1(V) = \ker(\mathrm{Br}(V) \rightarrow \mathrm{Br}(\overline{V}))$$

et

$$\mathrm{Br}_0(V) = \mathrm{Im}(\mathrm{Br}(F) \rightarrow \mathrm{Br}(V)).$$

Par abus de langage, un élément de $\mathrm{Br}(V) - \mathrm{Br}_1(V)$ est dit transcendant.

REMARQUE 2.3. — Notons que si V est géométriquement rationnelle, alors $\mathrm{Br}(\overline{V})$ est nul puisque c'est un invariant birationnel et le groupe $\mathrm{Br}(V)$ coïncide avec $\mathrm{Br}_1(V)$. En outre, le groupe $\mathrm{Pic}(\overline{V})$ est alors un \mathbf{Z} -module libre de rang fini (cf. [CTS5, corollaire 2.A.2]) et le groupe $H^1(F, \mathrm{Pic}(\overline{V}))$ est fini. Par conséquent, le groupe $\mathrm{Br}(V)/\mathrm{Br}_0(V)$ est fini dans ce cas.

2.3. Le critère de Brauer-Manin

NOTATIONS 2.4. — Soit V une bonne variété sur le corps de nombres K . Pour toute place v de K et tout point x de $V(K_v)$, la contravariance du groupe de Brauer fournit un morphisme

$$\mathrm{Br}(V) \rightarrow \mathrm{Br}(K_v) \\ A \mapsto A(x_v).$$

On obtient ainsi un accouplement

$$\begin{aligned} \mathrm{Br}(V) \times V(K_v) &\rightarrow \mathbf{Q}/\mathbf{Z} \\ (A, x) &\mapsto \langle A, x \rangle_v = \mathrm{inv}_v(A(x_v)). \end{aligned}$$

On peut montrer (cf [San, lemme 6.2]) que pour tout A de $\mathrm{Br}(V)$, l'application induite de $V(K_v)$ dans \mathbf{Q}/\mathbf{Z} est localement constante. Pour une place réelle, cette application est donc constante sur les composantes connexes. En outre, cette application est triviale pour presque toute place v de K . On obtient donc un accouplement

$$\begin{aligned} \langle \cdot, \cdot \rangle_{\mathbf{A}_K} : \mathrm{Br}(V) \times V(\mathbf{A}_K) &\rightarrow \mathbf{Q}/\mathbf{Z} \\ (A, (x_v)_{v \in M_K}) &\mapsto \sum_{v \in M_K} \langle A, x_v \rangle_v \end{aligned}$$

tel que, pour tout A de $\mathrm{Br}(V)$, l'application induite de $V(\mathbf{A}_K)$ dans \mathbf{Q}/\mathbf{Z} est localement constante.

L'idée cruciale du critère de Manin est alors la suivante : si x est un point rationnel de V sur K et A un élément de $\mathrm{Br}(V)$, alors

$$\langle A, x \rangle_{\mathbf{A}_K} = \sum_{v \in M_K} \langle A, x \rangle_v = \sum_{v \in M_K} \mathrm{inv}_v(A(x)) = 0$$

où la dernière égalité résulte du fait que la suite (4) est un complexe. On en déduit donc que, pour tout point x de l'adhérence de $V(K)$ dans $V(\mathbf{A}_K)$ et pour tout élément A de $\mathrm{Br}(V)$, on a

$$\langle A, x \rangle_{\mathbf{A}_K} = 0.$$

Ceci nous amène aux définitions suivantes :

DÉFINITION 2.5. — *Soit V une bonne variété sur le corps de nombres K . Pour tout élément x de $V(\mathbf{A}_K)$, le morphisme de groupes*

$$\begin{aligned} \mathrm{Br}(V) &\rightarrow \mathbf{Q}/\mathbf{Z} \\ A &\mapsto \langle A, x \rangle_{\mathbf{A}_K} \end{aligned}$$

s'annule sur l'image de $\mathrm{Br}(K)$ et définit par passage au quotient un morphisme ω_x de $\mathrm{Br}(V)/\mathrm{Br}_0(V)$ dans \mathbf{Q}/\mathbf{Z} . On appellera obstruction de Brauer-Manin en x le morphisme ω_x et espace de Brauer-Manin l'ensemble

$$V(\mathbf{A}_K)^{\mathrm{Br}} = \{x \in V(\mathbf{A}_K) \mid \omega_x = 0\}.$$

Les remarques qui précèdent montre que, pour tout point adélique x , l'obstruction de Brauer-Manin en x est une obstruction à ce que x appartienne à l'adhérence des points rationnels. Autrement dit, on a la proposition suivante :

PROPOSITION 2.6. — *Avec les notations qui précèdent, l'adhérence $\overline{V(K)}$ des points rationnels dans l'espace adélique est contenue dans $V(\mathbf{A}_K)^{\mathrm{Br}}$.*

REMARQUE 2.7. — (i) Si A est un élément de $\mathrm{Br}(V)$, l'ensemble

$$\{x \in V(\mathbf{A}_K) \mid \langle A, x \rangle_{\mathbf{A}_K} = 0\}$$

est à la fois ouvert et fermé dans $V(\mathbf{A}_K)$. Par conséquent, si $\mathrm{Br}(V)/\mathrm{Br}_0(V)$ est fini, il en est de même de l'ensemble de $V(\mathbf{A}_K)^{\mathrm{Br}}$. Par la remarque 2.3, c'est le cas si V est géométriquement rationnelle.

(ii) On ne dispose pas de méthode algorithmique pour décider de la vacuité de $V(\mathbf{A}_K)^{\mathrm{Br}}$ en général : pour une courbe de genre 1, cela revient en fait à calculer explicitement le groupe de Tate-Shafarevich de la jacobienne. Par contre, si V est géométriquement rationnelle (ou plus généralement géométriquement rationnellement connexe), alors le groupe quotient $\mathrm{Br}(V)/\mathrm{Br}_0(V)$ est fini et des méthodes sont disponibles pour calculer explicitement $V(\mathbf{A}_K)^{\mathrm{Br}}$.

EXEMPLE 2.8. — Dans l'exemple du paragraphe 2.1, soit \tilde{S} un modèle projectif et lisse de $\mathbf{Q}(S)$ sur \mathbf{Q} , soit A l'algèbre de quaternions $\left(\frac{-1, 3-x^2}{\mathbf{Q}(S)}\right)$ qui est l'algèbre associative unifère engendrée par deux éléments I et J avec les relations $I^2 = -1$, $J^2 = 3 - x^2$ et $IJ = -JI$. La classe de cette algèbre provient de $\mathrm{Br}(\tilde{S})$ et le raisonnement fait montre que

$$\{x \in \tilde{S}(\mathbf{A}_K) \mid \langle [A], x \rangle_{\mathbf{A}_K} = 0\}$$

est vide. Donc $\tilde{S}(\mathbf{A}_K)^{\mathrm{Br}} = \emptyset$ et $\tilde{S}(K) = \emptyset$.

REMARQUE 2.9. — La plupart des contre-exemples au principe de Hasse utilisent le quotient $\mathrm{Br}_1(V)/\mathrm{Br}_0(V)$. Toutefois quelques auteurs ont également fourni des exemples provenant de la partie transcendante du groupe de Brauer (Harari [Ha2] et Wittenberg [Wi]).

ABUS DE LANGAGE 2.10. — Suivant la terminologie couramment utilisée, nous dirons par abus de langage que l'obstruction de Brauer-Manin au principe de Hasse est la seule pour une bonne variété V si on a l'implication

$$V(\mathbf{A}_K)^{\mathrm{Br}} \neq \emptyset \Rightarrow V(K) \neq \emptyset.$$

On dit également que l'obstruction de Brauer-Manin à l'approximation faible est la seule pour V si et seulement si les points rationnels de V sont denses dans $V(\mathbf{A}_K)^{\mathrm{Br}}$.

Enfin si \mathcal{C} est une classe de bonnes variétés, on dit que l'obstruction de Brauer-Manin au principe de Hasse (resp. à l'approximation faible) est la seule pour \mathcal{C} si et seulement si c'est la seule pour toute variété V de \mathcal{C} .

REMARQUE 2.11. — La validité de l'implication

$$V(\mathbf{A}_K)^{\mathrm{Br}} \neq \emptyset \Rightarrow V(K) \neq \emptyset$$

ou la densité de $V(K)$ dans $V(\mathbf{A}_K)^{\mathrm{Br}}$ est un invariant birationnel des bonnes variétés. On peut donc étendre la terminologie précédente aux variétés géométriquement irréductibles sur le corps de nombres K en disant que l'obstruction de Brauer-Manin au principe

de Hasse (resp. à l'approximation faible) est la seule pour une variété géométriquement irréductible X si et seulement si c'est la seule pour un modèle projectif et lisse de $K(X)$.

3. L'OBSTRUCTION DE BRAUER-MANIN EST LA SEULE

Dans ce paragraphe, nous souhaitons décrire deux méthodes particulièrement efficaces pour montrer la densité des points rationnels dans l'espace de Brauer-Manin. Nous terminerons par une liste de cas pour lesquels cette densité a été démontrée.

3.1. Techniques de descente

Inspirée du travail de Châtelet [Ch] et de la méthode de descente classique pour les courbes de genre 1, la méthode de descente a été développée par Colliot-Thélène et Sansuc avec l'introduction de la notion de torseur universel d'une variété géométriquement rationnelle.

Si F est un corps algébriquement clos et V une bonne variété sur F dont le groupe de Picard est un \mathbf{Z} -module libre de rang fini, alors un torseur universel sur V peut être construit de la façon suivante : soient L_1, \dots, L_t des fibrés en droites sur V dont les classes forment une base du groupe de Picard de V . Pour $i \in \{1, \dots, t\}$, notons L_i^\times le complémentaire de la section nulle dans L_i et posons

$$\mathcal{T} = L_1^\times \times_V \cdots \times_V L_t^\times.$$

Le groupe algébrique \mathbf{G}_m^t agit sur \mathcal{T} et l'application naturelle $\pi : \mathcal{T} \rightarrow V$ fait de V un quotient de \mathcal{T} sous l'action de ce groupe. En outre, à isomorphisme près, les structures obtenues ne dépendent pas des choix effectués.

Dans le cas général, nous allons définir les torseurs universels comme solution à un problème universel.

DÉFINITIONS 3.1. — *On dit qu'un groupe algébrique G sur un corps F est de type multiplicatif si et seulement s'il existe un entier n tel que \overline{G} soit isomorphe à un sous-groupe fermé de $\mathbf{G}_{m, \overline{F}}^n$. Notons $\mathcal{G} = \text{Gal}(F^s/F)$ le groupe de Galois absolu de F . On a une équivalence de catégorie contravariante entre la catégorie des groupes de type multiplicatif et la catégorie des \mathbf{Z} -modules de type fini munis d'une action continue de \mathcal{G} , c'est-à-dire se factorisant via un quotient fini de \mathcal{G} . Cette équivalence associe à un groupe G le groupe des caractères de G*

$$X^*(G) = \text{Hom}_{F^s}(G^s, \mathbf{G}_{m, F^s}).$$

Si G est un groupe algébrique sur un corps F et X une variété sur F , on appelle torseur sur X sous G ou espace principal homogène sur X sous G la donnée d'une variété \mathcal{T} sur F munie d'un morphisme $\pi : \mathcal{T} \rightarrow X$ fidèlement plat et d'une action de G au-dessus de X de sorte que, localement pour la topologie plate, \mathcal{T} soit isomorphe au produit $G \times X$.

Autrement dit, si $m : G \times \mathcal{T} \rightarrow \mathcal{T}$ définit l'action de G , on a l'égalité $\pi \circ m = \pi \circ \text{pr}_2$ et l'application

$$\rho : G \times_F \mathcal{T} \rightarrow \mathcal{T} \times_X \mathcal{T}$$

définie comme la composée des applications

$$G \times_F \mathcal{T} \xrightarrow{\text{Id} \times \delta} G \times_F \mathcal{T} \times_X \mathcal{T} \xrightarrow{m \times \text{Id}} \mathcal{T} \times_X \mathcal{T}$$

où δ désigne la diagonale, est un isomorphisme.

Si X est une variété lisse et géométriquement intègre sur F telle que $\Gamma(\overline{X}, \mathcal{O}_{\overline{X}})^\times = \overline{F}^\times$ et si x est un point rationnel de X , un torseur universel au-dessus de (X, x) est une paire (\mathcal{T}, t) où \mathcal{T} est un torseur sur X sous un groupe de type multiplicatif T et t un point rationnel de \mathcal{T} au-dessus de x vérifiant en outre la propriété universelle suivante : pour toute paire (\mathcal{T}', t') où \mathcal{T}' est un torseur sur X sous un groupe de type multiplicatif T' et t' un point rationnel de \mathcal{T}' au-dessus de x , il existe un unique morphisme de groupes $\phi : T \rightarrow T'$ et un unique morphisme $\psi : \mathcal{T} \rightarrow \mathcal{T}'$ au-dessus de X , compatible avec l'action de T sur \mathcal{T} et de T' sur \mathcal{T}' et tel que $\psi(t) = t'$.

EXEMPLE 3.2. — Si G est un groupe semi-simple, et \tilde{G} le revêtement universel de G , alors (\tilde{G}, e) est un torseur universel au-dessus de (G, e) (cf [Sk4, §3.2]).

Expliquons maintenant comment obtenir de tels torseurs. Supposons que X soit une bonne variété sur F ; alors, pour tout groupe de type multiplicatif T de groupe de caractères $X^*(T)$, les classes d'isomorphismes de torseurs sur X sous T sont en bijection avec $H_{\text{pl}}^1(X, T)$. Colliot-Thélène et Sansuc construisent alors une suite exacte canonique [CTS5, (2.0.2)]

$$0 \rightarrow H^1(F, T) \rightarrow H_{\text{pl}}^1(X, T) \xrightarrow{\rho} \text{Hom}_{\mathcal{G}}(X^*(T), \text{Pic}(\overline{X})) \xrightarrow{\delta} H^2(F, T).$$

DÉFINITIONS 3.3. — Soit X une bonne variété sur un corps F de caractéristique 0. Si le groupe de Picard géométrique $\text{Pic}(\overline{X})$ de X est un \mathbf{Z} -module de type fini, alors on note T_{NS} le groupe de type multiplicatif qui lui est associé par l'équivalence de catégorie ci-dessus. Un torseur versel est un torseur \mathcal{T} sur X sous T_{NS} dont la classe $[\mathcal{T}]$ dans $H_{\text{pl}}^1(X, T_{\text{NS}})$ est telle que $\rho([\mathcal{T}])$ soit l'isomorphisme de $X^*(T_{\text{NS}})$ sur $\text{Pic}(\overline{X})$ donné par l'équivalence de catégorie.

REMARQUE 3.4. — Sous les hypothèses précédentes, un torseur versel muni d'un point rationnel t au-dessus de x est un torseur universel au-dessus de (X, x) . Nous avons choisi ici de distinguer la notion de torseur universel, solution du problème universel, de celle de torseur versel, ce qui nous amène à diverger de la terminologie usuelle introduite par Colliot-Thélène et Sansuc pour lesquels un torseur versel est dit universel.

Colliot-Thélène et Sansuc montrent que, sous les hypothèses qui précèdent, pour tout point rationnel x de X , il existe, à unique isomorphisme près, un unique torseur universel (\mathcal{T}, t) au-dessus de (X, x) . De plus si F est un corps de type fini sur son sous-corps

premier, alors les classes d'isomorphismes de toiseurs versels ayant un point rationnel sont en nombre fini [CTS4, proposition 2]. En notant $(\mathcal{T}_i)_{i \in I}$ une famille de représentants de ces classes d'isomorphisme, on obtient une partition canonique finie de l'ensemble des points rationnels de X :

$$X(F) = \coprod_{i \in I} \pi_i(\mathcal{T}_i(F)),$$

$\pi_i : \mathcal{T}_i \rightarrow X$ désignant le morphisme associé au toiseur \mathcal{T}_i .

L'intérêt des toiseurs versels est souligné par les deux résultats suivants :

THÉORÈME 3.5 (Colliot-Thélène, Sansuc [CTS5, théorème 2.1.2]). — *Soit V une bonne variété géométriquement rationnelle sur le corps de nombres K . Soit \mathcal{T} un toiseur versel au-dessus de X et \mathcal{T}^c une compactification projective et lisse de \mathcal{T} , alors*

$$\mathrm{Br}(\mathcal{T}^c) / \mathrm{Br}_0(\mathcal{T}^c) = \{0\}.$$

En particulier

$$\mathcal{T}^c(\mathbf{A}_K)^{\mathrm{Br}} = \mathcal{T}^c(\mathbf{A}_K).$$

THÉORÈME 3.6 (Colliot-Thélène, Sansuc [CTS5, corollaire 3.7.2]). — *Soit V une bonne variété géométriquement rationnelle sur le corps de nombres K . L'espace de Brauer-Manin $V(\mathbf{A}_K)^{\mathrm{Br}}$ est la réunion des images de $\prod_{v \in M_K} \mathcal{T}(K_v)$, où \mathcal{T} décrit un système de représentants des toiseurs versels au-dessus de V .*

REMARQUE 3.7. — Cet énoncé montre d'une part que l'implication

$$V(\mathbf{A}_K)^{\mathrm{Br}} \neq \emptyset \quad \Rightarrow \quad V(K) \neq \emptyset$$

est vraie si les toiseurs versels \mathcal{T} sur V vérifient l'implication

$$\prod_{v \in M_K} \mathcal{T}(K_v) \neq \emptyset \quad \Rightarrow \quad \mathcal{T}(K) \neq \emptyset$$

et d'autre part que les points rationnels de V sont denses dans $V(\mathbf{A}_K)^{\mathrm{Br}}$ si les compactifications projectives et lisses des toiseurs versels au-dessus de V vérifient l'approximation faible.

Donnons quelques exemples où cette méthode a été utilisée :

EXEMPLE 3.8. — Une surface de Châtelet S sur K est un modèle projectif et lisse d'une surface définie par une équation de la forme

$$Y^2 - aZ^2 = P(X),$$

où P est un polynôme séparable de degré 3 ou 4. L'exemple d'Iskovskih décrit au paragraphe 2.1 est une surface de ce type. Pour ces surfaces, Colliot-Thélène et Sansuc ont démontré dans [CTS4, §IV] que les toiseurs versels au-dessus de S sont stablement K -birationnellement équivalents au produit d'une conique et d'une variété X intersection complète géométriquement intègre et non conique de deux quadriques dans \mathbf{P}^7 . En outre,

X contient deux droites gauches conjuguées. Le principe de Hasse et l'approximation faible ayant été démontrés pour ces variétés via la méthode de fibration décrite plus loin, on obtient que les points rationnels de S sont denses dans $S(\mathbf{A}_K)^{\text{Br}}$ (Colliot-Thélène, Sansuc et Swinnerton-Dyer [CTSSD1] et [CTSSD2], cf. également [CT2]).

EXEMPLE 3.9. — Soit F un corps de groupe de Galois absolu \mathcal{G} . Un tore algébrique sur F est un groupe algébrique T sur F tel que \overline{T} soit isomorphe à un groupe de la forme $\mathbf{G}_{m,\overline{F}}^n$. L'équivalence de catégorie entre groupes de type multiplicatif et \mathbf{Z} -modules de type fini munis d'une action continue de \mathcal{G} envoie les tores algébriques sur les \mathcal{G} -réseaux, c'est-à-dire les \mathbf{Z} -modules libres de rang fini munis d'une action continue de \mathcal{G} . Une variété torique généralisée est une variété irréductible V munie d'une action d'un tore algébrique T avec une orbite ouverte U telle que \overline{U} soit isomorphe à \overline{T} .

Les toseurs universels au-dessus d'une variété torique projective et lisse ont d'abord été étudiés par Colliot-Thélène et Sansuc [CTS1, §4], ils ont ensuite été redécouverts par Delzant dans le cadre de la géométrie symplectique [Del]. Nous reprenons ici la construction de Cox [Co] qui en donne une description particulièrement élégante (cf. Salberger [Sal2] et Madore [Madore]).

Soit V une variété torique généralisée projective et lisse sur un corps F . Notons $\Sigma(1)$ l'ensemble des orbites de codimension 1 dans \overline{V} . On a alors une suite exacte naturelle de \mathcal{G} -réseaux

$$0 \rightarrow X^*(T) \rightarrow \mathbf{Z}^{\Sigma(1)} \rightarrow \text{Pic}(\overline{V}) \rightarrow 0$$

correspondant à une suite de tores algébriques

$$1 \rightarrow T_{\text{NS}} \rightarrow T_{\Sigma(1)} \rightarrow T \rightarrow 1.$$

On considère alors l'espace affine

$$\mathbf{A}_{\Sigma(1)} = \text{Spec}(F^s[X_\sigma, \sigma \in \Sigma(1)]^{\mathcal{G}}).$$

Pour toute partie I de $\Sigma(1)$, on note H_I le sous-espace affine de $\overline{\mathbf{A}}_{\Sigma(1)}$ défini par le système d'équations

$$X_\sigma = 0 \quad \text{pour} \quad \sigma \in I.$$

Pour tout σ de $\Sigma(1)$, notons $D_\sigma = \overline{\sigma}$ le diviseur correspondant de \overline{V} . On note alors \overline{X} le fermé réunion des sous-espaces H_I pour $I \subset \Sigma(1)$ tel que $\bigcap_{\sigma \in I} D_\sigma = \emptyset$. Ce fermé est défini sur F et l'ouvert $\mathcal{T} = \mathbf{A}_{\Sigma(1)} - X$ de l'espace affine $\mathbf{A}_{\Sigma(1)}$ est une variété torique pour le tore $T_{\Sigma(1)}$. Si x est un point rationnel dans l'orbite ouverte U de V , l'application naturelle $T \rightarrow V$ qui envoie l'élément neutre 1 de T sur x induit une application $T_{\Sigma(1)} \rightarrow V$ qui s'étend en un morphisme $\mathcal{T} \rightarrow V$. On vérifie que cela fait de $(\mathcal{T}, 1)$ un toseur universel au-dessus de (V, x) . Les toseurs versels au-dessus de V sont isomorphes à \mathcal{T} en tant que variétés. Ils vérifient donc tous l'approximation faible et les points rationnels de V sont denses dans $V(\mathbf{A}_K)^{\text{Br}}$. Notons que les équations de normes considérées par Hasse sont des exemples de variétés toriques.

Signalons enfin les travaux de Heath-Brown et Skorobogatov qui combinent la méthode du cercle avec les techniques de descente [HBS].

3.2. Techniques de fibration

Cette méthode apparaît déjà dans l'étude faite par Hasse du cas des quadriques : si $n \geq 5$, le passage du cas de $n - 1$ variables à celui de n variables peut se faire en utilisant des fibrations en quadriques.

Étant donné un morphisme $p : V \rightarrow B$, l'objectif est d'étudier si on peut déduire du fait que les obstructions de Brauer-Manin sont les seules pour les fibres de p que cela reste vrai pour V . Cette technique fut notamment explorée par Colliot-Thélène, Sansuc et Swinnerton-Dyer dans [CTSSD1] et [CTSSD2], puis par Harari et Skorobogatov pour des fibrations au-dessus de la droite projective (cf. [Sk1], [Ha1], [Ha3] et [Sk2]). Un archétype de ce que donne cette démarche est le résultat suivant :

THÉORÈME 3.10 (Harari [Ha3, proposition 3.1.1]). — *Soit V une bonne variété sur le corps de nombres K et $p : V \rightarrow \mathbf{P}_K^1$ un morphisme dominant de fibre générique géométriquement irréductible. Supposons que :*

- (i) *les fibres géométriques de p au-dessus de \mathbf{A}_K^1 sont irréductibles et de multiplicité 1,*
- (ii) *la fibre générique géométrique $\overline{V_{K(T)}}$ de p est rationnellement connexe,*
- (iii) *pour presque tout P appartenant à $\mathbf{P}^1(K)$ tel que la fibre V_P soit non singulière, on a*

$$\overline{V_P(K)} = V_P(\mathbf{A}_K)^{\text{Br}}.$$

Alors les points rationnels de V sont denses dans $V(\mathbf{A}_K)^{\text{Br}}$.

La difficulté est d'approcher les points de $V(\mathbf{A}_K)^{\text{Br}}$ par des points appartenant à l'espace de Brauer-Manin de fibres de l'application p au-dessus de points convenables de $\mathbf{P}^1(K)$. Identifions le corps des fonctions de \mathbf{P}_K^1 avec $K(T)$ et considérons le groupe

$$\text{Br}_{\text{nr}}(K(V)/K(T)) = \bigcap_{v \in \mathcal{P}(K(V)/K(T))} \ker(\partial_v)$$

où ∂_v est le morphisme résidu défini dans le paragraphe 2.2 et $\mathcal{P}(K(V)/K(T))$ désigne l'ensemble des valuations discrètes de rang un sur $K(V)$ qui sont triviales sur $K(T)$. Le groupe

$$\text{coker}(\text{Br}(K(T)) \rightarrow \text{Br}_{\text{nr}}(K(V)/K(T)))$$

est fini. Il existe donc un ouvert U de V de sorte que ce conoyau soit engendré par des éléments A_1, \dots, A_r dans l'image de l'application $\text{Br}(U) \rightarrow \text{Br}(K(V))$. Harari montre alors le lemme suivant :

LEMME FORMEL 3.11 (Harari [Ha1, corollaire 2.6.1], [CT6]). — Soient V une bonne variété sur le corps de nombres K et U un ouvert non vide de V . Soit B un sous-groupe fini de $\text{Br}(U)$. Soit $(P_v)_{v \in M_K} \in \prod_{v \in M_K} U(K_v)$ tel que pour tout A de $B \cap \text{Br}(V)$, on ait

$$\langle A, (P_v)_{v \in M_K} \rangle_{\mathbf{A}_K} = 0 ;$$

alors, pour toute partie finie S de M_K , il existe $(M_v)_{v \in M_K}$ appartenant à l'espace des adèles de U tel que $M_v = P_v$ pour $v \in S$ et

$$\forall A \in B, \quad \sum_{v \in M_K} \langle A, M_v \rangle_v = 0.$$

D'autre part, Harari montre qu'il existe un sous-ensemble hilbertien H de $\mathbf{P}^1(K)$ de sorte que pour tout P de H , le groupe $\text{Br}(V_P)/\text{Br}_0(V_P)$ pour la fibre soit engendré par les images des éléments A_1, \dots, A_r de $\text{Br}(K(V))$. La fin de la démonstration utilise un argument d'approximation forte pour les ensembles hilbertiens de la droite affine. Enfin l'énoncé donné ici utilise un résultat de Graber, Harris et Starr [GHS] qui montre que la condition de la proposition 3.1.1 de [Ha3] portant sur l'existence d'une section de la fibration sur \overline{K} est automatiquement vérifiée.

3.3. Une liste de résultats

Nous reprenons ici une liste de cas connus donnée par Colliot-Thélène dans un exposé récent [CT5] (cf. également [Sk4, §5.2]).

L'ensemble des points rationnels de V est dense dans $V(\mathbf{A}_K)^{\text{Br}}$ si V est une variété d'un des types suivants :

- un modèle projectif et lisse d'un espace homogène sous un groupe algébrique linéaire connexe si le stabilisateur d'un point géométrique est connexe (Voskresenskii, Sansuc [San], Borovoi [Bo2]), ou d'un espace homogène sous un groupe algébrique linéaire connexe et simplement connexe si le stabilisateur d'un point géométrique est abélien (Borovoi [Bo2]),
- un modèle projectif et lisse d'une intersection complète géométriquement irréductible et non conique de deux quadriques de \mathbf{P}_K^n si $n \geq 8$ (Colliot-Thélène, Sansuc, Swinnerton-Dyer, [CTSSD1] et [CTSSD2]) ; le groupe de Brauer étant trivial dans ce cas, ces variétés vérifient en fait le principe de Hasse et l'approximation faible,
- une surface fibrée en coniques sur la droite projective lorsque le nombre de fibres géométriques dégénérées est au plus 4 [CT1],
- une hypersurface cubique dans \mathbf{P}_K^n avec 3 points singuliers définis dans leur ensemble sur K , si $n \geq 3$ (Colliot-Thélène, Salberger [CTSal]) ; si, en outre, $n \neq 4$, alors l'approximation faible est vérifiée,
- une hypersurface cubique non singulière contenant une droite projective définie sur K dans \mathbf{P}_K^n si $n \geq 3$ (Salberger et Skorobogatov [Sal1] et [SaSk] pour $n = 3$, Harari [Ha3, §5.2.2] si $n \geq 4$). Pour $n \geq 4$, on a également l'approximation faible.

REMARQUE 3.12. — Parmi les cas pour lesquels la question de la densité de $V(K)$ dans $V(\mathbf{A}_K)^{\text{Br}}$ reste ouverte, on peut mentionner :

- les surfaces cubiques générales ; la question de savoir si l’obstruction de Brauer-Manin au principe de Hasse est la seule pour les surfaces cubiques diagonales sur \mathbf{Q} a été explorée de manière algorithmique dans [CTKS] ;
- l’intersection complète lisse de deux quadriques dans \mathbf{P}_K^n pour $7 \geq n \geq 4$, la difficulté étant de montrer que l’obstruction de Brauer-Manin au principe de Hasse est la seule dans ce cas.

REMARQUE 3.13. — Sous l’hypothèse de la finitude du groupe de Tate-Shafarevich des courbes elliptiques, Swinnerton-Dyer a obtenu les résultats qui suivent (cf. [SD2] et [CT6, §2 et 3]).

Si V est la surface projective cubique sur \mathbf{Q} définie par l’équation

$$\sum_{i=0}^3 a_i T_i^3 = 0$$

où les a_i sont des entiers non nuls sans facteur commun, non divisibles par un cube et vérifiant une des conditions suivantes :

- (i) il existe un nombre premier $p \neq 3$ divisant a_0 mais aucun des autres coefficients et un nombre premier $q \neq 3$ divisant a_1 mais aucun des autres coefficients,
- (ii) il existe un nombre premier $p \neq 3$ divisant a_0 mais aucun des autres coefficients et tel que les classes de a_1, a_2 et a_3 dans $\mathbf{F}_p^\times / \mathbf{F}_p^{\times 3}$ ne soient pas toutes égales,

alors V vérifie le principe de Hasse.

En outre, toujours sous l’hypothèse de la finitude du groupe de Tate-Shafarevich des courbes elliptiques sur un corps de nombres, il a montré que toute hypersurface cubique diagonale dans $\mathbf{P}_\mathbf{Q}^n$ pour $n \geq 4$ vérifie le principe de Hasse.

REMARQUE 3.14. — D’autres résultats ont été obtenus sous l’hypothèse de Schinzel. Cette hypothèse arithmétique forte s’énonce comme suit : soit $(f_i(x))_{1 \leq i \leq m} \in \mathbf{Z}[X]$ une famille de polynômes irréductibles dont les coefficients dominants sont positifs et telle que $\text{pgcd}_{n \in \mathbf{Z}}(\prod_{i=1}^m f_i(n)) = 1$; alors il existe une infinité de n tels que $f_i(n)$ soit premier pour $i = 1, \dots, m$. Le seul cas connu est le théorème de la progression arithmétique avec un polynôme de degré un, qui est utilisé par Hasse dans le passage de 3 à 4 variables pour les formes quadratiques. Ces résultats conditionnels concernent notamment des fibrations au-dessus de \mathbf{P}_K^1 .

3.4. Le cas des espaces principaux homogènes sous une variété abélienne

Jusqu’à maintenant nous sommes essentiellement resté dans le cadre des variétés géométriquement rationnellement connexes. Toutefois, Manin dans son exposé à Nice avait

déjà montré que, pour une variété abélienne de groupe de Tate-Shafarevich fini, l'obstruction au principe de Hasse qu'il construisait était la seule. Le cas de l'approximation faible fut traité plus tard par Wang [Wa], ce qui donne l'énoncé suivant :

THÉORÈME 3.15 (Manin, Wang). — *Soit A une variété abélienne sur le corps de nombres K et soit V un espace principal homogène sous A . Supposons que le groupe de Tate-Shafarevich de A*

$$\text{III}^1(K, A) = \ker \left(H^1(K, A) \rightarrow \prod_{v \in M_K} H^1(K_v, A) \right)$$

soit fini; alors

(i) $V(\mathbf{A}_K)^{\text{Br}} \neq \emptyset \Rightarrow V(K) \neq \emptyset$.

(ii) *Notons $M_{K, \infty}$ l'ensemble des places archimédiennes de K . Si l'adhérence de $A(K)$ dans $\prod_{v \in M_{K, \infty}} A(K_v)$ est une partie ouverte de ce produit, alors*

$$\overline{V(K)} = V(\mathbf{A}_K)^{\text{Br}}$$

REMARQUES 3.16. — (i) L'accouplement $\langle \alpha, \cdot \rangle_{\mathbf{A}_K}$ étant localement constant, la conclusion de la deuxième assertion ne peut être valide sans l'hypothèse faite.

(ii) La finitude du groupe de Tate-Shafarevich est une conjecture forte mais classique pour les contemplateurs des variétés abéliennes.

4. L'OBSTRUCTION DE BRAUER-MANIN N'EST PAS LA SEULE

4.1. Au-delà du critère de Manin

La notion de torseur versel donne une description alternative de l'obstruction de Brauer-Manin. Si \mathcal{T} est un torseur sur V sous un groupe de type multiplicatif T , alors pour toute extension L de K et pour tout point x de $V(L)$, l'image inverse de \mathcal{T} par x est un espace principal homogène sur $\text{Spec}(L)$ sous T . Notons $\mathcal{T}(x)$ sa classe dans $H^1(L, T)$. On a alors que $\overline{V(K)}$ est contenu dans

$$V(\mathbf{A}_K)^{\mathcal{T}} = \left\{ (x_v)_{v \in M_K} \in V(\mathbf{A}_K) \mid (\mathcal{T}(x_v))_{v \in M_K} \in \text{Im} \left(H^1(K, T) \rightarrow \prod_{v \in M_K} H^1(K_v, T) \right) \right\}.$$

Harari et Skorobogatov ont montré que cette inclusion subsiste si on remplace T par un groupe algébrique linéaire G arbitraire [HS]. Cet argument permet d'expliquer le premier exemple explicite de variété telle que

$$V(\mathbf{A}_K)^{\text{Br}} \neq \emptyset \quad \text{et} \quad V(K) = \emptyset$$

produit par Skorobogatov en 1999 dans [Sk3]. Cet exemple est donné par les équations affines

$$(X^2 + 1)Y^2 = (X^2 + 2)Z^2 = 3(T^4 - 54T^2 - 117T - 243);$$

un modèle projectif et lisse de cette surface est donné par une surface bielliptique quotient d'un produit $C \times E$ où C et E sont deux courbes de genre 1. Dans le même esprit, Harari dans [Ha4] a créé une méthode fournissant des exemples où $V(K)$ n'est pas dense dans $V(\mathbf{A}_K)^{\text{Br}}$ à partir de variétés dont le groupe fondamental géométrique n'est pas abélien.

REMARQUE 4.1. — Harari a montré dans [Ha5] que si G est un groupe linéaire abélien ou connexe, cette construction ne donne pas plus d'informations que l'obstruction de Brauer-Manin : l'espace obtenu contient $V(\mathbf{A}_K)^{\text{Br}}$.

4.2. Lien avec des conjectures de Lang

Soit $V \subset \mathbf{P}_K^n$ une hypersurface lisse de degré d et de dimension supérieure ou égale à trois. Le théorème de Lefschetz permet de montrer que, sous ces hypothèses, le quotient $\text{Br}(V)/\text{Br}_0(V)$ est trivial. Si $V(K)$ était dense dans $V(\mathbf{A}_K)^{\text{Br}}$, il le serait dans $V(\mathbf{A}_K)$. Pour toute telle hypersurface possédant un point rationnel, $V(K)$ serait dense pour la topologie de Zariski. Mais si $d > n$, V est de type général et cela contredirait une conjecture de Lang qui prédit que les points rationnels d'une variété de type général ne sont pas denses pour la topologie de Zariski [La, §3].

L'implication

$$V(\mathbf{A}_K)^{\text{Br}} \neq \emptyset \Rightarrow V(K) \neq \emptyset$$

est également en contradiction avec des conjectures de Lang, bien que cela soit plus délicat à montrer. En 1995, Sarnak et Wang [SW] considèrent l'hypersurface de $\mathbf{P}_\mathbf{Q}^5$ définie par l'annulation du polynôme

$$F(X_0, X_1, X_2, X_3, X_4, X_5) = X_5^{1130} + H(X_0, X_1, X_2, X_3, X_4)$$

avec

$$\begin{aligned} H(X_0, X_1, X_2, X_3, X_4) = & \sum_{i=0}^4 X_i^{1130} - 15(X_0 X_1^4)^{226} + (X_1 X_2^4)^{226} \\ & + (X_2 X_3^4)^{226} + (X_3 X_4^4)^{226} + (X_4 X_0^4)^{226} + (X_0^2 X_2^3)^{226} \\ & + (X_1^2 X_3^3)^{226} + (X_2^2 X_4^3)^{226} + (X_3^2 X_0^3)^{226} + (X_4^2 X_1^3)^{226}. \end{aligned}$$

Cette variété est hyperbolique au sens de Brody ou Kobayashi. Des conjectures de Lang prévoient que X ne possède qu'un nombre fini de points. Par conséquent, l'hypersurface X_k de \mathbf{P}_K^4 d'équation

$$F(kX_2, X_1, X_2, X_3, X_4, X_5) = 0$$

ne peut admettre de point rationnel pour k assez grand. Mais $X_k(\mathbf{A}_K)^{\text{Br}} \neq \emptyset$ pour une infinité de valeurs de k . Ces exemples ne peuvent pas être non plus expliqués par la construction d'Harari et Skorobogatov.

Plus récemment, Poonen dans [Po] montre que les conjectures de Lang entraînent l'existence d'intersections complètes lisses de dimension 3 dans \mathbf{P}_K^N pour lesquelles le principe

de Hasse n'est pas vérifié. Une telle variété est simplement connexe et de groupe de Brauer trivial et, à nouveau, l'absence de points rationnels ne peut donc être expliquée ni par la méthode de Manin ni par ses extensions non abéliennes.

5. L'OBSTRUCTION DE BRAUER-MANIN EST-ELLE LA SEULE ?

La formule de réciprocité du corps de classes permet également de construire des obstructions à l'existence d'un 0-cycle de degré 1. Plus généralement, ce paragraphe est consacré aux questions d'existence de cycles algébriques.

En combinant la dualité de Poitou-Tate avec la dualité de Poincaré, Saito [Sa] a construit pour toute bonne variété de dimension d sur K une suite exacte

$$\cdots \rightarrow H_{\text{ét}}^{2i}(V, \mu_n^{\otimes i}) \rightarrow \prod_{v \in M_K} \tilde{H}^{2i}(V_{K_v}, \mu_n^{\otimes i}) \rightarrow \text{Hom}(H_{\text{ét}}^{2j}(V, \mu_n^{\otimes j}), \mathbf{Q}/\mathbf{Z}) \rightarrow \cdots$$

où $j = d + 1 - i$, où $\tilde{H}^{2i}(V_{K_v}, \mu_n^{\otimes i})$ désigne $H_{\text{ét}}^{2i}(V_{K_v}, \mu_n^{\otimes i})$ pour une place v non archimédienne, vaut $\{0\}$ si v est complexe et est un groupe fini de 2-torsion muni d'un morphisme naturel

$$H_{\text{ét}}^{2i}(V_{K_v}, \mu_n^{\otimes i}) \rightarrow \tilde{H}_{\text{ét}}^{2i}(V_{K_v}, \mu_n^{\otimes i})$$

lorsque v est réelle et où $\prod_{v \in M_K} \tilde{H}^{2i}(V_{K_v}, \mu_n^{\otimes i})$ est le produit restreint des groupes précédents relativement aux images de $H_{\text{ét}}^{2i}(\mathcal{Y}_{\mathcal{O}_v}, \mu_n^{\otimes i})$ pour un modèle projectif et lisse \mathcal{Y} de V sur un ouvert $\text{Spec}(\mathcal{O}_S)$ de $\text{Spec}(\mathcal{O}_K)$ sur lequel n est inversible. Notons $\text{CH}^i(V)$ le groupe de Chow des cycles de codimension i de V modulo l'équivalence rationnelle. En utilisant l'application cycle

$$\text{cl}_n : \text{CH}^i(V_L) \rightarrow H_{\text{ét}}^{2i}(V_L, \mu_n^{\otimes i})$$

pour toute extension L de K , on obtient un accouplement (cf. [CT3])

$$(\cdot, \cdot) : H_{\text{ét}}^{2j}(V, \mathbf{Q}/\mathbf{Z}(j)) \times \prod_{v \in M_K} \text{CH}^i(V_{K_v}) \rightarrow \mathbf{Q}/\mathbf{Z}$$

qui est trivial sur l'image de $\text{CH}^i(V)$ dans $\prod_{v \in M_K} \text{CH}^i(V_{K_v})$. Colliot-Thélène [CT3] énonce alors la conjecture suivante :

CONJECTURE 5.1. — Soit $z = (z_v)_{v \in M_K}$ un élément de $\prod_{v \in M_K} \text{CH}^i(V_{K_v})$. Supposons que

$$\forall \xi \in H_{\text{ét}}^{2j}(V, \mathbf{Q}/\mathbf{Z}(j)), \quad (\xi, z) = 0;$$

alors, pour tout $n > 0$, il existe $y \in \text{CH}^i(V)$ tel que pour toute place non archimédienne v de K , on ait $\text{cl}_n(y) = \text{cl}_n(z_v)$ dans $H_{\text{ét}}^{2i}(V_{K_v}, \mu_n^{\otimes i})$.

REMARQUES 5.2. — (i) Si le groupe de Tate-Shafarevich de la variété de Picard de V est fini, la conjecture est vraie pour $i = 1$.

(ii) Pour $i = \dim(V)$, on obtient un accouplement

$$\mathrm{Br}(V) \times \prod_{v \in M_K} \mathrm{CH}_0(V_{K_v}) \rightarrow \mathbf{Q}/\mathbf{Z}$$

où CH_0 désigne le groupe des 0-cycles modulo l'équivalence rationnelle. Dans ce cas, il est induit par les accouplements naturels

$$\begin{aligned} \mathrm{Br}(V) \times Z_0(V_{K_v}) &\rightarrow \mathbf{Q}/\mathbf{Z} \\ \left(A, \sum_P n_P P \right) &\mapsto \mathrm{inv}_v \left(\sum_P n_P \mathrm{cores}_{K_v(P)/K_v}(A(P)) \right) \end{aligned}$$

définis par Manin [Ma1]. Une condition nécessaire pour l'existence d'un 0-cycle de degré 1 sur K est donc l'existence d'une famille de 0-cycles $z = (z_v)_{v \in M_K}$ tous de degré 1 tels que

$$\forall A \in \mathrm{Br}(V), \quad (A, z) = 0.$$

La conjecture impliquerait que cette condition est également suffisante; autrement dit, l'obstruction de Brauer-Manin à l'existence d'un 0-cycle de degré un serait la seule.

Cette conjecture a été vérifiée pour $i = \dim(V)$ dans le cas des bonnes variétés V pour lesquelles il existe un morphisme propre et surjectif $\pi : V \rightarrow C$, de fibre générique une variété de Severi-Brauer dont l'indice est sans facteur carré, au-dessus d'une courbe C projective et lisse, dont on suppose que sa jacobienne a un groupe de Tate-Shafarevich fini. Cela a été démontré par Frossard [Fr] en utilisant des résultats de Salberger [Sal1] et Colliot-Thélène [CT4] (cf. également l'article de van Hamel [vH]).

6. AUTRES CADRES

Mentionnons pour terminer que la problématique du principe de Hasse et de l'approximation faible a également été considérée dans un cadre fonctionnel :

- Dans le cas où le corps de base est $K = k(C)$ où C est une courbe projective, lisse et connexe sur un corps algébriquement clos. En particulier, Colliot-Thélène et Gille ont montré dans [CTG] que l'approximation faible vaut pour les K -variétés géométriquement rationnellement connexes qui se ramènent par des fibrations à des espaces homogènes sous des groupes linéaires connexes.
- Le cas fonctionnel réel où le corps de base est le corps des fonctions d'une courbe projective, lisse et géométriquement intègre sur le corps des réels \mathbf{R} . En particulier, Ducros [Du1] et Scheiderer [Sc] ont montré que le principe de Hasse vaut pour les espaces principaux homogènes sous un groupe semi-simple simplement connexe. D'autre part, Ducros a prouvé que l'analogie de l'obstruction de Brauer-Manin dans ce cadre est la seule pour les fibrés en coniques ou, plus généralement, en variétés de Severi-Brauer au-dessus de la droite projective ([Du2], [Du3]).

Remerciements. Je remercie chaleureusement ceux qui ont accepté de relire ce texte dans un délai incroyablement bref et, en particulier, J.-L. Colliot-Thélène, A. Ducros, D. Harari et G. Rémond.

RÉFÉRENCES

- [Az] J.-P. Azra, *Relations diophantiennes et la solution négative du 10ème problème de Hilbert (d'après M. Davis, H. Putnam, J. Robinson et I. Matiassevitch)*, Sémin. Bourbaki (1970/71), Exp. n° 383, Lect. Notes in Math. **244**, Springer-Verlag (1971), 11-28.
- [Bir] B.J. Birch, *Forms in many variables*, Proc. Roy. Soc. London **265A** (1962), 245–263.
- [Bo1] M.V. Borovoi, *Abelianization of the second nonabelian Galois cohomology*, Duke Math. J. **72** (1993), n° 1, 217–239.
- [Bo2] ———, *The Brauer-Manin obstructions for homogeneous spaces with connected or abelian stabilizer*, J. Reine Angew. Math. **473** (1996), 181–194.
- [Bki] N. Bourbaki, *Variétés différentielles et analytiques, fascicule de résultats*, Diffusion C.C.L.S., Paris, 1988.
- [Ca] J.W.S. Cassels, *Lectures on elliptic curves*, London mathematical society student texts, vol. 24, Cambridge university press, Cambridge, 1991.
- [Ch] F. Châtelet, *Points rationnels sur certaines courbes et surfaces cubiques*, Enseignement Math. (2) **5** (1959), 153–170.
- [Che] V.I. Chernousov, *The Hasse principle for groups of type E_8* , Dokl. Akad. Nauk SSSR **306** (1989), n° 5, 1059–1063; English transl. in Soviet Math. Dokl. **39** (1989), n° 3, 592–596.
- [CT1] J.-L. Colliot-Thélène, *Surfaces rationnelles fibrées en coniques de degré 4*, Séminaire de théorie des nombres (Paris 1988–1989) (C. Goldstein ed.), Progress in Math., vol. 91, Birkhäuser, Basel, 1990, pp. 43–53.
- [CT2] ———, *L'arithmétique des variétés rationnelles*, Ann. Fac. Sci. Toulouse (6) **1** (1992), n° 3, 295–336.
- [CT3] ———, *Conjectures de type local-global sur l'image des groupes de Chow dans la cohomologie étale*, Algebraic K-theory (Seattle, 1997) (W. Raskind et C. Weibel, eds.), Proc. Sympos. Pure Math., vol. 67, Amer. Math. Soc., Providence, 1999, pp. 1–12.

- [CT4] ———, *Principe local-global pour les zéro-cycles sur les surfaces réglées*, J. Amer. Math. Soc. **13** (2000), n° 1, 101–124.
- [CT5] ———, *The local-global principle for rational points and zero-cycles*, Raymond and Beverley Sackler distinguished lectures in mathematics, Tel Aviv University, 2003.
- [CT6] ———, *Points rationnels sur les fibrations*, Higher dimensional varieties and rational points (Budapest, 2001) (K. Böröczky, J. Kollár et T. Szamuely, eds.), Bolyai Soc. Math. Stud., vol. 12, Springer-Verlag, Berlin, 2003, pp. 171–221.
- [CTCS] J.-L. Colliot-Thélène, D. Coray et J.-J. Sansuc, *Descente et principe de Hasse pour certaines variétés rationnelles*, J. Reine Angew. Math. **320** (1980), 150–191.
- [CTG] J.-L. Colliot-Thélène et P. Gille, *Remarques sur l'approximation faible sur un corps de fonctions d'une variable*, Arithmetic of higher-dimensional algebraic varieties (Palo Alto, 2002) (B. Poonen et Y. Tschinkel eds.), Progress in Math., vol. 226, Birkhäuser, Basel, 2004, pp. 121–134.
- [CTKS] J.-L. Colliot-Thélène, D. Kanevsky et J.-J. Sansuc, *Arithmétique des surfaces cubiques diagonales*, Diophantine approximation and transcendence theory (Bonn, 1985), Lect. Notes in Math., vol. 1290, Springer-Verlag, Berlin, 1987, pp. 1–108.
- [CTSa] J.-L. Colliot-Thélène et P. Salberger, *Arithmetic on some singular cubic hypersurfaces*, Proc. London Math. Soc. (3) **58** (1989), n° 3, 519–549.
- [CTS1] J.-L. Colliot-Thélène et J.-J. Sansuc, *Torseurs sous des groupes de type multiplicatif; applications à l'étude des points rationnels de certaines variétés algébriques*, C. R. Acad. Sci. Paris Sér. A **282** (1976), 1113–1116.
- [CTS2] ———, *Variétés de première descente attachées aux variétés rationnelles*, C. R. Acad. Sci. Paris Sér. A **284** (1977), 967–970.
- [CTS3] ———, *La descente sur une variété rationnelle définie sur un corps de nombres*, C. R. Acad. Sci. Paris Sér. A **284** (1977), 1215–1218.
- [CTS4] ———, *La descente sur les variétés rationnelles*, Journées de géométrie algébrique d'Angers (1979) (A. Beauville, ed.), Sijthoff & Noordhoff, Alphen aan den Rijn, 1980, pp. 223–237.
- [CTS5] ———, *La descente sur les variétés rationnelles, II*, Duke Math. J. **54** (1987), n° 2, 375–492.
- [CTSSD1] J.-L. Colliot-Thélène, J.-J. Sansuc et H.P.F. Swinnerton-Dyer, *Intersections of two quadrics and Châtelet surfaces I*, J. für Math. **373** (1987), 37–107.

- [CTSSD2] ———, *Intersections of two quadrics and Châtelet surfaces II*, J. für Math. **374** (1987), 72–168.
- [Co] D. Cox, *The homogeneous coordinate ring of a toric variety*, J. Algebraic Geom. **4** (1995), n° 1, 17–50.
- [Del] T. Delzant, *Hamiltoniens périodiques et images convexes de l'application moment*, Bull. Soc. Math. France **116** (1988), n° 3, 315–339.
- [De] J.-M. Deshouillers, *L'étude des formes cubiques rationnelles via la méthode du cercle (d'après D.R. Heath-Brown, C. Hooley et R.C. Vaughan)*, Sémin. Bourbaki (1989/90), Exp. n° 720, Astérisque **189-190** (1990), 155-177.
- [Du1] A. Ducros, *Principe de Hasse pour les espaces principaux homogènes sous les groupes classiques sur un corps de dimension cohomologique virtuelle au plus 1*, Manuscripta Math. **89** (1996), n° 3, 335–354.
- [Du2] ———, *L'obstruction de réciprocité à l'existence de points rationnels pour certaines variétés sur le corps des fonctions d'une courbe réelle*, J. Reine Angew. Math. **504** (1998), 73–114.
- [Du3] ———, *Fibrations en variétés de Severi-Brauer au-dessus de la droite projective sur le corps des fonctions d'une courbe réelle*, C. R. Acad. Sci. Paris Sér. I Math. **327** (1998), n° 1, 71–75.
- [Ei] M. Eichler, *Über die Idealklassenzahl hyperkomplexer Systeme*, Math. Z. **43** (1938), 481–494.
- [Fr] E. Frossard, *Obstruction de Brauer-Manin pour les zéro-cycles sur des fibrations en variétés de Severi-Brauer*, J. Reine Angew. Math. **557** (2003), 81–101.
- [GHS] T. Graber, J. Harris et J. Starr, *Families of rationally connected varieties*, J. Amer. Math. Soc. **16** (2003), n° 1, 57–67.
- [Gr] A. Grothendieck, *Le groupe de Brauer III : Exemples et compléments*, Dix exposés sur la cohomologie des schémas, Adv. Stud. Pure Math., vol. 3, North-Holland, Amsterdam et Masson, Paris, 1968, pp. 88–188.
- [vH] J. van Hamel, *The Brauer-Manin obstruction for zero-cycles on Severi-Brauer fibrations over curves*, J. London Math. Soc. (2) **68** (2003), n° 2, 317–337.
- [Ha1] D. Harari, *Méthode des fibrations et obstruction de Manin*, Duke Math. J. **75** (1994), n° 1, 221–260.
- [Ha2] ———, *Obstructions de Manin transcendantes*, Number theory (Paris 1993-1994), London Math. Soc. Lecture Note Ser., vol. 235, Cambridge Univ. Press, Cambridge, 1996, pp. 75–87.

- [Ha3] ———, *Flèches de spécialisations en cohomologie étale et applications arithmétiques*, Bull. Soc. Math. France **125** (1997), 143–166.
- [Ha4] ———, *Weak approximation and non-abelian fundamental groups*, Ann. Sci. École Norm. Sup. (4) **33** (2000), n° 4, 467–484.
- [Ha5] ———, *Groupes algébriques et points rationnels*, Math. Ann. **322** (2002), n° 4, 811–826.
- [HS] D. Harari et A.N. Skorobogatov, *Non-abelian cohomology and rational points*, Compositio Math. **130** (2002), n° 3, 241–273.
- [Harder1] G. Harder, *Über die Galoiskohomologie halbeinfacher Matrizen­gruppen, I*, Math. Z. **90** (1965), 404–428.
- [Harder2] ———, *Über die Galoiskohomologie halbeinfacher Matrizen­gruppen, II*, Math. Z. **92** (1966), 396–415.
- [Harder3] ———, *Bericht über neuere Resultate der Galoiskohomologie halbeinfacher Gruppen*, Jber. Deutsche Math.-Verein. **70** 1967/68 Heft 4, Abt. 1, 182–216.
- [Hasse1] H. Hasse, *Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen*, J. Reine Angew. Math. **152** (1923), 129–148.
- [Hasse2] ———, *Beweis eines Satzes und Widerlegung einer Vermutung über das allgemeine Normenrestsymbol*, Nachrichten der Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Kl. H **1** (1931), 64–69.
- [HB] D.R. Heath-Brown, *Cubic forms in ten variables*, Proc. London Math. Soc. (3) **47** (1983), n° 2, 225–257.
- [HBS] D.R. Heath-Brown et A. Skorobogatov, *Rational solutions of certain equations involving norms*, Acta Math. **189** (2002), n° 2, 161–177.
- [Ho1] C. Hooley, *On nonary cubic forms*, J. Reine Angew. Math. **386** (1988), 32–98.
- [Ho2] ———, *On nonary cubic forms. III*, J. Reine Angew. Math. **456** (1994), 53–63.
- [Is] V.A. Iskovskih, *A counterexample to the Hasse principle for systems of two quadratic forms in five variables*, Mat. Zametki **10** (1971), 253–257; English transl. in Math. Notes **10** (1971), 575–577.
- [Kn] M. Kneser, *Hasse principle for H^1 of simply connected groups*, Algebraic groups and discontinuous subgroups (Boulder, 1965), Proc. Sympos. Pure Math., vol. 9, 1966, pp. 159–163.

- [Lan] W. Landherr, *Über einfache Liesche Ringe*, Abh. Math. Semin. Hamb. Univ. **11** (1935), 41–64.
- [La] S. Lang, *Number Theory III, diophantine geometry*, Encyclopaedia of Math. Sciences, vol. 60, Springer-Verlag, Berlin, 1991.
- [LW] S. Lang et A. Weil, *Number of points of varieties in finite fields*, Amer. J. Math. **76** (1954), 819–827.
- [Li] C.-E. Lind, *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven von Geschlecht Eins*, Diss. Uppsala, 1940.
- [Madore] D. Madore, *Very free R-equivalence on toric models* (2003).
- [Ma1] Y.I. Manin, *Le groupe de Brauer-Grothendieck en géométrie diophantienne*, Actes du congrès international des mathématiciens, Tome 1 (Nice, 1970), Gauthier-Villars, Paris, 1971, pp. 401–411.
- [Ma2] ———, *A course in mathematical logic*, Graduate Texts in Math., vol. 53, Springer-Verlag, New York, 1977.
- [Ma3] ———, *Cubic forms (second edition)*, North-Holland Math. Library, vol. 4, North-Holland, Amsterdam, 1986.
- [Mi] H. Minkowski, *Über die Bedingungen, unter welchen zwei quadratische Formen mit rationalen Koeffizienten ineinander rational transformiert werden können*, J. Reine Angew. Math. **106** (1890), 5–26.
- [NSW] J. Neukirch, A. Schmidt et K. Wingberg, *Cohomology of number fields*, Grundlehren der mathematischen Wissenschaften, vol. 323, Springer-Verlag, Berlin, 2000.
- [Ni] H. Nishimura, *Some remarks on rational points*, Mem. Coll. Sci. Univ. Kyoto Ser. A Math. **29** (1955), 189–192.
- [PR] V.P. Platonov et A. Rapinchuk, *Algebraic groups and number theory*, Pure and applied mathematics, vol. 139, Academic press, London, 1991.
- [Po] B. Poonen, *The Hasse principle for complete intersections in projective space*, Rational points on algebraic varieties, Progress in Math., vol. 199, Birkhäuser, Basel, 2001, pp. 307–311.
- [Re] H. Reichardt, *Einige im Kleinen überall lösbare, im Großen unlösbare diophantische Gleichungen*, J. Reine Angew. Math. **184** (1942), 12–18.

- [Sa] S. Saito, *A global duality theorem for varieties over global fields*, Algebraic K -theory : connections with geometry and topology (Lake Louise, 1987) (J.F. Jardine et V.P. Snaith, eds.), Kluwer Academic Publishers, Lake Louise, 1987, 1989, pp. 425–444.
- [Sal1] P. Salberger, *Zero-cycles on rational surfaces over number fields*, Invent. Math. **91** (1988), n° 3, 505–524.
- [Sal2] ———, *Tamagawa measures on universal torsors and points of bounded height on Fano varieties*, Nombre et répartition de points de hauteur bornée, Astérisque, vol. 251, SMF, Paris, 1998, pp. 91–258.
- [SaSk] P. Salberger et A.N. Skorobogatov, *Weak approximation for surfaces defined by two quadratic forms*, Duke Math. J. **63** (1991), n° 2, 517–536.
- [San] J.-J. Sansuc, *Groupe de Brauer et arithmétique des groupes algébriques linéaires sur un corps de nombres*, J. Reine Angew. Math. **327** (1981), 12–80.
- [SW] P. Sarnak et L. Wang, *Some hypersurfaces in \mathbf{P}^4 and the Hasse principle*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), 319–322.
- [Sc] C. Scheiderer, *Hasse principles and approximation theorems for homogeneous spaces over fields of virtual cohomological dimension one*, Invent. Math. **125** (1996), n° 2, 307–365.
- [Sch] W.M. Schmidt, *The density of integer points on homogeneous varieties*, Acta Math. **154** (1985), n° 3–4, 243–296.
- [Sei] A. Seidenberg, *A new decision method for elementary algebra*, Ann. of Math. (2) **60** (1954), 365–374.
- [Se1] J.-P. Serre, *Corps locaux*, Actualités scientifiques et industrielles, vol. 1296, Hermann, Paris, 1968.
- [Se2] ———, *Cours d'arithmétique*, Le mathématicien, PUF, Paris, 1988.
- [Skinner] C.M. Skinner, *Forms over number fields and weak approximation*, Compositio Math. **106** (1997), n° 1, 11–29.
- [Sk1] A.N. Skorobogatov, *On the fibration method for proving the Hasse principle and weak approximation*, Séminaire de théorie des nombres (Paris, 1988–1989) (C. Goldstein, ed.), Progress in Math., vol. 91, Birkhäuser, Boston, 1990, pp. 205–219.
- [Sk2] ———, *Descent on fibrations over the projective line*, Amer. J. Math. **118** (1996), n° 5, 905–923.

- [Sk3] ———, *Beyond the Manin obstruction*, Invent. Math. **135** (1999), n° 2, 399–424.
- [Sk4] ———, *Torsors and rational points*, Cambridge tracts in math., vol. 144, Cambridge University Press, 2001.
- [SD1] H.P.F. Swinnerton-Dyer, *Two special cubic surfaces*, Mathematika **9** (1962), 54–56.
- [SD2] ———, *The solubility of diagonal cubic surfaces*, Ann. Sci. École Norm. Sup. (4) **34** (2001), 891–912.
- [Ta] A. Tarski, *A decision method for elementary algebra and geometry*, RAND Corporation, Santa Monica, Calif., 1948.
- [Wa] L. Wang, *Brauer-Manin obstruction to weak approximation on abelian varieties*, Israel J. Math. **94** (1996), 189–200.
- [Wi] O. Wittenberg, *Transcendental Brauer-Manin obstruction on a pencil of elliptic curves*, Arithmetic of higher-dimensional algebraic varieties (Palo Alto, 2002) (B. Poonen et Y. Tschinkel eds.), Progress in Math., vol. 226, Birkhäuser, Basel, 2004, pp. 259–267.

Emmanuel PEYRE

Institut Fourier

UFR de Mathématiques, UMR 5582

Université de Grenoble I et CNRS

BP 74

F-38402 Saint-Martin d'Hères CEDEX

E-mail : Emmanuel.Peyre@ujf-grenoble.fr

COMPLÈTE RÉDUCTIBILITÉ

par Jean–Pierre SERRE

1. INTRODUCTION : LE CAS DU GROUPE LINÉAIRE

1.1. Rappels

Soient k un corps commutatif et Γ un groupe. Un Γ -module (ou *une représentation linéaire* de Γ) est un k -espace vectoriel V de dimension finie, muni d'une action linéaire de Γ .

On dit que :

(1.1.1) V est *irréductible* (ou *simple*) si $V \neq 0$ et si V ne contient aucun sous- Γ -module distinct de 0 et de V .

(1.1.2) V est *complètement réductible* (ou *semi-simple*) si V est somme directe de Γ -modules irréductibles.

(1.1.3) V est *indécomposable* si $V \neq 0$, et si V n'est pas somme directe de deux sous- Γ -modules $\neq 0$.

[Dans la suite, nous abrègerons en écrivant *ir*, *cr*, *ind* respectivement.]

Les propriétés suivantes sont bien connues :

(1.1.4) V est *cr* si et seulement si, pour tout sous-module W de V , il existe un sous-module W' de V tel que $V = W \oplus W'$.

(1.1.5) Si V est somme directe de sous-modules V_i , alors V est *cr* \Leftrightarrow tous les V_i sont *cr*.

On connaît moins bien les propriétés relatives au produit tensoriel $V \otimes V'$ de deux représentations V et V' . La plupart des cours d'Algèbre se bornent à définir le produit en question, et à en démontrer des propriétés évidentes. Aucun, à ma connaissance, ne signale le résultat très frappant suivant, dû à Chevalley :

THÉORÈME 1.1 ([Ch 55, p. 88]). — *Supposons k de caractéristique 0. Si V et V' sont des Γ -modules complètement réductibles, leur produit tensoriel $V \otimes V'$ est complètement réductible.*

(Noter que l'on ne fait aucune hypothèse sur le groupe Γ .)

Lorsqu'on veut étendre ce théorème à la caractéristique p (avec des conditions restrictives sur $\dim V$ et $\dim V'$, cf. prop. 5.8), il est utile de disposer d'une notion de *complète réductibilité* dans laquelle le groupe linéaire $\mathbf{GL}(V)$ est remplacé par un groupe réductif quelconque. Cette notion peut se définir, soit en termes de sous-groupes paraboliques, soit en termes d'*immeubles de Tits*, cf. [Se 97b], [Se 98]. C'est ce que nous allons voir. Les §§ 2,3 contiennent les énoncés généraux, et les §§ 4,5 donnent des critères plus précis, ainsi que diverses applications.

1.2. Exemple: l'immeuble de $\mathbf{GL}(V)$

Soit V un k -espace vectoriel de dimension finie n . On supposera $n \geq 2$ (sinon, l'immeuble correspondant est vide).

1.2.1. *Définition.* L'immeuble de $\mathbf{GL}(V)$, appelé aussi immeuble de V , est un complexe simplicial $X = X(V)$, de dimension $n - 2$, qui est défini de la manière suivante :

- les sommets de X correspondent bijectivement aux sous-espaces vectoriels de V distincts de 0 et de V . (Si W est un tel sous-espace, on note x_W le sommet correspondant.)
- un ensemble s de sommets de X est un simplexe de X si et seulement si les sous-espaces vectoriels correspondants forment un drapeau, i.e. une filtration strictement croissante de V . [Si l'on préfère la géométrie projective à la géométrie affine, on peut aussi voir les sommets de X comme les sous-variétés projectives de l'espace projectif $\mathbf{P}(V)$, distinctes de \emptyset et de $\mathbf{P}(V)$.]

1.2.2. *Type.* Si $x = x_W$ est un sommet de X , on note $\text{type}(x)$ la dimension de l'espace vectoriel W . L'ensemble des types de sommets est l'ensemble $I = \{1, \dots, n - 1\}$.

1.2.3. *Opposition.* Deux sommets $x = x_W$ et $x' = x_{W'}$ sont dits *opposés* si V est somme directe de W et W' ; leurs types se correspondent par l'involution $t \mapsto n - t$ de I . Deux simplexes sont *opposés* si tout sommet de l'un est opposé à un sommet de l'autre. En termes de filtrations, cela correspond à la notion usuelle de *filtrations opposées*.

1.2.4. *Sous-groupes paraboliques.* Les sous-groupes paraboliques de $\mathbf{GL}(V)$ sont les stabilisateurs des drapeaux de V . Les sous-groupes paraboliques *propres* (i.e. distincts de $\mathbf{GL}(V)$) correspondent donc aux simplexes non vides de l'immeuble X ; les paraboliques maximaux correspondent aux sommets de X et le simplexe \emptyset au groupe $\mathbf{GL}(V)$. Deux simplexes s et s' , correspondant aux paraboliques P et P' , sont opposés au sens du n° 1.2.3 si et seulement si P et P' sont opposés au sens usuel du terme, c'est-à-dire si $P \cap P'$ est un sous-groupe de Levi de chacun d'eux, cf. [BT 65, 4.8].

1.2.5. *Appartements.* Soit T un tore déployé maximal de $\mathbf{GL}(V)$. L'action de T sur V décompose V en somme directe de droites D_i . Si, dans la définition de X , on se restreint aux W qui sont sommes directes de certaines des D_i , on obtient un sous-complexe C de X qui est isomorphe au *complexe de Coxeter* du groupe symétrique S_n . D'un point

de vue combinatoire, c'est la subdivision barycentrique du bord d'un $(n - 1)$ -simplexe ; topologiquement, c'est une sphère de dimension $n - 2$. Un sous-complexe de X obtenu de cette façon est appelé un *appartement* de X . Par construction, les appartements correspondent aux tores déployés maximaux de $\mathbf{GL}(V)$.

1.2.6. *Exemple.* Prenons $n = 3$. L'immeuble correspondant est un *graphe*, qui a deux types de sommets : ceux qui correspondent aux points du plan projectif et ceux qui correspondent aux droites. Deux sommets sont voisins (i.e. sont les extrémités d'une arête) si et seulement si ils correspondent à un point situé sur une droite : la relation de voisinage est la relation d'incidence. Les appartements sont les hexagones $a-B-c-A-b-C-a$ associés aux triangles ABC , de côtés $a = BC$, $b = CA$, $c = AB$: la théorie de Tits transforme triangles en hexagones !

1.3. Traductions immobilières : le cas de $\mathbf{GL}(V)$

Si V est un Γ -module, le groupe Γ opère de façon naturelle sur l'immeuble $X = X(V)$ du n° 1.2. Cette action respecte les types (au sens du n° 1.2.2). En particulier, si un simplexe s de X est stable par Γ , il est fixé par Γ . Si P est le sous-groupe parabolique correspondant à s , P est normalisé par Γ , i.e. contient Γ (puisque'un parabolique est son propre normalisateur). Le sous-espace X^Γ des points fixes de Γ est un *sous-complexe simplicial* de X , et les définitions du n° 1.1 se traduisent de la façon suivante :

(1.3.1) V est *irréductible* $\Leftrightarrow X^\Gamma = \emptyset \Leftrightarrow \Gamma$ n'est contenu dans aucun sous-groupe parabolique propre de $\mathbf{GL}(V)$.

(1.3.2) V est *complètement réductible* \Leftrightarrow pour tout sommet x de X^Γ il existe un sommet x' de X^Γ qui est opposé à x (cf. 1.2.3) \Leftrightarrow pour tout parabolique maximal P contenant Γ , il existe un parabolique P' opposé à P qui contient Γ .

(1.3.3) V est *indécomposable* $\Leftrightarrow X^\Gamma$ ne contient aucun couple de sommets opposés \Leftrightarrow Il n'existe pas de couple (P, P') de paraboliques propres de $\mathbf{GL}(V)$ qui soient opposés et contiennent tous deux Γ .

Il n'est pas difficile de montrer que (1.3.2) équivaut à :

(1.3.2') pour tout simplexe s de X^Γ , il existe un simplexe s' de X^Γ qui est opposé à $s \Leftrightarrow$ pour tout parabolique P (maximal ou pas) contenant Γ , il existe un parabolique opposé à P qui contient Γ .

2. LA COMPLÈTE RÉDUCTIBILITÉ DANS LES IMMEUBLES SPHÉRIQUES

2.1. Immeubles sphériques

Un immeuble sphérique est un complexe simplicial X , muni d'une famille de sous-complexes appelés *appartements*. Je renvoie à [Ti 74] pour la liste des axiomes ; voir aussi [Br 89], [Ron 89] et [TW 02, § 40]. Voici quelques-unes des propriétés de ces immeubles :

2.1.1. *Dimension et rang.* Le complexe X est de dimension finie. L'entier $r = \dim(X) + 1$ est appelé le *rang* de X . Lorsque $X = \emptyset$, on convient que $\dim(X) = -1$, de sorte que $r = 0$. Tout simplexe maximal est de dimension $\dim(X)$.

2.1.2. *Types des sommets.* Soit $\text{som}(X)$ l'ensemble des sommets de X . Il existe une unique relation d'équivalence R sur $\text{som}(X)$ ayant les propriétés suivantes :

(a) L'ensemble quotient $\text{som}(X)/R$ a r éléments.

(b) Deux sommets appartenant à un même simplexe ne sont R -équivalents que s'ils sont égaux.

Si $x \in \text{som}(X)$, l'image de x dans $I = \text{som}(X)/R$ est appelée le *type* de x , et notée $\text{type}(x)$. On dit qu'un automorphisme f de X *préserve les types* si x et $f(x)$ ont même type quel que soit $x \in \text{som}(X)$.

Exemple. Lorsque X est l'immeuble $X(V)$ du n° 1.2, on peut identifier I à $\{1, \dots, n-1\}$, cf. 1.2.2.

2.1.3. *Appartements.* Un appartement est isomorphe au complexe de Coxeter d'un groupe de Coxeter fini qui ne dépend que de X ; c'est une sphère de dimension $r - 1$. Deux simplexes quelconques sont contenus dans un appartement.

2.1.4. *Opposition et géodésiques.* Soient x et y deux points de X (c'est-à-dire de sa réalisation géométrique), et soit A un appartement les contenant. On dit que x et y sont *opposés* dans X s'ils le sont dans A (ce qui a un sens puisque A est un complexe de Coxeter) ; cela ne dépend pas du choix de A . Deux simplexes sont dits opposés si chaque sommet de l'un est opposé d'un sommet de l'autre. Si x et y ne sont pas opposés, il y a une unique géodésique xy qui les joint dans A . Elle est indépendante (paramétrisation comprise) du choix de A .

2.1.5. *Convexité.* Une partie Y de X est dite *convexe*, si l'on a $xy \subset Y$ pour tout couple de points x, y de Y , non opposés. On dit que Y est *strictement convexe* si Y est convexe et ne contient aucun couple de points opposés (c'est la notion de «convexité» de [Ti 74] et de [Mu 65, p. 63]).

2.1.6. *Sphères de Levi.* Une sphère de Levi est un sous-complexe S de X , qui est contenue dans un appartement A , et qui est l'intersection de A (vu comme sphère) avec un sous-espace vectoriel.

2.1.7. *Remarque.* Ces définitions se présentent de façon un peu plus naturelle si l'on introduit l'immeuble vectoriel X^{vect} associé à X , dans lequel les points de X sont remplacés par des demi-droites, ayant en commun un point «0» (cf. [Rou 78]). Les appartements deviennent alors des espaces vectoriels de dimension r , et les sphères de Levi des sous-espaces vectoriels définis par l'annulation de certaines racines. Si x et y sont deux points de X^{vect} , leur somme $x + y$ a un sens, et l'on a

$$x + (y + z) = (x + y) + z$$

pourvu que x, y et z appartiennent à un même appartement. Deux points x et x' sont opposés si $x + x' = 0$. Une partie de X est convexe si le cône correspondant de X^{vect} est stable par $(x, y) \mapsto x + y$, autrement dit si son intersection avec tout appartement est un cône convexe au sens usuel du terme.

2.1.8. *Immeuble résiduel.* Soit s un simplexe de X de dimension m . On note X_s , ou $\text{St}(s)$, l'immeuble résiduel («link») de X en s , cf. [Ti 74]). Rappelons que les simplexes de X_s de dimension d correspondent bijectivement aux simplexes de X contenant s de dimension $d + m + 1$ (de sorte que le simplexe vide de X_s correspond à s). Si Y est un sous-complexe de X , les simplexes de Y contenant s définissent un sous-complexe Y_s de X_s ; si Y est convexe, il en est de même de Y_s .

Soit S une sphère de Levi, et soient s et s' deux simplexes de S de dimension maximale. Il y a un isomorphisme canonique $\text{proj} : X_s \rightarrow X_{s'}$ (défini dans [Ti 74, §3.19]). Ces isomorphismes satisfont à la condition de transitivité usuelle, ce qui permet d'écrire X_S à la place de X_s . L'immeuble X_S peut être appelé *l'immeuble* de S . Les sphères de Levi de X_S correspondent bijectivement aux sphères de Levi de X contenant S .

2.2. Complète réductibilité et contractibilité

2.2.1. *Définition.* Une partie Y de X est dite *complètement réductible* (en abrégé : X -cr, ou simplement cr) si elle est convexe, et si, pour tout point $y \in Y$, il existe $y' \in Y$ qui est opposé à y .

Dans la suite, nous nous intéresserons surtout au cas où Y est un sous-complexe convexe de X (ou, parfois, de sa subdivision barycentrique); dans ce cas, la condition cr équivaut à dire que, pour tout simplexe s de Y , il existe un simplexe s' de Y qui est opposé à s (il suffit même que tout sommet de Y ait un opposé dans Y , cf. th. 2.2 ci-dessous).

Exemple de sous-complexe convexe. Prenons pour X l'immeuble $X(V)$ du n° 1.2. Soit L un ensemble de sous-espaces vectoriels de V tel que :

$$W, W' \in L \implies W \cap W' \in L \text{ et } W + W' \in L.$$

Soit Y_L le sous-complexe plein de $X(V)$ dont les sommets sont les x_W , pour $W \in L$ et $W \neq 0, V$ (cf. 1.2.1). Alors Y_L est *convexe*, et l'on obtient ainsi tous les sous-complexes convexes de $X(V)$.

2.2.2. Nous allons donner un critère topologique permettant de reconnaître si un sous-complexe convexe est cr. Précisons que nous munissons X de la topologie limite inductive : une partie de X est ouverte si ses intersections avec les sous-complexes finis de X le sont. (Une autre topologie est souvent utile : celle définie par la distance angulaire ; elle n'interviendra pas ici.)

Rappelons d'autre part qu'un espace est *contractile* s'il a le type d'homotopie d'un point ; un espace discret est contractile si et seulement si son cardinal est égal à 1 (l'ensemble vide n'est pas contractile !).

THÉORÈME 2.1 ([Se 97b]). — *Soit Y un sous-complexe convexe de X . Les propriétés suivantes sont équivalentes :*

- (a) Y est X -cr.
- (b) Y contient un couple (s, s') de simplexes opposés ayant même dimension que Y .
- (c) Y contient une sphère de Levi S de dimension $\dim(Y)$.
- (d) Y n'est pas contractile.

Démonstration.— (a) \Rightarrow (b) est clair.

(b) \Rightarrow (c) : le plus petit sous-complexe convexe $C(s, s')$ contenant s et s' est une sphère de Levi.

(c) \Rightarrow (d) : la sphère S définit un cycle dans Y qui n'est pas homologue à 0 puisqu'il est de même dimension que Y (ceci ne vaut que si $\dim(Y) \geq 1$, mais le cas $\dim(Y) \leq 0$ est immédiat).

(d) \Rightarrow (a) : si Y n'est pas cr, il existe un point $y \in Y$ qui n'a pas d'opposé dans Y , et l'on contracte Y grâce aux géodésiques issues de y .

THÉORÈME 2.2. — *Les conditions (a), ..., (d) du th. 2.1 sont équivalentes à :*

- (e) *Pour tout sommet x de Y , il existe un sommet x' de Y qui est opposé à x .*

La démonstration sera donnée au n° 2.2.5 ci-dessous.

2.2.3. *Remarques.* 1) Un résultat analogue au th. 2.1 vaut pour les sous-complexes convexes de la subdivision barycentrique de X , à condition d'utiliser d'autres sphères que les sphères de Levi.

2) Supposons que Y soit cr, et non vide. On peut préciser sa structure topologique de la façon suivante :

Choisissons un simplexe s de Y de dimension maximum, et soit $U(s)$ l'ensemble des simplexes de Y qui sont opposés à s . Si $t \in U(s)$, soit $S_t = C(s, t)$ la sphère de Levi définie par s et t . Soit $B = *S_t$ le bouquet des sphères S_t ($t \in U(s)$) ; le choix d'un point x de s permet d'envoyer B dans Y .

PROPOSITION 2.3. — *L'application $B \rightarrow Y$ ainsi définie est une équivalence d'homotopie.*

(Autrement dit, Y a le type d'homotopie d'un bouquet de n -sphères, où $n = \dim(Y)$.)

Démonstration.— Soit x un point intérieur à s , et soit Y' le complexe obtenu en retirant de Y les intérieurs des simplexes appartenant à $U(s)$. Si $y \in Y'$, il est clair que y n'est pas opposé à x ; de plus, on peut montrer que la géodésique xy est contenue dans Y' . Il en résulte que Y' est contractile. On peut donc contracter Y' en un point sans changer le type d'homotopie de Y . On obtient ainsi le bouquet de sphères B , et la prop.2.3 s'en déduit. (Noter que cet argument est le même que celui employé par Solomon-Tits [So 69] dans le cas particulier où $Y = X$.)

On peut aussi se placer à un point de vue homologique. Il est commode d'utiliser les groupes d'homologie réduits $\tilde{H}_i(Y, \mathbf{Z})$; rappelons que ces groupes sont égaux aux groupes d'homologie usuels si $i > 0$ et que $\tilde{H}_0(Y, \mathbf{Z})$ et $\tilde{H}_{-1}(Y, \mathbf{Z})$ sont respectivement le noyau et le conoyau de l'homomorphisme $H_0(Y, \mathbf{Z}) \rightarrow \mathbf{Z}$. En particulier $\tilde{H}_{-1}(Y, \mathbf{Z})$ est 0 si $Y \neq \emptyset$ et est \mathbf{Z} si $Y = \emptyset$. Si Y est contractile, tous les $\tilde{H}_i(Y, \mathbf{Z})$ sont nuls.

PROPOSITION 2.4. — *Si Y est un sous-complexe convexe de dimension n , on a $\tilde{H}_i(Y, \mathbf{Z}) = 0$ pour $i \neq n$, et $\tilde{H}_n(Y, \mathbf{Z})$ est un groupe abélien libre de rang égal à $\text{Card}(U(s))$; il est $\neq 0$ si et seulement si Y est cr.*

Cela résulte de la prop. 2.3.

2.2.4. *Réduction.* L'énoncé suivant permet souvent de passer de X à l'un des immeubles résiduels X_s du n° 2.1.8.

PROPOSITION 2.5. — *Soit Y un sous-complexe convexe de X , et soit S une sphère de Levi contenue dans Y . Soit X_S l'immeuble associé à S , et soit Y_S le sous-complexe de X_S défini par Y . Pour que Y soit X -cr, il faut et il suffit que Y_S soit X_S -cr.*

(Précisons comment est défini Y_S : on choisit un simplexe s de S de dimension maximum, et l'on prend l'image de Y_s par l'isomorphisme naturel $X_s \rightarrow X_S$.)

La démonstration utilise le lemme suivant ([Ti 74, p. 54]) :

LEMME 2.6. — *Soit $\{s, s'\}$ un couple de simplexes opposés, et soient t_1, t_2 deux simplexes de X_s (identifiés à des simplexes de X contenant s). Soit t'_1 le simplexe de $X_{s'}$ correspondant à t_1 par l'isomorphisme $\text{proj} : X_s \rightarrow X_{s'}$. Alors :*

$$t_1 \text{ et } t_2 \text{ sont opposés dans } X_s \Leftrightarrow t'_1 \text{ et } t_2 \text{ sont opposés dans } X.$$

Démonstration de la prop. 2.5. Soit s un simplexe de S de dimension maximum, et soit t_1 un simplexe de Y contenant s et de dimension égale à $\dim(Y)$. Supposons d'abord que Y_s soit X_s -cr. Soit s' l'unique simplexe de S opposé à s . Puisque Y_s est cr, il existe un simplexe t_2 de Y contenant s , qui est opposé à t_1 dans Y_s . Le simplexe t'_1 du lemme 2.6 est opposé à t_2 , et est contenu dans Y du fait que Y est convexe (utiliser la définition de l'isomorphisme $\text{proj} : X_s \rightarrow X_{s'}$ donnée dans [Ti 74, §3.19]). On en déduit que Y contient la sphère de Levi définie par $\{t_2, t'_1\}$, d'où le fait que Y est cr d'après le th. 2.1. Inversement, si Y est cr, il contient un simplexe opposé à t_1 , d'où une sphère de Levi S'

de même dimension que Y ; le sous-complexe S'_s de X_s est une sphère de Levi contenue dans Y_s et de même dimension ; il en résulte que Y_s est cr.

2.2.5. *Démonstration du théorème 2.2.* Il s'agit de prouver que (e) \Leftrightarrow (a). L'implication (a) \Rightarrow (e) est évidente. On prouve (e) \Rightarrow (a) par récurrence sur $\dim(X)$. Si $Y = \emptyset$, l'énoncé est clair. Sinon, choisissons un sommet y de Y , que nous identifions à un simplexe de dimension 0. Puisque Y satisfait à (e), il contient un sommet y' opposé à y . Le couple $\{y, y'\}$ est une sphère de Levi de dimension 0. Vu la prop. 2.5, pour prouver que Y est X -cr il suffit de montrer que le sous-complexe Y_y de l'immeuble résiduel X_y est X_y -cr. Comme $\dim(X_y) = \dim(X) - 1$, on peut appliquer l'hypothèse de récurrence au couple (X_y, Y_y) ; il suffit donc de prouver que Y_y a la propriété (e). Cela revient à montrer que, pour toute arête yz de Y d'extrémité y , il existe une autre arête yz_1 qui est opposée à la précédente dans X_y . Choisissons un sommet z' de Y opposé à z , ce qui est possible d'après (e). Il existe une sphère de Levi D de dimension 1 (« cercle de Levi ») contenant z' et yz . De plus, z' et yz sont contenus dans le demi-cercle formé de la réunion de yz et de la géodésique yz' . Soit z_1 le sommet de la géodésique yz' qui est le plus proche de y , tout en étant distinct de y . L'image D_y de D dans X_y est égale à $\{yz, yz_1\}$, et c'est une sphère de Levi de dimension 0 de X_y ; il en résulte que yz et yz_1 sont opposés dans X_y , comme on le désirait.

2.3. Groupes agissant sur X

2.3.1. Soit Γ un groupe agissant sur X , i.e. muni d'un homomorphisme $\Gamma \rightarrow \text{Aut}(X)$. Soit X^Γ le sous-espace de X fixé par Γ . Il est clair que X^Γ est convexe : si Γ fixe deux points x et y qui ne sont pas opposés, il fixe la géodésique xy . Par analogie avec le n° 1.3, nous dirons que l'action de Γ sur X est :

- *irréductible*, si $X^\Gamma = \emptyset$;
- *complètement réductible*, si X^Γ est cr ;
- *indécomposable* si X^Γ est strictement convexe (cf. n° 2.1.5).

Comme précédemment, nous utiliserons les abréviations ir, cr et ind.

Remarque : L'espace X^Γ est un sous-complexe de la subdivision barycentrique de X . Si l'action de Γ préserve les types (ce qui sera le cas dans les §§ 3,4,5), c'est même un sous-complexe de X . D'après le th. 2.1 (complété par la Remarque 2.2.3.1), X^Γ est contractile si et seulement si l'action de Γ sur X n'est pas cr.

2.3.2. Voici une propriété de réduction, au sens du n° 2.2.4 :

PROPOSITION 2.7. — *Supposons que Γ préserve les types, et qu'il fixe une sphère de Levi S , auquel cas il opère sur l'immeuble X_S correspondant. On a l'équivalence suivante :*

L'action de Γ sur X est cr \Leftrightarrow L'action de Γ sur X_S est cr.

Cela résulte de la prop. 2.5, appliquée au sous-complexe $Y = X^\Gamma$ de X .

2.4. La conjecture du point fixe

CONJECTURE 2.8. — *Soit Y un sous-complexe de X (ou de sa subdivision barycentrique), convexe et contractile. Il existe alors un point de Y qui est fixé par tout automorphisme de X qui stabilise Y .*

(Un tel point mérite d'être appelé un *centre* de Y .)

Cette conjecture a été faite par Tits dans les années 50, sous l'hypothèse supplémentaire que Y est strictement convexe ; son but était, semble-t-il, de prouver un résultat sur les groupes unipotents que Borel et lui ont démontré ensuite par une méthode différente, cf. [BT 71]. Sous cette forme plus restrictive, la conjecture est signalée par Mumford [Mu 65, p. 64] à cause de ses relations avec la «Geometric Invariant Theory» (G.I.T.). En fait, le cas particulier utile pour G.I.T. a été démontré en 1978 par Kempf [Ke 78] et Rousseau [Rou 78]. Il y a d'ailleurs beaucoup d'autres cas où 2.8 a été démontrée, cf. [Rou 78], [Ti 97] et [Mü 97].

PROPOSITION 2.9. — *Admettons la conjecture 2.8. Soit Y un sous-complexe convexe contractile de X et soit Γ un groupe d'automorphismes de X préservant les types, et stabilisant Y . Alors Y^Γ est contractile.*

Il est clair que Y^Γ est un sous-complexe convexe de X . D'après 2.8, il est non vide. Il s'agit de montrer qu'il n'est pas cr. S'il l'était, il contiendrait une sphère de Levi S de même dimension, cf. th. 2.1. L'image Y_S de Y dans X_S est stable par Γ , et est contractile (prop. 2.5). En lui appliquant la conjecture 2.8, on en déduirait que $(Y^\Gamma)_S = (Y_S)^\Gamma$ est non vide, ce qui est impossible puisque S et Y^Γ ont la même dimension.

PROPOSITION 2.10. — *La prop. 2.9 est vraie (sans supposer que la conjecture 2.8 le soit) dans chacun des deux cas suivants :*

- (a) $\dim(Y) \leq 1$.
- (b) *L'image de Γ dans $\text{Aut}(Y)$ est un groupe résoluble fini.*

Le cas (a) est clair si $\dim(Y) = 0$, car Y est réduit à un point ; il est facile si $\dim(Y) = 1$ car Y est un arbre de diamètre borné, et un tel arbre a un centre, à savoir le milieu des chemins sans aller-retour de longueur égale au diamètre.

Pour (b), on se ramène par dévissage au cas où Γ agit sur Y par un groupe cyclique d'ordre premier p . Comme les $\tilde{H}_i(Y, \mathbf{Z}/p\mathbf{Z})$ sont tous nuls, il en est de même des $\tilde{H}_i(Y^\Gamma, \mathbf{Z}/p\mathbf{Z})$ d'après la théorie de Smith; en effet, cette théorie dit que, si $\Gamma \simeq \mathbf{Z}/p\mathbf{Z}$ opère sur un complexe Y de dimension finie, et si les $\tilde{H}_i(Y, \mathbf{Z}/p\mathbf{Z})$ sont nuls pour $i > N$ (où N est un entier fixé), il en est de même des $\tilde{H}_i(Y^\Gamma, \mathbf{Z}/p\mathbf{Z})$ pour $i > N$. Or, si Y n'était pas contractile, il serait cr, et le groupe $\tilde{H}_i(Y^\Gamma, \mathbf{Z}/p\mathbf{Z})$ correspondant à $i = \dim(Y^\Gamma)$ serait non nul d'après la prop. 2.4.

Action de sous-groupes normaux

PROPOSITION 2.11. — *Admettons la conjecture 2.8. Soit Γ un groupe d'automorphismes de X respectant les types, et soit Γ' un sous-groupe normal de Γ . Si l'action de Γ est cr, il en est de même de celle de Γ' .*

(Comparer avec le résultat bien connu suivant : si une représentation linéaire d'un groupe Γ est complètement réductible, il en est de même de ses restrictions aux sous-groupes normaux de Γ , cf. e.g. [Se 94, lemme 5].)

Posons $Y' = X^{\Gamma'}$ et $Y = X^{\Gamma} = Y'^{\Gamma/\Gamma'}$. Il s'agit de montrer que Y' n'est pas contractile. S'il l'était, la prop. 2.9, appliquée à (X, Y', Γ) , montrerait que $Y = X^{\Gamma}$ est contractile, contrairement à l'hypothèse faite.

Un argument analogue, utilisant la prop. 2.10, démontre :

PROPOSITION 2.12. — *Si Γ/Γ' est un groupe résoluble fini, la prop. 2.11 est vraie sans supposer que la conjecture 2.8 le soit.*

Nous verrons au § 3.3 un autre cas du même genre.

3. COMPLÈTE RÉDUCTIBILITÉ DES SOUS-GROUPES D'UN GROUPE RÉDUCTIF

Dans ce qui suit, G désigne un groupe algébrique réductif sur un corps k (cf. [BT 65]). Rappelons qu'un tel groupe est lisse et connexe.

Par un «sous-groupe algébrique» de G , on entend un k -sous-groupe algébrique («défini sur k »).

3.1. L'immeuble de G

3.1.1. *Sous-groupes paraboliques.* Un sous-groupe parabolique de G est un sous-groupe algébrique P tel que G/P soit une variété projective (définition équivalente : après extension des scalaires, P contient un sous-groupe de Borel). Un tel groupe est lisse, connexe, et coïncide avec son normalisateur dans G . Les propriétés de ces groupes dont nous aurons besoin se trouvent dans [BT 65] ; voir aussi [DG 70, XXVI].

3.1.2. *L'immeuble de G .* Sa définition est donnée dans [Ti 74, § 5]. On le notera $X(G)$ ou simplement X . Ses simplexes correspondent aux sous-groupes paraboliques de G ; si s est un simplexe, on note P_s le sous-groupe parabolique correspondant. Les sommets de X correspondent aux paraboliques propres maximaux ; le simplexe vide correspond à G . Le rang r de X est égal au k -rang semi-simple de G , c'est-à-dire au k -rang (ou *rang relatif*) de $G^{\text{ad}} = G/C_G$, où C_G est le centre de G . On a $r = 0$ (et $X = \emptyset$) si G^{ad} est anisotrope.

3.1.3. *Appartements*. Ils correspondent aux tores déployés maximaux (de G , ou de G^{ad} , c'est la même chose).

3.1.4. *Types de sommets*. L'ensemble I des types de sommets peut être identifié à l'ensemble des sommets du k -diagramme de Dynkin de G .

3.1.5. *Opposition*. Deux simplexes s et s' de X sont opposés si et seulement si les paraboliques P_s et $P_{s'}$ sont opposés au sens de [BT 65, § 4], i.e. si $P_s \cap P_{s'}$ est réductif, auquel cas c'est un sous-groupe de Levi de chacun d'eux.

3.1.6. *Action de $G(k)$* . Le groupe $G(k)$ des k -points de G opère sur X (par conjugaison des paraboliques). Cette action respecte les types. Si s est un simplexe de X , le sous-groupe de $G(k)$ fixant (ou stabilisant) s est $P_s(k)$.

3.1.7. *Sphères de Levi*. Soit L un sous-groupe de Levi d'un parabolique. Les sous-groupes paraboliques contenant L correspondent aux simplexes d'une *sphère de Levi* S_L , et l'on obtient ainsi une bijection entre les L et les sphères de Levi (ce qui explique la terminologie utilisée au § 2). Les paraboliques ayant L pour sous-groupe de Levi correspondent aux simplexes de dimension maximale de la sphère S_L . Si s est l'un de ces simplexes, l'immeuble résiduel X_s (cf. 2.1.8) peut être identifié à l'immeuble de L .

3.1.8. *Critère de convexité*. Soit Y un sous-complexe de X , et soit H l'ensemble des paraboliques correspondant aux simplexes de Y .

PROPOSITION 3.1. — *Pour que Y soit convexe, il faut et il suffit que H satisfasse à la propriété suivante :*

(C) *Si trois paraboliques P, P', Q sont tels que $P \in H, P' \in H$, et $Q \supset P \cap P'$, alors $Q \in H$.*

(Attention : l'inclusion $Q \supset P \cap P'$ est une inclusion de groupes algébriques. Il ne suffit pas que $Q(k)$ contienne $P(k) \cap P'(k)$.)

L'énoncé revient à déterminer l'enveloppe convexe de la réunion de deux simplexes (ceux correspondant à P et P'), ce qui se fait au moyen d'un appartement les contenant tous deux.

3.1.9. *Relations avec les sous-groupes multiplicatifs à 1 paramètre (G.I.T.)*. Soit $\lambda : \mathbf{G}_m \rightarrow G$ un homomorphisme. On peut lui associer de façon naturelle deux points opposés $h_+(\lambda)$ et $h_-(\lambda)$ de l'immeuble vectoriel X^{vect} , cf. 2.1.7. Si λ est à valeurs dans le centre de G , on a $h_+(\lambda) = h_-(\lambda) = 0$. Sinon, les demi-droites engendrées par $h_+(\lambda)$ et $h_-(\lambda)$ définissent deux points $x_+(\lambda)$ et $x_-(\lambda)$ de l'immeuble X ; ces points sont opposés. Soient $s_+(\lambda)$ et $s_-(\lambda)$ les plus petits simplexes de X contenant respectivement $x_+(\lambda)$ et $x_-(\lambda)$, et soient $P_+(\lambda)$ et $P_-(\lambda)$ les sous-groupes paraboliques correspondants. Le groupe $P_+(\lambda)$ est formé des points g de G qui sont *contractés* par λ , i.e. tels que $\lambda(t).g.\lambda(t^{-1})$ ait une limite pour $t \rightarrow 0$ (cf. [Mu 65, p. 55] ou [Ri 88, § 2]). De même, $P_-(\lambda)$ est l'ensemble

des g tels que $\lambda(t).g.\lambda(t^{-1})$ ait une limite pour $t \rightarrow \infty$, et le groupe de Levi $P_+(\lambda) \cap P_-(\lambda)$ est le *centralisateur* de l'image de λ .

Remarque.— Les $h_+(\lambda)$ jouent le rôle de *points entiers* pour l'immeuble vectoriel, et les $x_+(\lambda)$ sont les *points rationnels* de l'immeuble sphérique. L'interprétation des sous-groupes paraboliques en termes de contractions est à la base de la "Geometric Invariant Theory" ; elle joue un rôle essentiel dans les résultats de Richardson et de Bate-Martin-Röhrle cités plus loin.

3.2. Les propriétés G -ir, G -cr et G -ind

3.2.1. À partir de maintenant, Γ désigne un sous-groupe de $G(k)$. Son action sur X permet de lui appliquer les définitions du n° 2.3.1 : irréductibilité, complète réductibilité et indécomposabilité. Pour mettre G en évidence, nous écrirons G -ir, G -cr et G -ind. Autrement dit :

Γ est G -ir $\Leftrightarrow \Gamma$ n'est contenu dans aucun sous-groupe parabolique propre de G .

Γ est G -cr \Leftrightarrow Pour tout parabolique P de G contenant Γ , il existe un sous-groupe de Levi de P contenant Γ (ou, ce qui revient au même, il existe un parabolique P' opposé à P tel que $\Gamma \subset P(k) \cap P'(k)$).

Γ est G -ind $\Leftrightarrow \Gamma$ n'est contenu dans aucun sous-groupe de Levi d'un sous-groupe parabolique propre de G .

3.2.2. *Exemples.* Lorsque G est un groupe classique (ou un groupe de type G_2) la notion de « G -cr» peut se traduire très concrètement :

(a) Lorsque $G = \mathbf{GL}(V)$, elle signifie que le Γ -module V est semi-simple, cf. n° 1.3.

(b) Supposons k de caractéristique $\neq 2$, et prenons pour G un groupe $\mathbf{SO}(V)$ (ou $\mathbf{Sp}(V)$), relatif à une forme bilinéaire symétrique (ou alternée) B sur V , non dégénérée. La définition de « G -cr» donnée ci-dessus dit que Γ est G -cr si et seulement si, pour tout sous- Γ -module totalement isotrope W de V , il existe un autre sous- Γ -module totalement isotrope W' , de même dimension, tel que la restriction de B à $W + W'$ soit non dégénérée. Un argument élémentaire permet de montrer que *cela se produit si et seulement si le Γ -module V est semi-simple* (c'est aussi une conséquence de la théorie de Richardson, du moins quand k est algébriquement clos, cf. [Ri 88, cor. 16.10]).

(c) Supposons G de type G_2 , et k de caractéristique $\neq 2$. Soit V l'unique représentation irréductible de G de dimension 7. Ici encore, on peut montrer que Γ est G -cr si et seulement si le Γ -module V est semi-simple. Cela se voit en utilisant la description des paraboliques donnée dans [As 86].

On verra au § 5 des résultats analogues pour d'autres représentations – mais on devra alors éviter d'autres caractéristiques que la caractéristique 2.

3.2.3. La proposition suivante est une conséquence immédiate de la prop. 2.5 :

PROPOSITION 3.2. — *Supposons que Γ soit contenu dans un sous-groupe de Levi L d'un sous-groupe parabolique de G . On a alors*

$$\Gamma \text{ est } G\text{-cr} \iff \Gamma \text{ est } L\text{-cr.}$$

(Lorsque $G = \mathbf{GL}_n$, cela redonne (1.1.5).)

3.2.4. On peut définir un « G -analogue» de la *semi-simplification* d'une représentation :

Choisissons un parabolique P contenant Γ et minimal pour cette propriété (cela revient à choisir un simplexe de dimension maximum de X^Γ). Soit L un sous-groupe de Levi de P et soit $\pi : P \rightarrow L$ la projection de P sur L de noyau le radical unipotent $R_u(P)$ de P .

PROPOSITION 3.3. — (a) *Le groupe $\pi(\Gamma) \subset L(k)$ est L -ir et G -cr.*

(b) *Différents choix de (P, L) donnent des homomorphismes $\Gamma \rightarrow L(k) \rightarrow G(k)$ qui sont conjugués par $G(k)$.*

(Lorsque $G = \mathbf{GL}(V)$, l'homomorphisme $\Gamma \rightarrow L(k) \rightarrow G(k)$ est le semi-simplifié de $\Gamma \rightarrow G(k)$, et l'assertion d'unicité de (b) est le *théorème de Jordan-Hölder*.)

Dans (a), le fait que $\pi(\Gamma)$ soit L -ir provient de ce que P est minimal ; on en déduit que Γ est G -cr en appliquant la prop. 3.2. On prouve (b) en remarquant que, pour P fixé, le choix de L n'a pas d'importance puisque deux L différents sont conjugués par $R_u(P)(k)$; et, pour un autre choix P' de P , on utilise le fait que P et P' ont un sous-groupe de Levi commun (c'est une propriété générale des simplexes maximaux d'un sous-complexe convexe).

3.2.5. Voici un autre résultat, inspiré par des arguments de [Ri 88] et de [BMR 04] :

PROPOSITION 3.4. — *Soit C_Γ le centralisateur de Γ dans G . Soit T un tore déployé maximal de C_Γ , et soit L le centralisateur de T dans G . On a $\Gamma \subset L(k)$. De plus :*

(a) *L est un sous-groupe de Levi d'un parabolique de G ; il est minimal parmi tous les Levi de paraboliqes contenant Γ .*

(b) *Γ est L -ind.*

(c) *Pour que Γ soit G -cr, il faut et il suffit qu'il soit L -ir.*

(d) *Si k est parfait, les différents choix de L sont conjugués par $C_\Gamma(k)$.*

(Lorsque $G = \mathbf{GL}(V)$, le choix de L correspond à une décomposition du Γ -module V en somme directe de modules indécomposables, et (d) est le *théorème de Krull-Remak-Schmidt*.)

L'assertion (a) se déduit du fait que les sous-groupes de Levi de paraboliqes sont les centralisateurs des tores déployés de G , cf. [BT 65, 4.16]. L'assertion (b) provient de la minimalité de L , et (c) se déduit de (b) et de la prop. 3.2. Quant à (d), il résulte de la conjugaison des tores déployés maximaux de C_Γ , qui est valable quand k est parfait, d'après [BT 65, 11.6].

3.3. La propriété de « forte réductivité » de Richardson

On suppose maintenant que k est algébriquement clos, ce qui assure que tous les tores sont déployés. La prop. 3.4 (c) s'énonce alors de la façon suivante :

THÉORÈME 3.5 ([BMR 04]). — Soit T un tore maximal du centralisateur de Γ , et soit L le centralisateur de T . Les deux propriétés suivantes sont équivalentes :

- (i) Γ est G -cr.
- (ii) Γ n'est contenu dans aucun sous-groupe parabolique propre de L .

La propriété (ii) a été introduite en 1988 par Richardson [Ri 88, § 16] sous le nom de « strong reductivity ». Ce n'est que tout récemment que Bate, Martin et Röhrle ont démontré qu'elle équivaut à la propriété (i). Ils en ont tiré de nombreuses conséquences, que l'on trouvera dans [BMR 04]. En voici quelques unes :

THÉORÈME 3.6 ([Ma 03b] et [BMR 04]). — Soit Γ' un sous-groupe normal de Γ . Si Γ est G -cr, il en est de même de Γ' .

(Comparer avec la prop. 2.11 du § 2.)

THÉORÈME 3.7 ([Ri 88] et [BMR 04]). — Supposons que Γ soit engendré topologiquement par des éléments x_1, \dots, x_m . Soit $f : G \rightarrow G \times \dots \times G$ (m copies) l'application

$$g \mapsto (gx_1g^{-1}, \dots, gx_mg^{-1}).$$

Pour que Γ soit G -cr, il faut et il suffit que $f(G)$ soit une partie fermée de $G \times \dots \times G$.

Dans le cas particulier $m = 1$, on retrouve le fait qu'une classe de conjugaison est fermée si et seulement si ses éléments sont semi-simples.

THÉORÈME 3.8 ([BMR 04]). — Soit G' un groupe réductif contenant G , et tel que :

- (a) Le centralisateur (schématique) de G dans G' est lisse.
- (b) Il existe un sous-espace vectoriel \mathfrak{m} de Lie G' , stable par conjugaison par G et tel que $\text{Lie } G' = \mathfrak{m} \oplus \text{Lie } G$.

Alors, si Γ est G' -cr, il est G -cr.

(Dans [BMR 04], la condition (a) est appelée « séparabilité » ; quant à (b), elle exprime que (G', G) est un « couple réductif » au sens de Richardson.)

COROLLAIRE 3.9. — Soit V un G -module fidèle. Supposons que la forme bilinéaire $(x, y) \mapsto \text{Tr}(x_V \cdot y_V)$ sur Lie G soit non dégénérée. Alors, si V est Γ -semi-simple, le groupe Γ est G -cr.

Cela résulte du th. 3.8, appliqué à $G' = \mathbf{GL}(V)$. La condition (a) est satisfaite. La condition (b) l'est aussi : on prend pour \mathfrak{m} l'orthogonal de Lie G dans Lie $\mathbf{GL}(V)$ pour la forme trace.

4. CRITÈRES DE COMPLÈTE RÉDUCTIBILITÉ

Dans ce qui suit, G est un groupe réductif sur un corps algébriquement clos k , et Γ est un sous-groupe de $G(k)$. À partir du n° 4.2, on suppose que la caractéristique p de k est > 0 .

On se propose de donner des critères, aussi explicites que possible, permettant de reconnaître si Γ possède la propriété G -cr. Si $\bar{\Gamma}$ est l'adhérence de Γ pour la topologie de Zariski, il est clair que Γ est G -cr $\iff \bar{\Gamma}$ est G -cr. Cela nous permettra souvent de supposer que Γ est fermé, i.e. que c'est un sous-groupe algébrique (lisse) de G .

[Il serait intéressant de considérer aussi le cas d'un sous-groupe algébrique *non nécessairement lisse*. Il n'y a pas de difficulté à étendre à de tels groupes la définition de « G -cr», non plus que celle du sous-complexe convexe « X^Γ ». Ce qui est moins clair, c'est ce qui doit remplacer la *saturation* du § 5. Une fois cet obstacle surmonté, on peut espérer que les résultats des n°s 5.2 et 5.3 s'étendent sans changement.]

4.1. Une première condition

Notons $R_u(\Gamma)$ le radical unipotent de Γ , i.e. son plus grand sous-groupe unipotent normal (connexe ou non). (Lorsque k est de caractéristique $p > 0$, et que Γ est fini, on a $R_u(\Gamma) = O_p(\Gamma)$, avec les notations usuelles de la théorie des groupes finis.)

PROPOSITION 4.1. — *Si Γ est G -cr, on a $R_u(\Gamma) = 1$.*

Quitte à remplacer G par un Levi de l'un de ses sous-groupes paraboliques, on peut supposer que Γ est G -ir (cf. prop. 3.3). Soit alors U l'adhérence de $R_u(\Gamma)$. D'après [BT 71, prop. 3.1], il existe un sous-groupe parabolique P de G , avec $U \subset R_u(P)$, qui est stable par tout automorphisme du couple (G, U) . Comme Γ normalise U , il normalise P , donc est contenu dans P . Puisque Γ est G -ir, cela entraîne $P = G$, d'où $U = 1$ puisque $R_u(G) = 1$.

PROPOSITION 4.2. — *Supposons k de caractéristique 0, et supposons que Γ soit fermé. Les propriétés suivantes sont équivalentes :*

- (i) Γ est G -cr.
- (ii) $R_u(\Gamma) = 1$.
- (iii) La composante neutre Γ^0 de Γ est un groupe réductif.

L'équivalence de (ii) et (iii) provient de ce que tout sous-groupe unipotent de Γ est contenu dans Γ^0 . L'implication (i) \implies (ii) est la prop. 4.1. L'implication (iii) \implies (i) provient du fait bien connu suivant (spécial à la caractéristique 0) : toute extension de Γ par un groupe unipotent est scindée, et deux scindages quelconques sont conjugués (on se ramène, par dévissage à une annulation de groupes de cohomologie, cf. [Mo 56] et [DG 70, p. 393]).

COROLLAIRE 4.3. — *Supposons k de caractéristique 0. Soit $f : G \rightarrow G'$ un homomorphisme de G dans un groupe réductif G' . Si Γ est G -cr, alors $f(\Gamma)$ est G' -cr ; la réciproque est vraie si f est presque fidèle.*

(Un homomorphisme f est dit *presque fidèle* si $\text{Ker}(f)$ est un groupe de type multiplicatif.)

C'est clair, grâce à (iii).

Remarques.— 1) Le cor. 4.3 redonne le théorème de Chevalley cité au n° 1.1 : il suffit de l'appliquer à l'homomorphisme naturel $f : \mathbf{GL}(V) \times \mathbf{GL}(V') \rightarrow \mathbf{GL}(V \otimes V')$.

2) La prop. 4.2 montre que la propriété « G -cr» n'a pas grand intérêt en caractéristique 0. C'est pour cela que, à partir de maintenant, *on supposera que le corps k est de caractéristique $p > 0$* . On donnera alors des conditions sur p (du genre « p est assez grand») permettant d'avoir des résultats analogues à ceux de la caractéristique 0, cf. th. 4.4, th. 4.5 et th. 5.3.

4.2. Le cas où Γ est connexe

Définissons un entier $a(G)$ par la recette suivante :

(1) Si G est simple :

$$a(G) = 1 + \text{rang}(G).$$

(2) Si $\{G_1, \dots, G_r\}$ sont les quotients simples de G , on pose :

$$a(G) = \sup(1, a(G_1), \dots, a(G_r)).$$

THÉORÈME 4.4 (Jantzen, McNinch, Liebeck-Seitz). — *Supposons que $p \geq a(G)$, que Γ soit un sous-groupe fermé de G , et que $(\Gamma : \Gamma^0)$ soit premier à p . Il y a alors équivalence entre :*

(i) Γ est G -cr.

(ii) Γ^0 est réductif.

L'implication (i) \Rightarrow (ii) a été démontrée plus haut. Supposons que la condition (ii) soit satisfaite. En utilisant le fait que $(\Gamma : \Gamma^0)$ est premier à p , on démontre facilement que Γ^0 est G -cr $\Rightarrow \Gamma$ est G -cr. On peut donc supposer que Γ est connexe, autrement dit que c'est un sous-groupe réductif de G ; on peut aussi supposer que G est quasi-simple. Le cas où G est de type exceptionnel est traité dans [LS 96] (sauf pour $p = 3$ et G de type G_2 , mais ce cas n'offre pas de difficultés). Lorsque G est de type A_n , le théorème signifie que toute représentation linéaire de degré $\leq p$ d'un groupe réductif est semi-simple, ce qui a été démontré par Jantzen [Ja 97]. De même, si G est de type B_n, C_n ou D_n , on est ramené à montrer que toute représentation self-duale d'un groupe réductif est semi-simple si sa dimension est $< 2p$; cela a été démontré récemment par McNinch (non publié : le cas crucial est celui où le groupe réductif est de type A_1 – les autres cas se déduisent de [Mc 98]).

Remarque.— La borne $p \geq a(G)$ du th. 4.4 est essentiellement optimale lorsque le groupe Γ^0 est de rang 1. Elle peut par contre être améliorée lorsque les facteurs simples de Γ^0 sont de rang > 1 , cf. [LS 96] et [Mc 98].

4.3. Le cas non connexe

On se borne au cas où $G^{\text{ad}} = G/C_G$ est un groupe simple. On remplace l'entier $a(G)$ du n° 4.2 par un entier $b(G)$ un peu plus grand :

$$\begin{aligned} b(G) &= 2, 3, 5 && \text{si } G^{\text{ad}} \text{ est de type } A_1, A_2, B_2 ; \\ b(G) &= n + 3 && \text{si } G^{\text{ad}} \text{ est de type } A_n \ (n \geq 3) ; \\ b(G) &= 2n + 3 && \text{si } G^{\text{ad}} \text{ est de type } B_n, C_n \ (n \geq 3) \text{ ou } D_n \ (n \geq 4) ; \\ b(G) &= 11, 29, 29, 59, 251 && \text{si } G^{\text{ad}} \text{ est de type } G_2, F_4, E_6, E_7, E_8. \end{aligned}$$

THÉORÈME 4.5. — *Supposons que Γ soit un sous-groupe fermé de $G(k)$, avec G^{ad} simple, et $p \geq b(G)$. Il y a équivalence entre :*

(i) Γ est G -cr.

(ii) $R_u(\Gamma) = 1$.

(Autrement dit, on a le même énoncé que 4.2, pourvu que $p \geq b(G)$.)

Ici encore, il suffit de montrer que (ii) \Rightarrow (i). Le cas essentiel est celui où $G = \mathbf{GL}_n$; il est dû à Guralnick, cf. [Gu 99, th. C]. (La démonstration est loin d'être élémentaire : elle utilise non seulement la classification des groupes finis simples, mais aussi la liste des caractères modulaires irréductibles des groupes sporadiques donnée dans [JLPW 95].) Les autres cas s'en déduisent en appliquant le th. 5.4 ci-dessous à des représentations linéaires de G de basse dimension.

Remarques. 1) Voici quelques exemples de couples (Γ, G) montrant que, pour les groupes classiques, la condition $p \geq b(G)$ du th. 4.5 ne peut guère être améliorée :

Type A : $\Gamma = \mathcal{S}_p$; $G = \mathbf{SL}_{p-1}$

Types B, C, D : $\Gamma = \mathbf{SL}_2(\mathbf{F}_p)$ ou $\mathbf{PSL}_2(\mathbf{F}_p)$; $G = \mathbf{SO}_{p+2}, \mathbf{Sp}_{p-1}, \mathbf{SO}_{p+1}$.

Supposons $p > 3$. Dans chaque cas, on a $R_u(\Gamma) = 1$, et l'on peut construire un plongement de Γ dans $G(k)$ qui donne une représentation linéaire non semi-simple, ce qui signifie que Γ n'est pas G -cr, d'après 3.2.2. On a $p = b(G) - 2, b(G) - 4, b(G) - 4, b(G) - 2$ respectivement.

2) La situation est différente pour les groupes exceptionnels ; la borne $p \geq b(G)$ peut être grandement améliorée, par exemple en utilisant les méthodes de [LS 96]. Ainsi, pour le type G_2 , on peut remplacer « $p \geq 11$ » par « $p \geq 5$ », qui est optimal. J'ignore quelles sont les bornes optimales pour les types F_4, E_6, E_7 et E_8 .

5. SATURATION ET REPRÉSENTATIONS LINÉAIRES

5.1. Exponentielle et saturation

On note $h(G)$ la borne supérieure des nombres de Coxeter des quotients simples de G . (S'il n'y en a aucun, i.e. si G est un tore, on convient que $h(G) = 1$.) Rappelons que, si G est simple, on a $h(G) = \dim(G)/\text{rang}(G) - 1$; les valeurs de h pour les différents types sont :

$A_n : h = n + 1$; $B_n, C_n : h = 2n$; $D_n : h = 2n - 2$; $G_2 : h = 6$; $F_4, E_6 : h = 12$; $E_7 : h = 18$; $E_8 : h = 30$.

Supposons maintenant que $p \geq h(G)$. Soit u un élément unipotent de G . D'après [Te 95], on a $u^p = 1$. De plus, si t est un élément de k , on peut définir de façon canonique (c'est là un point essentiel) la « t -ième puissance» u^t de u , cf. [Se 98] (voir aussi [Sei 00] qui traite un cas plus général). L'application $t \mapsto u^t$ est un homomorphisme du groupe additif \mathbf{G}_a dans le groupe G . Lorsque $G = \mathbf{GL}_n$, l'hypothèse $p \geq h(G)$ signifie que $p \geq n$, de sorte que $u = 1 + v$ avec $v^p = 0$, et u^t est donné par le développement binomial : $(1 + v)^t = 1 + t.v + \dots$

[L'hypothèse « k algébriquement clos» faite au début du §4 n'intervient pas dans la définition de l'exponentielle u^t . En fait, le cas crucial est celui d'un schéma en groupes semi-simples, déployé et simplement connexe, sur le localisé $\mathbf{Z}_{(p)}$ de \mathbf{Z} en p . Les autres cas s'en déduisent par descente, en utilisant les méthodes de [DG 70], cf. [Sei 00, §5].]

DÉFINITION 5.1. — *Un sous-groupe Γ de $G(k)$ est dit saturé s'il est fermé et si l'on a $u^t \in \Gamma$ pour tout élément unipotent u de Γ , et tout $t \in k$.*

[Par exemple, tout sous-groupe parabolique est saturé ; tout centralisateur d'un sous-groupe est saturé.]

On démontre facilement :

PROPOSITION 5.2. — *Si Γ est saturé, l'indice de Γ^0 dans Γ est premier à p .*

Pour tout sous-groupe Γ de $G(k)$, il existe un plus petit sous-groupe saturé le contenant ; on l'appelle le saturé de Γ et on le note Γ^{sat} . (Lorsque $G = \mathbf{GL}_n$, on retrouve la notion utilisée dans [No 87] et [Se 94].)

THÉORÈME 5.3 ([Se 98, th. 8]). — *Il y a équivalence entre :*

- (i) Γ est G -cr.
 - (ii) Γ^{sat} est G -cr.
 - (iii) La composante neutre de Γ^{sat} est un groupe réductif.
- (Rappelons que l'on suppose $p \geq h(G)$.)

L'équivalence (i) \Leftrightarrow (ii) est claire : les sous-groupes paraboliques et leurs sous-groupes de Levi sont saturés. L'équivalence de (ii) et (iii) résulte du th. 4.4 et de la prop. 5.2 ; noter que le th. 4.4 est applicable car $a(G) \leq h(G)$.

5.2. Représentations linéaires: l'invariant $n(V)$

Choisissons un tore maximal T de G , ainsi qu'un sous-groupe de Borel B de G contenant T . Soit $X(T) = \text{Hom}(T, \mathbf{G}_m)$ le groupe des caractères de G , et soit $Y(T) = \text{Hom}(\mathbf{G}_m, T)$ son dual. Notons $R \subset X(T)$ le système de racines de (G, T) . Si $a \in R$, notons a^* la racine duale ; c'est un élément de $Y(T)$.

Pour tout $\chi \in X(T)$, on pose

$$n(\chi) = \sum \langle \chi, a^* \rangle,$$

où a parcourt les éléments > 0 de R (pour la relation d'ordre associée à B).

Si V est un G -module, on définit un *invariant* $n(V)$ de V par la formule :

$$n(V) = \sup n(\chi), \text{ où } \chi \text{ parcourt l'ensemble des poids de } T \text{ dans } V, \text{ cf. [Dy 52, n}^\circ \text{ 12].}$$

C'est un entier ≥ 0 . Voici quelques unes de ses propriétés (on en trouvera d'autres dans [Dy 52], [Se 98] et [IMP 03]) :

(5.2.1) Si V a une suite de composition dont les facteurs successifs sont V_1, \dots, V_r , on a $n(V) = \sup n(V_i)$.

(5.2.2) Si V est irréductible de plus grand poids λ , on a $n(V) = n(\lambda)$.

(5.2.3) On a $n(V) = 0$ si et seulement si le groupe dérivé de G opère trivialement sur V , i.e. si l'image de $G \rightarrow \mathbf{GL}(V)$ est un tore.

(5.2.4) Si V est presque fidèle (au sens du cor. 4.3), on a $n(V) \geq h(G) - 1$.

(5.2.5) Si $V = V_1 \otimes V_2$, on a $n(V) = n(V_1) + n(V_2)$.

(5.2.6) On a $n(\text{Lie } G) = 2h(G) - 2$.

Remarque.— L'invariant $n(V)$ «se calcule sur \mathbf{SL}_2 » au sens suivant :

Soit $f : \mathbf{SL}_2 \rightarrow G$ un homomorphisme tel que le composé $\mathbf{G}_m \rightarrow \mathbf{SL}_2 \rightarrow G$ soit à valeurs dans T , et soit égal dans $Y(T)$ à la somme $\sum a^*$, où a parcourt les racines > 0 . Un tel f existe : cela se démontre par réduction à partir de la caractéristique 0. Grâce à f , le G -module V peut être vu comme un \mathbf{SL}_2 -module, et son invariant $n(V)$ comme G -module est le même que son invariant $n(V)$ comme \mathbf{SL}_2 -module. [Autre interprétation de $n(V)$: à un facteur 2 près, c'est le degré en la variable « q » de la *dimension quantique* de V .]

L'intérêt de l'invariant $n(V)$ provient du théorème suivant, qui relie la semi-simplicité du Γ -module V à la propriété « G -cr» :

THÉORÈME 5.4. — *Supposons $p > n(V)$. Soit Γ un sous-groupe de $G(k)$.*

(i) *Si Γ est G -cr, alors V est Γ -semi-simple (i.e. semi-simple comme Γ -module).*

(ii) *Inversement, si V est Γ -semi-simple, et si V est presque fidèle, alors Γ est G -cr.*

(Si l'on regarde V comme un \mathbf{SL}_2 -module – cf. *Remarque* ci-dessus – l'hypothèse $p > n(V)$ signifie que V est un \mathbf{SL}_2 -module *restreint* («restricted») : ses poids sont $< p$.)

On peut supposer que V est presque fidèle. D'après (5.2.4), on a alors $p \geq h(G)$, ce qui permet d'utiliser le th. 5.3. Les détails de la démonstration se trouvent dans [Se 98] (voir aussi [IMP 03]).

COROLLAIRE 5.5. — *Supposons $p > 2h(G) - 2$. Les deux propriétés suivantes sont équivalentes :*

(cr) Γ est G -cr.

(ad) L 'algèbre de Lie $\text{Lie } G$ de G est un Γ -module semi-simple.

Cela résulte du th. 5.4 et de la formule (5.2.6).

Remarque 5.6. — La condition $p > 2h(G) - 2$ est presque optimale pour l'implication (cr) \Rightarrow (ad). En effet, supposons que cette condition ne soit pas satisfaite, et que G soit simple ; on peut alors construire (à deux exceptions près, cf. ci-après) un sous-groupe Γ de G , isomorphe à \mathbf{SL}_2 ou \mathbf{PGL}_2 , qui satisfait à (cr) mais pas à (ad). (Les deux exceptions sont : $p = 3, G$ de type B_2 , et $p = 5, G$ de type G_2 .)

Par contre, l'implication (ad) \Rightarrow (cr) est en général valable sous des conditions bien moins restrictives que $p > 2h(G) - 2$. Par exemple, si G est de type E_8 , le cor.3.9 montre que la condition $p > 2h(G) - 2 = 58$ peut être remplacée par $p > 5$.

5.3. Applications

Le th. 5.4 peut être utilisé pour prouver des énoncés où la notion de « G -cr» n'intervient pas explicitement. Par exemple :

PROPOSITION 5.7. — *Soit $\Gamma \subset G(k)$, où G est de type E_8 . Soient V_1, \dots, V_8 les huit représentations irréductibles fondamentales de G . Supposons que l'un des Γ -modules V_i soit semi-simple. Alors tous les autres le sont pourvu que $p > 270$.*

(J'ignore si la minoration $p > 270$ est optimale.)

Soit (α_i) une base du système de racines et soit $\sum c_i \alpha_i^*$ la somme des duales des racines positives. Il résulte de (5.2.2) que l'on a $n(V_i) = c_i$ pour tout i . Dans le cas de E_8 , cela donne :

$$n(V_i) = 92, 136, 182, 270, 220, 168, 114, 58 \text{ pour } i = 1, \dots, 8.$$

D'où le résultat, d'après le th. 5.4.

Voici deux autres applications. La première est l'analogie en caractéristique p du théorème de Chevalley cité au n° 1.1 :

PROPOSITION 5.8 ([Se 94]). — *Soient V_i des représentations linéaires semi-simples d'un groupe Γ . Si $p > \sum(\dim(V_i) - 1)$, la représentation $\otimes V_i$ est Γ -semi-simple.*

On applique le th. 5.4 avec $G = \prod \mathbf{GL}(V_i)$. L'hypothèse que les V_i sont semi-simples signifie que l'image de Γ dans G est G -cr. D'autre part, il résulte de (5.2.5) que l'invariant $n(V)$ de la G -représentation $\otimes V_i$ est égal à $\sum(\dim(V_i) - 1)$. D'où le résultat.

Autre énoncé du même goût:

PROPOSITION 5.9 ([Se 98] et [Mc 00]). — *Si V est une représentation linéaire semi-simple d'un groupe Γ , il en est de même de $\wedge^i V$, pourvu que $p > i(\dim(V) - i)$.*

On applique le th. 5.4 avec $G = \mathbf{GL}(V)$. On peut supposer que $0 \leq i \leq \dim(V)$. On a alors $n(\wedge^i V) = i(\dim(V) - i)$. D'où le résultat.

Remarque.— Dans les deux cas ci-dessus, on peut se proposer de prouver des réciproques. Par exemple, si $\wedge^i V$ est semi-simple, est-il vrai (si $0 < i < \dim(V)$ et si p est assez grand) que V est semi-simple ? Le th. 5.4 dit que «oui» si $p > i(n - i)$ où $n = \dim(V)$. En fait, un argument tannakien élémentaire ([Se 97a]) donne un résultat nettement meilleur : il suffit que p ne divise aucun des entiers $n - 2, n - 3, \dots, n - i$.

BIBLIOGRAPHIE

- [AS 86] M. ASCHBACHER – *Chevalley groups of type G_2 as the group of a trilinear form*, J. Algebra **109** (1986), 193-259.
- [BMR 04] M. BATE, B.M.S. MARTIN et G. RÖHRLE – *A geometric approach to complete reducibility*, preprint, Univ. Birmingham, 2004.
- [BT 65] A. BOREL et J. TITS – *Groupes réductifs*, Publ. Math. IHES **27** (1965), 55-150.
- [BT 71] A. BOREL et J. TITS – *Éléments unipotents et sous-groupes paraboliques de groupes réductifs I*, Invent. math. **12** (1971), 95-104.
- [Br 89] K. BROWN – *Buildings*, Springer-Verlag, 1989.
- [Ch 55] C. CHEVALLEY – *Théorie des Groupes de Lie*, vol. III, Hermann, Paris, 1955.
- [DG 70] M. DEMAZURE et A. GROTHENDIECK – *Structure des schémas en groupes réductifs* (SGA 3 III), Lect. Notes **153**, Springer-Verlag, 1970.
- [Dy 52] E.B. DYNKIN – *Sous-groupes maximaux des groupes classiques* (en russe), Trudy Mosc. Mat. Obsh. **1** (1952), 39-116 ; trad. anglaise : *Selected Papers*, A.M.S., 2000, 37-170.
- [Gu 99] R.M. GURALNICK – *Small representations are completely reducible*, J. Algebra **220** (1999), 531-541.
- [IMP 03] S. ILANGOVAN, V.B. MEHTA et A.J. PARAMESWARAN – *Semistability and semisimplicity in representations of low height in positive characteristic*, in *A Tribute to C.S. Seshadri*, edited by V. Lakshmibai et al, Hindustani Book Ag., New Delhi, 2003.
- [JLPW 95] C. JANSEN, K. LUX, R. PARKER et R. WILSON – *An atlas of Brauer characters*, LMS Monographs, Clarendon Press, Oxford, 1995.

- [Ja 97] J.C. JANTZEN – *Low dimensional representations of reductive groups are semisimple*, in *Algebraic Groups and Lie Groups*, Cambridge U. Press, Cambridge, 1997, 255-266.
- [Ke 78] G.R. KEMPF – *Instability in invariant theory*, *Ann. Math.* **108** (1978), 299-316.
- [LS 96] M.W. LIEBECK et G.M. SEITZ – *Reductive Subgroups of Exceptional Algebraic Groups*, *Memoirs A.M.S.* **580**, 1996.
- [Ma 03a] B.M.S. MARTIN – *Reductive subgroups of reductive groups in nonzero characteristic*, *J. Algebra* **262** (2003), 265-286.
- [Ma 03b] B.M.S. MARTIN – *A normal subgroup of a strongly reductive subgroup is strongly reductive*, *J. Algebra* **265** (2003), 669-674.
- [Mc 98] G.J. MCNINCH – *Dimensional criteria for semisimplicity of representations*, *Proc. London Math. Soc.* **76** (1998), 95-149.
- [Mc 00] G.J. MCNINCH – *Semisimplicity of exterior powers of semisimple representations of groups*, *J. Algebra* **225** (2000), 646-666.
- [Mo 56] G.D. MOSTOW – *Fully reducible subgroups of algebraic groups*, *Amer. J. Math.* **78** (1956), 200-221.
- [Mu 65] D.MUMFORD – *Geometric Invariant Theory*, Springer-Verlag, 1965 ; third enlarged edit. (D. Mumford, J. Fogarty, F. Kirwan), 1994.
- [Mü 97] B. MÜHLHERR – *Complete reducibility in projective spaces and polar spaces*, preprint, Dortmund, 1997.
- [No 87] M.V. NORI – *On subgroups of $\mathbf{GL}(n, \mathbf{F}_p)$* , *Invent. math.* **88** (1987), 257-275.
- [Ri 88] R.W. RICHARDSON – *Conjugacy classes of n -tuples in Lie algebras and algebraic groups*, *Duke Math. J.* **57** (1988), 1-35.
- [Ron 89] M. RONAN – *Lectures on Buildings*, Acad. Press, San Diego, 1989.
- [Rou 78] G. ROUSSEAU – *Immeubles sphériques et théorie des invariants*, *C.R.A.S.* **286** (1978), 247-250.
- [S 00] G.M. SEITZ – *Unipotent elements, tilting modules, and saturation*, *Invent. math.* **141** (2000), 467-502.
- [Se 94] J-P. SERRE – *Sur la semi-simplicité des produits tensoriels de représentations de groupes*, *Invent. math.* **116** (1994), volume dédié à Armand Borel, 513-530.
- [Se 97a] J-P. SERRE – *Semisimplicity and tensor products of group representations: converse theorems* (with an Appendix by Walter Feit), *J. Algebra* **194** (1997), 496-520.
- [Se 97b] J-P. SERRE – *La notion de complète réductibilité dans les immeubles sphériques et les groupes réductifs*, Séminaire au Collège de France, résumé dans [Ti 97, pp. 93-98].

- [Se 98] J-P. SERRE – *The notion of complete reducibility in group theory*, Moursund Lectures Part II, Eugene 1998 (Notes by W.E. Duckworth), <http://darkwing.uoregon.edu/~math/serre/index.html>
- [So 69] L. SOLOMON – *The Steinberg character of a finite group with BN-pair*, Theory of Finite Groups, Benjamin, 1969, 213-221.
- [Te 95] D.M. TESTERMAN – *A_1 -type overgroups of elements of order p in semisimple algebraic groups and the associated finite groups*, J. Algebra **177** (1995), 34-76.
- [Ti 74] J. TITS – *Buildings of spherical type and finite BN-pairs*, LN **386**, Springer-Verlag, 1974.
- [Ti 97] J. TITS – *Résumé des cours de 1996-1997*, Annuaire du Collège de France **97** (1997), 89-102.
- [TW 02] J. TITS et R.M. WEISS – *Moufang Polygons*, Springer-Verlag, 2002.

[Texte révisé en juin 2004]

Jean-Pierre SERRE

Collège de France

11 place M. Berthelot

F-75231 PARIS Cedex 05

E-mail : `serre@dma.ens.fr`

ON THE LONG TIME BEHAVIOR OF KDV TYPE EQUATIONS
[after Martel-Merle]

by **Nickolay TZVETKOV**

1. INTRODUCTION

A central problem in the theory of dispersive PDE's is to understand the interplay between nonlinearity and dispersion. In the context of the water waves problem (see e.g. [1]) the Korteweg-de Vries (KdV) equation

$$(1) \quad u_t + u_{xxx} + \partial_x(u^2) = 0, \quad x \in \mathbb{R}$$

appears to be the simplest (asymptotic) model where both dispersive and nonlinear effects are taken into account. If we neglect the nonlinear interaction $\partial_x(u^2)$ we deal with the Airy equation

$$(2) \quad u_t + u_{xxx} = 0.$$

The solutions of (2) are known to “disperse” in the sense that every solution u of (2) issued from $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ initial data $u(0, \cdot)$, has its L^2 mass conserved but

$$\lim_{t \rightarrow \infty} \|u(t, \cdot)\|_{L^\infty(\mathbb{R})} = 0.$$

If we neglect the dispersive term u_{xxx} , we deal with the Burgers equation which is known to develop singularities in finite time, even for smooth initial data. The KdV equation (1) displays a balance between dispersion and nonlinearity since the dynamics of (1) is well defined, globally in time, for a very large class of initial data and moreover the solutions of (1) enjoy a rich dynamics as $t \rightarrow \infty$. A very special role among the solutions of (1) is played by the so-called solitary wave solution

$$(3) \quad u_c(t, x) = Q_c(x - ct) = \frac{3c}{2} \operatorname{ch}^{-2}\left(\frac{\sqrt{c}}{2}(x - ct)\right), \quad c > 0.$$

The solution (3) *does not disperse* and represents the displacement of the profile Q_c with speed c from left to right as the time t increases. Using the inverse scattering method (see [16, 29, 58]), it turns out that for sufficiently large t , any solution of (1) issued from well localized smooth initial data decomposes as a sum of solitary waves of type (3) plus

a radiation term moving in the opposite direction. A natural generalization of (1), with stronger nonlinear effects, is the equation

$$(4) \quad u_t + u_{xxx} + \partial_x(u^p) = 0,$$

where p is a positive integer. The case $p = 3$ (modified KdV) is a very special case since, as in the case of (1), it can be treated with the inverse scattering method. Unfortunately, the integrability machinery does not seem to apply anymore for the equation (4) when $p \neq 2, 3$. Therefore the qualitative study of (4) in these cases is much less understood.

The equation (4) is a Hamiltonian PDE and its solutions enjoy, at least formally, the conservation laws

$$(5) \quad \|u(t, \cdot)\|_{L^2} = \|u(0, \cdot)\|_{L^2}$$

and

$$(6) \quad \frac{1}{2}\|u_x(t, \cdot)\|_{L^2}^2 - \frac{1}{p+1} \int_{-\infty}^{\infty} u^{p+1}(t, x) dx = \frac{1}{2}\|u_x(0, \cdot)\|_{L^2}^2 - \frac{1}{p+1} \int_{-\infty}^{\infty} u^{p+1}(0, x) dx.$$

Using the Gagliardo-Nirenberg inequalities

$$\|u(t, \cdot)\|_{L^{p+1}(\mathbb{R})}^{p+1} \leq C \|u(t, \cdot)\|_{L^2(\mathbb{R})}^{\frac{p+3}{2}} \|u_x(t, \cdot)\|_{L^2(\mathbb{R})}^{\frac{p-1}{2}},$$

we deduce from (5) and (6) that, for $p < 5$, the H^1 norm of $u(t, \cdot)$ is bounded independently of t as $u(0, \cdot) \in H^1(\mathbb{R})$. Consequently, the H^1 local well-posedness result of Kenig-Ponce-Vega [27] implies the existence of well-defined global dynamics of (4), for $p < 5$, in the energy space $H^1(\mathbb{R})$.

If $p \geq 5$, the H^1 local well-posedness result of Kenig-Ponce-Vega still applies (see Theorem 2.1 below) but the conservation laws (5), (6) provide no longer an H^1 control and hence solutions developing singularities in finite time may appear. The existence of such solutions has been a long standing open problem. In the case $p = 5$, this problem has been solved by Martel-Merle in a series of recent papers. The goal of this exposé is to discuss the main ideas developed by Martel-Merle, together with a presentation of previously known closely related results. One can extract from the results of Martel-Merle the following statement.

THEOREM 1.1 (Martel-Merle [35, 44, 36, 37]). — *Let $p = 5$. There exists $u_0 \in H^1(\mathbb{R})$ such that the local solution of (4) with initial data u_0 blows up in finite time. More precisely there exists $T > 0$ such that $\lim_{t \rightarrow T} \|u(t, \cdot)\|_{H^1(\mathbb{R})} = \infty$.*

We refer to section 8 below for a more precise statement. Let us make a comment on the choice of the initial data u_0 . Equation (4) still has solutions of type (3). Namely, the solitary waves of (4) have the form $u_c(t, x) = Q_c(x - ct)$, $c > 0$ with $Q_c(x) = c^{\frac{1}{p-1}} Q(\sqrt{cx})$ and

$$Q(x) = \left[\frac{p+1}{2 \operatorname{ch}^2\left(\frac{p-1}{2}x\right)} \right]^{\frac{1}{p-1}}.$$

The crux of the Martel-Merle analysis is the deep understanding of the flow of (4) *close to a solitary wave*. It turns out that the solutions developing singularities in finite time constructed by Martel-Merle are issued from initial data close to $Q(x)$ and are essentially of the form $Q_{c(t)}(x + x(t))$ with $c(t) \rightarrow \infty$ as $t \rightarrow \infty$.

The study of solutions of PDE's developing singularities in finite time is an active research field. Let us briefly recall a few of the existing results and compare them with the analysis in the context of (4). In the case of semi-linear wave equations, due to the "finite propagation speed", the blow-up dynamics can be approximated by an ODE developing singularities in finite time (see [2] and the references therein). In the case of quasi-linear wave equations, a Burgers type behavior is behind the blow up dynamics (see [15] and the references therein). The equation (4) does not enjoy similar finite propagation speed properties and the qualitative study of (4) offers new features. Probably the closest models to (4) are the nonlinear Schrödinger equations (NLS). In the case of NLS, we have a functional (viriel functional) giving a simple obstruction for the existence of global dynamics available (see [59] and the references therein). A similar functional is not known to exist in the context of (4). Due to a conformal invariance¹ of some Nonlinear Schrödinger equations, one can construct explicit blow-up solutions (see [41, 43, 60]). Similar invariance is not known in the context of (4).

The rest of this text is organized as follows. In the next section we recall some basic facts on the Cauchy problem for (4). Next, we recall results on the stability of the solitary waves for (4). Starting from section 4, we concentrate on the case $p = 5$. In section 4, we present a characterization of the solitary waves among the solutions with data close to the profile Q . Then, in sections 5 and 6, we present two applications of that characterization result. Section 5 is devoted to an asymptotic stability result while in section 6 we present a result showing the existence of solutions blowing up in finite or infinite time. The last two sections are devoted to the existence of solutions blowing up in finite time. In section 7, we present a result on the blow-up profile which is essential to prove the blow-up in finite time. Section 8 is devoted to the argument providing finite-time blow-up solutions. Finally, in section 9 we present some remarks and open problems.

Acknowledgments. It is a pleasure to thank Anne de Bouard, Khaled El Dika and Jean-Claude Saut for many valuable discussions on the subject. I am also indebted to Anne de Bouard, Laurent Clozel, Yvan Martel and Frank Merle for their remarks on previous versions of this text.

¹The viriel functional is a consequence of that invariance too.

2. THE CAUCHY PROBLEM

In this section, we collect some preliminary results on the Cauchy problem

$$(7) \quad \begin{cases} u_t + u_{xxx} + \partial_x(u^p) = 0, \\ u(0, x) = u_0(x), \end{cases}$$

where $p \geq 1$ is an integer. The following theorem², which can be extracted from the work of Kenig-Ponce-Vega [27], is the starting point for the study of (7) in H^1 .

THEOREM 2.1 ([27]). — *For every $u_0 \in H^1(\mathbb{R})$, there exist $T \in]0, +\infty]$, bounded from below by a positive constant which only depends on $\|u_0\|_{H^1}$, and a functional space X_T continuously embedded in $C([0, T]; H^1(\mathbb{R}))$ such that the Cauchy problem (7) has a unique maximal solution $u \in X_T$. Moreover, if $T < +\infty$ then $\lim_{t \rightarrow T} \|u(t, \cdot)\|_{H^1} = \infty$.*

Of course, a similar statement holds for negative times t . One can also prove the local well-posedness of (7) in H^s for suitable $s < 1$. This fact plays an important role in the Martel-Merle work. For example, it is used to prove that the flow enjoys a continuity property with respect to the weak H^1 topology.

Let us give some indications on the proof of Theorem 2.1 in the case $p = 5$. The proof of the other cases follows similar lines. In the case $p = 5$, one can prove that (7) is well-posed for data in H^s , $s > 0$. The proof is based on applying the contraction mapping principle to the integral formulation (Duhamel principle) of (7)

$$(8) \quad u(t) = S(t)u_0 - \int_0^t S(t - \tau) \partial_x(u^p(\tau)) d\tau.$$

In (8), $S(t) = \exp(-t\partial_x^3)$ is the generator of the free evolution. This is the operator of convolution with respect to x with $(3t)^{-1/3} \text{Ai}(x(3t)^{-1/3})$, where Ai is the Airy function. Let us recall that the Airy function is exponentially decaying on the right and it decays as $|x|^{-1/4}$ on the left (see e.g. [24]). Using the smoothing properties of $S(t)$ one can prove (see [27, Corollary 2.11]) that for $u_0 \in H^1$, the right-hand side of (8) is a contraction in a suitable ball of the space X_T of functions defined on $[0, T] \times \mathbb{R}$, equipped with the norm

$$\|u\|_{X_T} = \|u\|_{L_T^\infty H_x^s} + \|D_x^s u\|_{L_x^5 L_T^{10}} + \|D_t^{s/3} u\|_{L_x^5 L_T^{10}} + \|D_x^s u_x\|_{L_x^\infty L_T^2} + \|D_t^{s/3} u_x\|_{L_x^\infty L_T^2}.$$

The argument relies on some methods from harmonic analysis (restriction phenomena, maximal function estimates, etc.). In the case $s = 0$ the argument breaks down. However, in that case we are able to insure the contraction property, if $\|u_0\|_{L^2}$ is small enough. Therefore, if $p = 5$, the equation (4) is L^2 -critical.

Another very important aspect in the study of (7) is the Kato smoothing effect (see [25]). Let $\varphi \in C^3(\mathbb{R})$ be bounded with all its derivatives. If u is a solution of (4) then,

²We refer to [56, 7, 25, 21] for earlier results on the well-posedness theory of (7).

multiplying (4) with φu and integrating by parts, we obtain the formal³ identity

$$(9) \quad \frac{d}{dt} \int_{-\infty}^{\infty} u^2(t)\varphi = -3 \int_{-\infty}^{\infty} u_x^2(t)\varphi' + \int_{-\infty}^{\infty} u^2(t)\varphi^{(3)} + \frac{2p}{p+1} \int_{-\infty}^{\infty} u^{p+1}(t)\varphi'.$$

Note that, if φ is increasing then the first term in the right-hand side of (9) is negative. This fact was used by Kato [25] to show a remarkable local smoothing effect for (7), if $p < 5$. Namely the solution turns out to be one derivative smoother than the data, locally in space. In [25] well-posedness results in weighted Sobolev spaces are also obtained. The article of Kato was a great source of inspiration for many further works on the subject. It is also the case in the papers by Martel-Merle. For example, the crucial monotonicity properties (see section 5 below) are strongly related to identity (9).

3. STABILITY AND INSTABILITY OF THE SOLITARY WAVES

The initial data giving rise to blow-up solutions in the work of Martel-Merle belong to a small neighborhood of the function $Q(x)$ which is the initial data for a solitary wave. Thus the question of long time stability (or instability) of the solution $Q(x-t)$ of (4) is closely related to Martel-Merle analysis. This question has a long history starting from the pioneering work of Benjamin [4]. The aim of this section is to briefly summarize the state of the art on the stability of $Q(x-t)$. Similar discussion is valid for the solitary wave $Q_c(x-ct)$ (recall that $Q = Q_1$).

Let us first notice that there exist data for (4) arbitrary close to $Q(x)$ such that the corresponding solution does not stay close to $Q(x-t)$ for long times. This is clearly the case of $Q_c(x)$ with c close but different from 1. Indeed, if c is close to 1 then $Q(x)$ is close to $Q_c(x)$, but, because of the different propagation speed, $Q(x-t)$ and $Q_c(x-ct)$ separate from each other for $t \gg 1$.

Notice however that in the previous example the solution issued from Q_c remains close to spatial translates of Q . Hence this example does not exclude orbital stability of Q (up to the action of the group of spatial translations). Indeed, it turns out that for $p < 5$ the solution $Q(x-t)$ is orbitally stable under small H^1 perturbations. Here is the precise statement.

THEOREM 3.1. — *Let $p < 5$. For every $\varepsilon > 0$ there exists $\delta > 0$ such that if the initial data of (7) satisfies $\|u_0 - Q\|_{H^1} < \delta$ then there exists a C^1 function $x(t)$ such that, for every $t \in \mathbb{R}$, the corresponding solution⁴ of (7) satisfies*

$$\|u(t, \cdot + x(t)) - Q(\cdot - t)\|_{H^1} < \varepsilon.$$

³The rigorous justification for sufficiently “nice solutions” u can be obtained by approximation arguments thanks to a propagation of regularity property of the local flow of (7).

⁴The existence is ensured from Theorem 2.1 and the Gagliardo-Nirenberg inequality as shown in the introduction.

The proof of Theorem 3.1 can be found in [8] as an application of the general theory developed in [23]. See also [4, 6, 60, 61, 62] for earlier closely related results. Let us also mention the work of Cazenave-Lions [14] for an apparently different approach based on global variational arguments. In the case $p = 2$, we have L^2 solutions of (4) (see [12]) and it is a natural question whether the space H^1 in Theorem 3.1 could be replaced with L^2 . The answer to that question is positive, as shown in the work by Merle-Vega [50].

Let us give the main idea of the proof of Theorem 3.1. Denote by $E(u)$ and $N(u)$ the functionals

$$(10) \quad E(u) := \frac{1}{2} \|u_x\|_{L^2(\mathbb{R})}^2 - \frac{1}{p+1} \int_{-\infty}^{\infty} u^{p+1}(x) dx, \quad N(u) := \frac{1}{2} \|u\|_{L^2(\mathbb{R})}^2.$$

Recall that if u is a H^1 solution of (4) then the quantities $E(u(t))$ and $N(u(t))$ are time independent. Hence the functional $H(u) := E(u) + N(u)$ defines another conservation law. Since Q solves $(-\partial_x^2 + 1)Q = Q^p$, we infer that Q is a critical point of H . Using the implicit function theorem, we can find a C^1 function $x(t)$, defined a priori at least for small t such that $u(t, x + x(t)) = Q(x) + \mathcal{E}(t, x)$ and $\langle \mathcal{E}(t), Q' \rangle = 0$. Here and in the sequel $\langle \cdot, \cdot \rangle$ stands for the $L^2(\mathbb{R})$ inner product. Then clearly,

$$H(u(0)) = H(u(t)) = H(u(t, \cdot + x(t))) = H(Q + \mathcal{E}(t)),$$

and since Q is a critical point of H , by Taylor expansion, we deduce

$$(11) \quad H(u(0)) - H(Q) = H(Q + \mathcal{E}(t)) - H(Q) = \frac{1}{2} \langle L\mathcal{E}(t), \mathcal{E}(t) \rangle + o(\|\mathcal{E}(t)\|_{H^1}^2),$$

where $L = -\partial_x^2 - pQ^{p-1} + 1$. The next lemma is crucial.

LEMMA 3.2. — Let $p < 5$. Then there exists $C > 0$ such that for every $v \in H^1$ satisfying $\langle v, Q \rangle = \langle v, Q' \rangle = 0$, one has $\langle Lv, v \rangle \geq C\|v\|_{H^1}^2$.

The proof of Lemma 3.2 uses the explicit form of Q . Actually we have a complete understanding of the spectrum of L , namely one simple negative eigenvalue associated to a positive eigenfunction, the simple eigenvalue zero associated to Q_x and all the rest of the spectrum is included in $[\gamma, \infty]$, for some $\gamma > 0$. The relevant fact related to the assumption $p < 5$ is that $d''(1) > 0$ where $d(c) = E(Q_c) + cN(Q_c)$. Unfortunately, the function $\mathcal{E}(t)$ involved in the decomposition of $u(t, x + x(t))$ is not orthogonal to Q and Lemma 3.2 alone does not suffice to complete the proof. However, by writing $\mathcal{E}(t) = \alpha Q + \mathcal{E}_1(t)$ with \mathcal{E}_1 satisfying the assumptions of Lemma 3.2, we can apply the lemma to \mathcal{E}_1 , the L^2 conservation law to evaluate α and to conclude that

$$\langle L\mathcal{E}(t), \mathcal{E}(t) \rangle \geq C\|\mathcal{E}(t)\|_{H^1}^2 + o(\|\mathcal{E}(t)\|_{H^1}^2),$$

provided $\|u\|_{L^2} = \|Q\|_{L^2}$. Hence Q is a local minimizer of the energy E on the sphere of L^2 centered at the origin and of radius $\|Q\|_{L^2}$. Finally using the continuity of E and N on H^1 allows one to complete the stability proof.

It turns out that the restriction $p < 5$ in Theorem 3.1 is sharp.

THEOREM 3.3. — *Let $p \geq 5$. Then the solitary wave $Q(x - t)$ is not stable in the sense of Theorem 3.1.*

The proof of Theorem 3.3 for $p > 5$ can be found in [8] while the more involved analysis in the critical case $p = 5$ is performed in [33]. The argument is based on the construction of a suitable Lyapunov functional. If we set $\psi_c(x) := c^{\frac{1}{p-1}}Q(c^{\frac{2}{p-1}}x)$ then $\|\psi_c\|_{L^2} = \|Q\|_{L^2}$, and for $p > 5$, $\psi_1 = Q$ is a local maximum of $E(\psi_c)$ for c near 1. Thus, in contrast with the case $p < 5$, if $p > 5$ then Q is a saddle point of $E(u)$ subject to the constraint $\|u\|_{L^2} = \|Q\|_{L^2}$. Define a function y as

$$y(x) := \frac{1}{p-1}Q(x) + \frac{2x}{p-1}Q'(x).$$

The relevant fact about y is that $y = \frac{\partial \psi_c}{\partial c}|_{c=1}$. As in the stability proof, we “modulate” the solution of (4) with data close to Q as $u(t, x + x(t)) = Q(x) + \mathcal{E}(t, x)$ with $\langle \mathcal{E}(t), Q' \rangle = 0$. Such a decomposition of the solution is possible as far as it stays in a small H^1 neighborhood of the spatial translates of Q . For $p > 5$, the Lyapunov functional is

$$J(t) = \int_{-\infty}^{\infty} Y(x - x(t))u(t, x)dx,$$

where $Y(x) = \int_{-\infty}^x y(z)dz$. We refer to [23] for the general construction of Lyapunov functionals in the context of solitary waves for PDE's in the presence of symmetry. The discussion of [23] greatly clarifies the nature of the functional $J(t)$.

For $p > 5$, we have that $J'(t) \geq \kappa > 0$, if the initial data is ψ_c with c close but different from one, due to the property of the curve $\{\psi_c, c \sim 1\}$ described above. On the other hand, using some properties of the Airy function, by arguments in the spirit of the Cauchy problem analysis, we can obtain that $J(t) \leq C(t^{-\frac{2}{3}} + t^{\frac{2}{3}})$, $t > 0$, which in view of the lower bound for $J'(t)$ shows that instability holds.

In the critical case $p = 5$, the Lyapunov functional used in [33] is a suitable combination of $J(t)$ and the viriel functional

$$I(t) = \int_{-\infty}^{\infty} (x - x(t))u^2(t, x)dx.$$

The quantities $J(t)$ and $I(t)$ play a central role in Martel-Merle work. They are both measures for the loss of some mass during the time evolution. Notice that a priori these quantities do not make sense for H^1 solutions. The rigorous justification of the existence of $J(t)$ and $I(t)$ is one of the important analytic aspects in the analysis of (7) in a small neighborhood of the solitary waves.

Actually, the critical nature of the exponent $p = 5$ can be “predicted” from the spectral analysis of Pego-Weinstein [53]. Let us linearize (4) around the solution $Q(x - t)$. We set $u(t, x) = Q(x - t) + v(t, x - t)$. If we take $(t, x - t)$ as new variables that we note again with (t, x) , we obtain that v solves the equation

$$v_t - \partial_x L v + R(v) = 0,$$

where $L = -\partial_x^2 - pQ^{p-1} + 1$ and the remainder $R(v)$ contains terms which are at least quadratic in v . Note that L is the same operator as in the expansion of the functional $H(u)$ (see (11) above). As usual, the spectral properties⁵ of $\partial_x L$ are an indicator for the nonlinear stability of $Q(x-t)$ as a solution of (4). In [53], it is shown that $\partial_x L$, considered as an operator on L^2 with domain H^3 , has the following spectrum :

- If $p \leq 5$ then the spectrum coincides with the imaginary axis.
- If $p > 5$ then the spectrum consists in the imaginary axis together with two simple, real eigenvalues, $-\lambda(p) < 0 < \lambda(p)$.

Therefore, if $p \neq 5$, the spectral analysis of [53] agrees with (and further clarifies) the stability theory presented above. The eigenvalue $\lambda(p)$ seems to be responsible for the instability in the case $p > 5$. This assertion *could* become rigorous by combining [53] with the ideas of [20].

Let us now turn to the concept of *asymptotic stability* which is at the heart of the work of Martel-Merle. The result of Theorem 3.1 says that the shape of the solitary wave is stable under small H^1 perturbations. But it is not clear whether $u(t, x + x(t))$ converges, in an appropriate sense, to a limit solitary wave. If it is indeed the case, we say that the family of solitary waves is asymptotically stable. Of course, the choice of the functional setting where one measures the convergence is crucial in that discussion. There are many results on asymptotic stability in the context of dissipative PDE's since in that case one can directly split the dynamics into noninteracting parts due to a spectrum of the linearized operator which does not quite meet the imaginary axis. It seems that the first results on asymptotic stability for Hamiltonian PDE's are those of M. Weinstein and collaborators. In the context of (4), Pego-Weinstein obtained asymptotic stability with convergence in the weighted Sobolev spaces H_a^1 , where $a > 0$ and H_a^1 is equipped with the norm $\|u\|_{H_a^1} = \|e^{ax}u(x)\|_{H^1}$. One can similarly define L_a^2 . The following result is due to Pego-Weinstein.

THEOREM 3.4 ([54]). — *Let $p = 2, 3$ and let a, b be two positive numbers such that $a^3 < 1/3$ and $b < a - a^3$. Then there exist $C > 0$ and $\varepsilon > 0$ such that if $u_0 \in H^2(\mathbb{R})$ satisfies*

$$\|u_0 - Q\|_{H^1} + \|u_0 - Q\|_{H_a^1} \leq \varepsilon$$

then there exist $x(t)$ and c_∞ such that, for all $t \geq 0$, the solution of (7) satisfies

$$\|u(t, \cdot + x(t)) - Q_{c_\infty}\|_{H^1} \leq C\varepsilon$$

and

$$\|u(t, \cdot + x(t)) - Q_{c_\infty}\|_{H_a^1} \leq C\varepsilon e^{-bt}.$$

The approach of Pego-Weinstein also works for some other $p < 5$ which are not integers (see [54] for a precise statement). The proof of Theorem 3.4 uses heavily the spectral

⁵This leads to the notion of *spectral stability*.

analysis of $\partial_x L$ that we already discussed. In the case $p < 5$, the eigenvalue $\lambda(p)$ “becomes” a resonance (resolvent pole) and the corresponding mode plays an important role in the dynamics. If we consider $\partial_x L$ as an operator on L_a^2 then the spectrum shifts from the imaginary axis and the situation becomes very similar to the case of a dissipative PDE. The number b involved in the statement of Theorem 3.4 is essentially the distance between the resonance and the imaginary axis. Let us also mention that the modulation parameter $x(t)$ in Theorem 3.4 is an *affine* function of t .

In [34], Martel-Merle prove an asymptotic stability result where the weighted norm H_a^1 is replaced by a weak H^1 convergence. Here is the precise statement.

THEOREM 3.5 ([34]). — *Consider (7) with $p = 2, 3, 4$. There exists $\varepsilon > 0$ such that if $\|u_0 - Q\|_{H^1} \leq \varepsilon$ then there exist $x(t)$ and c_∞ such that $u(t, \cdot + x(t)) - Q_{c_\infty}$ converges to zero, weakly in H^1 , as $t \rightarrow \infty$.*

Note that in contrast with Theorem 3.4, the result of Theorem 3.5 does not give an estimate for the rate of convergence. On the other hand the assumptions on the data u_0 in Theorem 3.5 are less restrictive than in Theorem 3.4. Notice also that in Theorem 3.5 one can replace the weak H^1 convergence with strong L_{loc}^2 convergence. In the proof of Theorem 3.5, we modulate the solution u as

$$(12) \quad \lambda^{\frac{2}{p-1}}(t)u(t, \lambda(t)x + x(t)) = Q(x) + \mathcal{E}(t, x),$$

where $\lambda(t)$ and $x(t)$ are chosen so that the remainder \mathcal{E} satisfies the orthogonality conditions

$$(13) \quad \langle \mathcal{E}(t), Q \rangle = \langle \mathcal{E}(t), Q' \rangle = 0.$$

We note that, comparing to the stability analysis above, a new modulation parameter⁶ $\lambda(t)$ appears in (12). That parameter is closely related to the *scaling* invariance of (4) which means that if $u(t, x)$ solves (4) then so does $\lambda^{\frac{2}{p-1}}u(\lambda^3 t, \lambda x)$. In that context, the modulation parameter $x(t)$ is related to the translation invariance of (4) which means that if $u(t, x)$ solves (4) then, for every $x_0 \in \mathbb{R}$, the equation (4) is also solved by $u(t, x + x_0)$. The orthogonality conditions (13) are clearly linked with Lemma 3.2 and can be achieved thanks to the implicit function theorem due to the following non degeneracy properties of Q ,

$$(14) \quad \frac{d}{d\lambda} \langle Q_\lambda, Q \rangle \Big|_{\lambda=1} = \frac{5-p}{4(p-1)} \|Q\|_{L^2}^2, \quad \frac{d}{dx_0} \langle Q(\cdot + x_0), Q'(\cdot) \rangle \Big|_{x_0=0} = \|Q'\|_{L^2}^2.$$

Recall from the introduction that Q_λ is the initial data of a solitary wave of (4) propagating with speed λ . In view of (14), in the critical case $p = 5$, we are not able to modulate u so that $\langle \mathcal{E}(t), Q \rangle = 0$ which makes the analysis in that case quite different. To prove Theorem 3.5, one needs to show the convergence of $\lambda(t)$ as $t \rightarrow \infty$ and the convergence of $\mathcal{E}(t)$ to zero in H^1 weak. The proof of these two facts follows lines similar to the analysis of

⁶The introduction of this parameter is in fact the main new point in the instability proof for $p = 5$.

the critical case $p = 5$ to which the next two sections are devoted. In particular it heavily relies on a classification of L^2 compact solutions with data close to Q . It is important to mention that, in contrast with previous works on the subject, in the proof of Theorem 3.5 the bounds on \mathcal{E} rely much less on solving the equation for \mathcal{E} by an iteration scheme in a suitable functional setting.

In a very recent paper of Martel-Merle [39], a different proof of Theorem 3.5 is presented. It is based on the use of a localized viriel type functional which provides a control on \mathcal{E} . Moreover, in [39] a strong H^1 convergence on the right of the solitary wave is obtained. In the proof of Theorem 3.5 it appears that the derivative of the modulation parameter $x(t)$ converges to c_∞ . A natural question is whether, similarly to the result of Pego-Weinstein, in Theorem 3.5 one may take $x(t)$ an affine function of t . The answer to that question is negative, at least in the case $p = 2$, due to an example constructed in [39].

The approach of Theorem 3.5 is further extended in [40, 32] where the asymptotic stability in H^1 of suitable sums of N solitary waves is obtained (see [31] for earlier results). In the case $p = 2$, the result applies to the exact N -soliton solution of the KdV equation (see e.g. [52]). The N -solitons of KdV can be used to show that, for $p = 2$, in Theorem 3.5 strong L^2 convergence is impossible. The existence of such solutions also “explains” the use of weighted norms in the work of Pego-Weinstein.

4. CLASSIFICATION OF L^2 -COMPACT GLOBAL SOLUTIONS WITH DATA CLOSE TO Q

From now on we only consider (4) in the case $p = 5$, i.e. we analyze the Cauchy problem

$$(15) \quad \begin{cases} u_t + u_{xxx} + \partial_x(u^5) = 0, \\ u(0, x) = u_0(x). \end{cases}$$

Let us notice that for $p = 5$, we have $E(Q_c) = 0$ and $\|Q_c\|_{L^2} = \|Q\|_{L^2}$ for all c . Hence in that case the conservation laws (5) and (6) do not present an obstruction for the existence of blow-up solutions of the form $Q_{c(t)}(x + x(t))$ with $c(t) \rightarrow \infty$.

The following rigidity result, called by the authors “Liouville property”, is the main tool in the proof of existence of simple asymptotic objects for the dynamics close to $Q(x - t)$.

THEOREM 4.1 ([35]). — *Suppose that :*

- (1) *The function $u(t, x)$ is a solution global in time of (15) such that*

$$c_1 \leq \|u_x(t)\|_{L^2} \leq c_2,$$

for some positive constants c_1 and c_2 .

- (2) *The solution $u(t, x)$ is L^2 compact which means that there exists a function $y(t)$ such that for every $\varepsilon > 0$ there exists $R > 0$ such that for every t ,*

$$\|u(t, x)\|_{L^2(|x-y(t)|>R)} \leq \varepsilon.$$

Then there exists $\alpha > 0$ such that if $\|u_0 - Q\|_{H^1} \leq \alpha$ then there exist λ and x_0 such that

$$u(t, x) = \lambda^{\frac{1}{2}} Q(\lambda(x - x_0) - \lambda^3 t).$$

REMARK 4.2. — It seems that Theorem 4.1 is the first result of this type for a Hamiltonian PDE. On the other hand, results in the spirit of Theorem 4.1 were previously known in the context of parabolic PDE's (see [51, 19] and the references therein).

Let us give the main lines of the proof of Theorem 4.1. We modulate the solutions $u(t, x)$ of (15) with data close to Q in H^1 as

$$(16) \quad \lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t)) = Q(x) + \mathcal{E}(t, x),$$

where the geometric parameters $\lambda(t) > 0$ and $x(t)$ are defined so that \mathcal{E} satisfies the orthogonality conditions

$$(17) \quad \langle \mathcal{E}(t), Q^3 \rangle = \langle \mathcal{E}(t), Q' \rangle = 0.$$

The reason to chose (17) as orthogonality conditions is that if \mathcal{E} satisfies (17) then $\langle L\mathcal{E}(t), \mathcal{E}(t) \rangle \geq C\|\mathcal{E}(t)\|_{H^1}^2$, where $L = -\partial_x^2 - 5Q^4 + 1$ is the operator arising in the linearization of the energy functional around Q . It seems however that (17) is not the only possible choice of orthogonality conditions on \mathcal{E} which makes the proof of Theorem 4.1 work. The assumption (1) and the conservation laws imply the smallness of $\mathcal{E}(t)$ in H^1 , if u_0 is close to Q in H^1 . If we change the time variable as

$$(18) \quad s = s(t) := \int_0^t \lambda^{-3}(\tau) d\tau$$

then we obtain that $\mathcal{E}(s, x)$ solves the following equation

$$(19) \quad \mathcal{E}_s - (L\mathcal{E})_x - \frac{\lambda_s}{\lambda} \left(\frac{Q}{2} + xQ' \right) - \left(\frac{x_s}{\lambda} - 1 \right) Q' = \frac{\lambda_s}{\lambda} \left(\frac{\mathcal{E}}{2} + x\mathcal{E}_x \right) + \left(\frac{x_s}{\lambda} - 1 \right) \mathcal{E}_x - (10Q^3\mathcal{E}^2 + 10Q^2\mathcal{E}^3 + 5Q\mathcal{E}^4 + \mathcal{E}^5)_x.$$

The new time variable s is clearly related to the scaling of the equation. An advantage of introducing it is that even if t ranges in a finite interval (blow-up regime) the variable s ranges in the whole real line. Indeed, from the H^1 well-posedness and a scaling argument one easily obtains that if a solution $u(t)$ of (15) blows up in finite time T , then

$$\|u_x(t, \cdot)\|_{L^2} \geq C|T - t|^{-\frac{1}{3}},$$

for $t \sim T$. In the context of the decomposition (16), $\lambda(t) \sim \|u_x(t, \cdot)\|_{L^2}^{-1}$ and therefore, in view of (18) the variable s takes all real values. Notice also that the blow-up of $u(t)$ can be simply expressed as $\lambda(t) \rightarrow 0$, while the assumption (1) of Theorem 4.1 can be seen as $\tilde{c}_1 \leq \lambda(t) \leq \tilde{c}_2$.

The parameter $x(t)$ involved in the decomposition (16) is different from $y(t)$ in the assumption (2) of the theorem. But one easily obtains that $|x(t) - y(t)| \leq C$ and that $\mathcal{E}(s, x)$ is L^2 -compact, i.e. for every $\varepsilon > 0$ there exists $R > 0$ such that for every s one has

$\|\mathcal{E}(s, x)\|_{L^2(|x|>R)} \leq \varepsilon$. It turns out that the L^2 -compactness together with the properties of the equation (19) give much stronger estimates on $\mathcal{E}(s, x)$.

LEMMA 4.3. — Let a and b be defined as

$$(20) \quad a := \sup_{s \in \mathbb{R}} \|\mathcal{E}(s, \cdot)\|_{H^1(\mathbb{R})}, \quad b := \sup_{s \in \mathbb{R}} \|\mathcal{E}(s, \cdot)\|_{L^2(\mathbb{R})}.$$

Then there exist $a_0 > 0$, $C > 0$ and $\theta > 0$ such that if $a < a_0$, then $|\mathcal{E}(s, x)| \leq C\sqrt{ab}e^{-\theta|x|}$.

REMARK 4.4. — The exponential decay displayed in Lemma 4.3 is related to a general property of L^2 -compact solutions of (4) and is not restricted only to data close to Q (see [30] for more details).

To prove Lemma 4.3 one needs to observe that equation (19) is essentially a critical generalized KdV equation with exponentially decaying source term. Thus to prove Lemma 4.3, in addition to the L^2 -compactness property one also has to use the L^2 small data scattering theory for (15) developed by Kenig-Ponce-Vega in [27] and the persistence of the decay by the linear KdV flow, a fact already observed in the work of Kato [25].

The next step is to use Lemma 4.3 to get the equivalence between the H^1 and L^2 norms of $\mathcal{E}(s)$.

LEMMA 4.5. — Let a and b be defined as in (20). Then there exist $a_1 > 0$ and $C > 0$ such that if $a < a_1$ then $a \leq Cb$.

The exponential decay obtained in Lemma 4.3 allows us to use viriel type identities. Indeed, if we set

$$I(s) = \frac{1}{2} \int_{-\infty}^{\infty} x \mathcal{E}^2(s, x) dx$$

we can obtain the estimate

$$\frac{d}{ds}(\lambda(s)I(s)) \leq C_1 b^2 - C_2 \|\mathcal{E}(s)\|_{H^1}^2$$

which easily provides the bound $a \leq Cb$.

With Lemma 4.3 and Lemma 4.5 in hand we can turn to the proof of Theorem 4.1. Let us first notice that if $\mathcal{E} = 0$ in (16) and u is a solution of (15) then, in view of the equation solved by \mathcal{E} , we deduce that there exist two constants α and β such that $\lambda(t) = \alpha$ and $x(t) = \alpha^{-2}t + \beta$ which in turn implies that $u(t, x)$ can be represented as claimed in the statement of Theorem 4.1.

The proof of Theorem 4.1 is indirect. Consider a sequence (u_n) of solutions of (15) with $\lim_{n \rightarrow \infty} \|u_n(0) - Q\|_{H^1} = 0$ such that u_n satisfies assumptions (1) and (2) of Theorem 4.1. We can thus represent u_n , for $n \gg 1$ as in (16) with corresponding modulation parameters $\lambda_n(t)$, $x_n(t)$ and remainder \mathcal{E}_n . We suppose that for every n the function \mathcal{E}_n is not identically zero and we look for a contradiction.

It turns out that a suitably renormalized subsequence of \mathcal{E}_n converges to a solution of a linear problem.

LEMMA 4.6. — Let $b_n := \sup_s \|\mathcal{E}_n(s)\|_{L^2}$. Then there exist a sequence (s_n) of real numbers and a subsequence $(\mathcal{E}_{n'})$ such that $b_{n'}^{-1}\mathcal{E}_{n'}(s_{n'}+s)$ converges in $L_{loc}^\infty(\mathbb{R}; L^2(\mathbb{R}))$ to $w(s)$ which is not identically zero and satisfies $w \in C(\mathbb{R}; H^1(\mathbb{R})) \cap L^\infty(\mathbb{R}; H^1(\mathbb{R}))$. Moreover w solves the equation

$$(21) \quad w_s - (Lw)_x = \alpha(s)\left(\frac{Q}{2} + xQ'\right) + \beta(s)Q'$$

for some continuous functions α and β . In addition w satisfies the orthogonality conditions

$$(22) \quad \langle w(s), Q^3 \rangle = \langle w(s), Q' \rangle = 0$$

and the exponential decay

$$(23) \quad |w(s, x)| \leq C e^{-\theta|x|}$$

with a suitable choice of the positive constants C and θ .

At a formal level one easily verifies that the linear equation (21) appears as a limit model for the nonlinear equation (19). We have the estimate

$$(24) \quad \left| \frac{\lambda_s}{\lambda} \right| + \left| \frac{x_s}{\lambda} - 1 \right| \leq C \sup_s \|\mathcal{E}(s, \cdot)\|_{L^2}$$

as a very basic property of the decomposition (16). Indeed, it suffices to multiply (19) with Q^3 and Q' and to integrate on x . Since \mathcal{E}_n solves (19), in view of the bound (24), we obtain that at the limit $b_n \rightarrow 0$ the limit equation for $b_n^{-1}\mathcal{E}_n$ is (21). Indeed, all terms in the right-hand side of (19) disappear either because of (24) or because \mathcal{E} appears in higher powers.

In order to make the previous formal reasoning rigorous one heavily relies on the L^2 small data global existence theory of Kenig-Ponce-Vega [27]. The fact that $\|\mathcal{E}_n\|_{L^2}$ controls $\|\mathcal{E}_n\|_{H^1}$ is of great importance in the limit process.

The last step in the proof of Theorem 4.1 is the analysis of the linear equation (21). Theorem 4.1 follows from the following rigidity property of (21).

LEMMA 4.7. — Let $w \in C(\mathbb{R}; H^1(\mathbb{R})) \cap L^\infty(\mathbb{R}; H^1(\mathbb{R}))$ be a solution of (21) satisfying the orthogonality conditions (22) and the decay estimate (23). Then w is identically zero.

Lemma 4.7 could be seen as a result of unique continuation *at infinity* for the equation (21) (the decay of w is essential for the proof). Let us give an outline of the proof of Lemma 4.7. It turns out that $w(s)$ satisfies an additional orthogonality condition. One can directly show that $\langle w(s), Q \rangle$ is a quantity independent of s . In order to show that it is zero, one appeals to the functional⁷

$$J(s) = \int_{-\infty}^{\infty} w(s, x) \left(\int_0^x \left(\frac{Q(y)}{2} + yQ'(y) \right) dy \right) dx.$$

⁷Recall that a *similar* functional is involved in the instability analysis of the previous section.

A direct computation shows that $J'(s) = 2\langle w(0), Q \rangle$. Due to the exponential decay of $w(s)$, $|J(s)|$ is uniformly bounded and therefore $\langle w(s), Q \rangle = 0$. Thus, for all s , $w(s)$ is orthogonal to Q .

We next consider the viriel functional

$$I(s) = \frac{1}{2} \int_{-\infty}^{\infty} x w^2(s, x) dx.$$

A direct computation shows that

$$(25) \quad I'(s) = H(w(s), w(s)) + \alpha(s) \langle x \left(\frac{Q}{2} + xQ' \right), w(s) \rangle + \beta(s) \langle xQ', w(s) \rangle,$$

where $H(w, w) = \langle (Lw)_x, xw \rangle$. In view of (25), we slightly modify $w(s)$ by setting

$$(26) \quad \tilde{w}(s) = w(s) + \gamma(s) \left(\frac{Q}{2} + xQ' \right) + \delta(s)Q'.$$

It turns out that with a suitable choice of $\gamma(s)$ and $\delta(s)$, \tilde{w} solves an equation of type

$$\tilde{w}_s - (L\tilde{w})_x = \tilde{\alpha}(s) \left(\frac{Q}{2} + xQ' \right) + \tilde{\beta}(s)Q',$$

satisfies the orthogonality conditions

$$(27) \quad \langle \tilde{w}(s), \left(\frac{xQ}{2} + x^2Q' \right) \rangle = \langle \tilde{w}(s), xQ' \rangle = \langle \tilde{w}(s), Q \rangle = 0,$$

and, if we set $I_1(s) = \int_{-\infty}^{\infty} x \tilde{w}^2(s, x) dx$, then $I_1'(s) = H(\tilde{w}(s), \tilde{w}(s))$.

The next step is to use that due to the orthogonality conditions satisfied by \tilde{w} , one has

$$(28) \quad -H(\tilde{w}(s), \tilde{w}(s)) \geq \frac{1}{10} \langle L\tilde{w}(s), \tilde{w}(s) \rangle.$$

Once again, in order to prove (28), one needs to make explicit calculations based on the very particular form of Q . Using the equation solved by $\tilde{w}(s)$, one can easily check that $\langle L\tilde{w}(s), \tilde{w}(s) \rangle = \langle L\tilde{w}(0), \tilde{w}(0) \rangle$ and therefore $I_1'(s) \leq -\frac{1}{10} \langle L\tilde{w}(0), \tilde{w}(0) \rangle \leq 0$. Since $\tilde{w}(s)$ is exponentially decaying we obtain that $\langle L\tilde{w}(s), \tilde{w}(s) \rangle = 0$.

But as a matter of fact, if $u \in H^1(\mathbb{R})$ is orthogonal to Q and satisfies $\langle Lu, u \rangle = 0$, then u is necessarily a linear combination of Q' and $\frac{Q}{2} + xQ'$. In view of (27), we directly conclude that \tilde{w} is identically zero. Finally, we obtain that w is identically zero thanks to (26) and the orthogonality conditions (22).

5. ASYMPTOTIC STABILITY IN THE REGULAR REGIME

A first consequence of Theorem 4.1 is an asymptotic stability result under the assumption that the solution with data close to Q is globally defined and uniformly bounded in $H^1(\mathbb{R})$.

THEOREM 5.1 ([35]). — *Suppose that $u(t)$ is a solution global in time of (15) such that $c_1 \leq \|u_x(t)\|_{L^2} \leq c_2$, for some positive constants c_1 and c_2 . Then there exists $\alpha > 0$ such that if $\|u_0 - Q\|_{H^1} \leq \alpha$ then there exist $\lambda(t)$ and $x(t)$ such that $\lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t))$ converges to $Q(x)$, weakly in $H^1(\mathbb{R})$, as $t \rightarrow \infty$.*

Notice that the result of Theorem 5.1 displays a weaker form of asymptotic stability compared with the case $p < 5$. Indeed, in contrast with the situation for $p < 5$, it is not clear whether the modulation parameter $\lambda(t)$, involved in the statement of Theorem 5.1, converges to some limit as $t \rightarrow \infty$.

Let us give the main ideas of the proof of Theorem 5.1. The two main ingredients are :

- Continuity of the flow of (15) with respect to the weak H^1 topology.
- An almost monotonicity property of the L^2 mass for solutions of (15) with data close to Q in H^1 .

The continuity property of the flow with respect to the weak H^1 topology is a consequence of the well-posedness of the Cauchy problem below H^1 and of a viriel identity argument. This type of results seems to appear first in a paper by Glangetas-Merle [22]. The monotonicity property of the mass is closely related to the Kato identity discussed in section 2. This monotonicity property separates the dynamics into two noninteracting parts and it is related to the dispersion relation (the symbol) of the *linear part* of the equation. It is worth noticing that similar monotonicity properties hold for a fairly large class of equations such as the Benjamin-Bona-Mahony equation (see [17]), the Kadomtsev-Petviashvili II equation (see [57, 10]) but, it does not seem to hold for other models as the Kadomtsev-Petviashvili I equation or the nonlinear Schrödinger equation. We refer to [54, page 308] for a very clear explanation whether an equation in hand may enjoy the crucial separation of the dynamics property.

The proof of Theorem 5.1 is again indirect. We modulate $u(t, x)$ as in (16)

$$(29) \quad \lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t)) = Q(x) + \mathcal{E}(t, x).$$

Let us take a sequence $t_n \rightarrow \infty$ such that $\lambda^{\frac{1}{2}}(t_n)u(t_n, \lambda(t_n)x + x(t_n))$ converges, weakly in H^1 , to \tilde{u}_0 and $\lambda(t_n)$ converges to $\tilde{\lambda}_0$. We suppose $\tilde{u}_0 \neq Q$ and we seek for a contradiction by means of Theorem 4.1. Let \tilde{u} be the local solution of the critical generalized KdV equation with data \tilde{u}_0 . Since \tilde{u}_0 is close to Q in H^1 , we can modulate \tilde{u} , at least for small times, with modulation parameters $\tilde{\lambda}(t)$, $\tilde{x}(t)$ and remainder $\tilde{\mathcal{E}}(t, x)$ satisfying the same orthogonality conditions as \mathcal{E} . The continuity property of the flow with respect to the weak H^1 topology implies the following statement.

LEMMA 5.2. — The solution $\tilde{u}(t)$ is defined for all $t \in \mathbb{R}$, and the sequence $\mathcal{E}(t_n + t)$ converges, weakly in $H^1(\mathbb{R})$, to $\tilde{\mathcal{E}}(t)$. Moreover, for every $T > 0$, $\lambda(t_n + t) - \tilde{\lambda}(t)$ and $x(t_n + t) - \tilde{x}(t)$ converge to zero in $C([-T, T]; \mathbb{R})$.

Notice that one first proves the lemma for small times which implies a H^1 bound on $\tilde{u}(t)$ thanks to the H^1 boundedness assumption on u . Then we extend $\tilde{u}(t)$ for all times due to the H^1 local well-posedness of (15).

The asymptotic solution \tilde{u} being constructed, the aim is to show that it satisfies the assumptions of Theorem 4.1, i.e. we need to check that \tilde{u} is L^2 -compact. Let us notice that the leading idea at this point is that the solution \tilde{u} enjoys more properties than the original solution u itself.

5.1. L^2 -compactness of \tilde{u} on the right of the solitary wave

There are different ways to prove the L^2 -compactness of \tilde{u} on the right. In this subsection, we discuss an argument which is based on a direct analysis of the limit solution \tilde{u} . In the next subsection, we present a more involved method, using Lemma 5.2, providing compactness on both sides.

To get the L^2 -compactness of \tilde{u} on the right, it is sufficient to show that small solutions of (15) can not travel too fast to the right. Such a result would imply the needed compactness since the main part of the solution is moving to the right with speed uniformly bounded from below. Let us state precisely a lemma giving the L^2 -compactness on the right. We introduce a function $\psi \in C^\infty(\mathbb{R})$ by setting

$$\psi(x) = c_0 \int_{-\infty}^x Q\left(\frac{x}{K}\right) dy,$$

where c_0 is chosen so that $\lim_{x \rightarrow \infty} \psi(x) = 1$. Consider the functional

$$\mathcal{I}_\sigma(t) := \int_{-\infty}^{\infty} \psi(x - \sigma t) u^2(t, x) dx$$

which measures the distribution of the L^2 -mass on the right with respect to a frame moving with speed σ . Since the quantity on the right of the solitary wave is essentially a “small solution” of (15), the next statement is the crucial point in the proof of the L^2 -compactness of \tilde{u} on the right.

LEMMA 5.3. — Let $\sigma > 0$ and $K > \sqrt{2/\sigma}$. There exists a positive constant C_σ such that, if $\|u_0\|_{L^2} \leq C_\sigma$, then $\mathcal{I}_\sigma(t)$ is a non-increasing function on the trajectories of (15).

The proof of Lemma 5.3 is an application of the Kato identity (9).

5.2. L^2 -compactness of \tilde{u} on the left of the solitary wave

The analysis in that case is more delicate. The main point is to show that the loss of mass at the left is “irreversible”. Notice that Lemma 5.3 does not hold for large solutions since there are solitary waves moving with arbitrary large speed. It turns out however

that a weaker form of Lemma 5.3 survives for large data. Let $u(t)$ be a solution of (15) which is decomposed as in (16). For $(t_0, x_0) \in \mathbb{R}^2$ and $t_0 \geq t$, we introduce the functional

$$\mathcal{I}_{x_0,t_0}(t) := \int_{-\infty}^{\infty} \psi(x - x(t) - x_0 - \frac{3}{4}(x(t_0) - x(t))) u^2(t, x) dx,$$

where the function ψ is defined in the previous subsection. The following statement is now the substitute of Lemma 5.3.

LEMMA 5.4. — Suppose that there exist two positive numbers c_1 and c_2 such that $c_1 \leq \lambda(t) \leq c_2$. Then there exist $\delta > 0$, $K > 0$ and $C > 0$ such that, if $\sup_t \|\mathcal{E}(t)\|_{H^1} < \delta$, then for every $x_0 \geq 0$, and $0 \leq t \leq t_0$,

$$I_{x_0,t_0}(t_0) - I_{x_0,t_0}(t) \leq Ce^{-\frac{x_0}{K}}.$$

REMARK 5.5. — The very particular structure of the functional $\mathcal{I}_{x_0,t_0}(t)$ is important for the proof. The term $\frac{3}{4}(x(t) - x(t_0))$ is strongly needed. The number $\frac{3}{4}$ can be replaced by any number between $\frac{1}{2}$ and 1.

We now explain how the L^2 -compactness of \tilde{u} can be obtained by combining Lemma 5.4 and Lemma 5.2. Define the functional

$$m_r(u(t)) := \int_{-\infty}^{\infty} \psi(x - x(t) - x_0) u^2(t, x) dx$$

which measures the L^2 -mass on the right of the solitary wave. It is easy to see that Lemma 5.4 implies that $m_r(t)$ is an “almost decreasing quantity”. More precisely,

$$(30) \quad m_r(u(t)) - m_r(u(t')) \leq Ce^{-\frac{x_0}{K}}, \quad t \geq t'.$$

Indeed, we have that $m_r(u(t)) = \mathcal{I}_{x_0,t}(t)$ and due to the monotonicity of ψ and $x(t)$, $m_r(u(t')) \geq \mathcal{I}_{x_0,t}(t')$. Similarly, since if $u(t, x)$ solves (15) then so does $u(-t, -x)$, we deduce that the quantity

$$m_l(u(t)) := \int_{-\infty}^{\infty} (1 - \psi(x - (x(t) - x_0))) u^2(t, x) dx,$$

measuring the L^2 -mass on the left of the solitary wave is “almost increasing”. More precisely,

$$(31) \quad m_l(u(t)) - m_l(u(t')) \geq -Ce^{-\frac{x_0}{K}}, \quad t \geq t'.$$

Assume that \tilde{u} is not L^2 -compact. It means that there exists $\delta > 0$ such that for every bounded interval $I \subset \mathbb{R}$, there exists t_0 such that

$$\int_I \tilde{u}^2(t_0, x) dx \leq \|\tilde{u}(0)\|_{L^2}^2 - \delta.$$

For a solution $u(t)$ of (15), we define the quantity $m_{loc}(u(t))$, measuring the L^2 -mass in a moving frame, via the identity

$$(32) \quad \|u(t)\|_{L^2}^2 = m_l(u(t)) + m_{loc}(u(t)) + m_r(u(t)).$$

Taking $x_0 \gg 1$, we can assume

$$(33) \quad m_{loc}(\tilde{u}(0)) > \|\tilde{u}(0)\|_{L^2}^2 - \frac{\delta}{4}.$$

The assumption of lack of L^2 -compactness of \tilde{u} implies the existence of $t_0 \in \mathbb{R}$ such that

$$(34) \quad m_{loc}(\tilde{u}(t_0)) < \|\tilde{u}(0)\|_{L^2}^2 - \frac{\delta}{2},$$

if $x_0 \gg 1$. We suppose that $t_0 > 0$, the case $t_0 < 0$ being similar. From (33) and (34), we get the estimate

$$m_{loc}(\tilde{u}(0)) - m_{loc}(\tilde{u}(t_0)) > \frac{\delta}{4}.$$

Since the weak H^1 -convergence implies the strong L^2_{loc} -convergence, using Lemma 5.2, we deduce that, if $x_0 \gg 1$ (independently of t_0) then there exists N such that for $n \geq N$,

$$(35) \quad m_{loc}(u(t_n)) - m_{loc}(u(t_n + t_0)) > \frac{\delta}{8}.$$

The L^2 -conservation law for u can now be written as

$$m_l(u(t_n + t_0)) - m_l(u(t_n)) = m_r(u(t_n)) - m_r(u(t_n + t_0)) + m_{loc}(u(t_n)) - m_{loc}(u(t_n + t_0))$$

which, using (30) and (35), yields

$$(36) \quad m_l(u(t_n + t_0)) - m_l(u(t_n)) > \frac{\delta}{8} - Ce^{-\frac{x_0}{K}}.$$

We can clearly suppose that $t_{n+1} > t_n + t_0$. Therefore using (31) and (36), we get

$$\begin{aligned} m_l(u(t_{n+1})) - m_l(u(t_n)) &= m_l(u(t_{n+1})) - m_l(u(t_n + t_0)) + m_l(u(t_n + t_0)) - m_l(u(t_n)) \\ &> \frac{\delta}{8} - 2Ce^{-\frac{x_0}{K}}. \end{aligned}$$

Hence, if $x_0 \gg 1$, it follows that $m_l(u(t_n)) \rightarrow \infty$ which is clearly impossible in view of the L^2 -conservation law. We obtain a contradiction coming from the assumption of lack of L^2 -compactness of \tilde{u} . Therefore \tilde{u} is L^2 -compact.

REMARK 5.6. — Notice that Theorem 5.1 gives an important information about negative energy solutions with data close to Q . Namely, such solutions cannot be uniformly bounded in $H^1(\mathbb{R})$. Indeed, let u be a negative energy solution of (15) with data close to Q which is uniformly bounded in $H^1(\mathbb{R})$. Then similarly to above, we can construct an asymptotic solution \tilde{u} which satisfies the assumptions of Theorem 4.1. Thus \tilde{u} is necessarily a rescaled and translated version of Q . In particular, $E(\tilde{u}) = 0$ since in the critical case $p = 5$, $E(Q_c) = 0$ for all c . On the other hand \tilde{u} is obtained as a weak H^1 limit from solutions close to $Q(x - t)$ with negative energies which implies that $E(\tilde{u}) < 0$. We thus get a contradiction with the assumption of uniform H^1 -boundedness of u .

Let us finally remark that since $E(Q) = 0$ and $\nabla E(Q) = -Q$, we obtain that there exists a large set of negative energy initial data for (15) which is close to Q . An example of such data is clearly $u_0(x) = \pm(1 + \varepsilon)Q$, where $0 < \varepsilon \ll 1$.

6. BLOW-UP IN FINITE OR INFINITE TIME

In this section, we present a second consequence of Theorem 4.1 which is the existence of solutions of (15) blowing up in finite or infinite time.

THEOREM 6.1 ([44]). — *There exists $\alpha > 0$ such that if $u_0 \in H^1(\mathbb{R})$ satisfies*

$$(37) \quad \|u_0\|_{L^2} \leq \|Q\|_{L^2} + \alpha$$

and $E(u_0) < 0$ (negative energy) then the solution of (15) blows up in finite or infinite time which means that there exists $T \in]0, \infty]$ such that $\lim_{t \rightarrow T} \|u(t, \cdot)\|_{H^1} = \infty$.

REMARK 6.2. — Under the assumptions of Theorem 6.1, the initial data $u_0(x)$ is close in $H^1(\mathbb{R})$ to $\pm \lambda^{\frac{1}{2}} Q(\lambda(x + x_0))$ for some constants λ_0 and x_0 (see [41, 42, 60]).

Notice that if u_0 satisfies the assumptions of Theorem 6.1, then so does $-u_0$. We also remark that the new point in Theorem 6.1 with respect to Theorem 5.1 is that we have the existence of the limit as t goes to T of $\|u(t, \cdot)\|_{H^1}$ and not only the existence of a sequence (t_n) such that $\|u(t_n, \cdot)\|_{H^1}$ goes to infinity.

The approach of Theorem 6.1 is similar to that of Theorem 5.1. The new ingredients are :

- Extension of the L^2 -compactness of the limit solution to the case when there is no lower bound on the scaling modulation parameter $\lambda(t)$.
- The use of a third conservation law of (15) which provides a control on the size of $\lambda(t)$.

The starting point in Theorem 6.1 is to modulate a negative energy solution of (15), with data satisfying (37) for α small, as

$$(38) \quad \lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t)) = \pm Q(x) + \mathcal{E}(t, x)$$

with $\mathcal{E}(t, x)$ satisfying the orthogonality conditions (17). Without loss of generality, we may assume that the sign in front Q is plus. The decomposition (38) is the same as in (16) but the proof of the control on the modulation parameters $\lambda(t)$, $x(t)$ and the remainder $\mathcal{E}(t, x)$ is different under the assumptions of Theorem 6.1. Using the variational nature of Q one can show that the decomposition (38) holds with smallness estimates⁸ on $\mathcal{E}(t, x)$ and bounds on $\lambda(t)$ and $x(t)$, as far as the solution u exists. Here, “the variational nature of Q ” means that if $u \in H^1(\mathbb{R})$ is such that $E(u) = 0$, $\|u\|_{L^2} = \|Q\|_{L^2}$ and $\|u'\|_{L^2} = \|Q'\|_{L^2}$ then $u(x) = \pm Q(x + x_0)$ for some constant $x_0 \in \mathbb{R}$ (see [41, 42, 60]).

The proof is by contradiction. Take a sequence (u_n) of negative energy global solutions of (15) such that $\|u_n(0)\|_{L^2}$ tends to $\|Q\|_{L^2}$ as $n \rightarrow \infty$. We suppose that for each n there exist a sequence $(t_{n,m})$ and a constant c_n so that $\|\partial_x u_n(t_{n,m})\|_{L^2} \leq c_n$, uniformly in m . We seek for a contradiction under this assumption. Similarly to the previous section, we

⁸Depending on the smallness of $(\|u_0\|_{L^2} - \|Q\|_{L^2})$.

define a limit object $\widetilde{u}_n(0)$ from the sequence $(t_{n,m})$ and the decomposition (38) applied to u_n . We denote by $\widetilde{u}_n(t)$ the local solution of (15) with initial data $\widetilde{u}_n(0)$, defined on a time interval $(-T_1(n), T_2(n))$. The closeness to $Q(x-t)$ and the weak H^1 -convergence yield

$$(39) \quad E(\widetilde{u}_n) < 0.$$

Similarly to the considerations on the L^2 -compactness of \widetilde{u} in the previous section, one can show that \widetilde{u}_n is L^2 -compact and satisfies a crucial exponential decay property. This allows one to use a third conservation law of (15) applied to \widetilde{u}_n . Namely,

$$(40) \quad \int_{-\infty}^{\infty} \widetilde{u}_n(t, x) dx = \int_{-\infty}^{\infty} \widetilde{u}_n(0, x) dx,$$

if $t \in (-T_1(n), T_2(n))$. The conservation law (40) shows that the scaling modulation parameter $\widetilde{\lambda}_n(t)$, involved in the decomposition of $\widetilde{u}_n(t)$, is uniformly bounded from below. Using that $\widetilde{\lambda}_n(t) \sim \|\partial_x \widetilde{u}_n(t, \cdot)\|_{L^2}^{-1}$, we obtain a uniform bound on $\|\widetilde{u}_n(t, \cdot)\|_{H^1}$ which, thanks to the H^1 well-posedness of (15) implies that $\widetilde{u}_n(t)$ is globally defined. Since \widetilde{u}_n is L^2 -compact, using Theorem 4.1, we conclude that \widetilde{u}_n is a rescaled and translated Q . In particular, $E(\widetilde{u}_n) = 0$ which is in contradiction with (39).

7. BLOW-UP PROFILE

In this section, we present an extension of the asymptotic stability result of Theorem 5.1 to the singular regime of Theorem 6.1. It turns out that the blow-up solutions with data close to Q converge in H^1 weak, after a suitable *singular renormalization* to the profile Q . Thus the concept of asymptotic stability naturally extends to the singular regime as shows the next statement.

THEOREM 7.1 ([36]). — *There exists $\alpha > 0$ such that if $\|u_0\|_{L^2} \leq \|Q\|_{L^2} + \alpha$ and if the solution $u(t)$ of (15) blows up in finite or infinite time $T \in]0, \infty]$ then there exist $\lambda(t) > 0$ and $x(t)$ such that $\lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t))$ converges as $t \rightarrow T$, weakly in $H^1(\mathbb{R})$, either to $Q(x)$ or to $-Q(x)$.*

Notice that, if $E(u_0) < 0$ then the result of Theorem 7.1 applies to the blow-up solutions considered in the previous section. It is also worth noticing that the energy conservation and the weak H^1 convergence imply the strong convergence in the homogeneous Sobolev space $\dot{H}^1(\mathbb{R})$.

The proof of Theorem 7.1 follows a similar strategy to that of Theorem 5.1. However, the classification result of Theorem 4.1 alone is not sufficient to conclude because the asymptotic solution may be singular ($\lambda(t) \rightarrow 0$). For that purpose a new rigidity result adapted to the singular regime has to be established. Again, the viriel functional $I(t)$ and

the functional $J(t)$, appeared already several times in our discussion, are the key for the argument.

In the proof of Theorem 7.1, we write once again

$$(41) \quad \lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t)) = \pm Q(x) + \mathcal{E}(t, x),$$

for t near the blow-up time T and α small enough. The difference is that now the modulation parameters $\lambda(t)$ and $x(t)$ are chosen so that \mathcal{E} satisfies the orthogonality conditions

$$(42) \quad \langle \mathcal{E}(t), (\frac{xQ}{2} + x^2Q') \rangle = \langle \mathcal{E}(t), xQ' \rangle = 0.$$

Notice that we already considered orthogonality conditions of type (42) in the linear analysis of section 4 (see (27)). The orthogonality conditions (42) are used to cancel some second order terms in variation of the viriel functional $I(t) = \frac{1}{2} \int_{-\infty}^{\infty} x \mathcal{E}^2(t, x) dx$. More precisely, at least formally,

$$\begin{aligned} \lambda^3(t)\dot{I}(t) &= \lambda^2(t)\dot{\lambda}(t) \langle \mathcal{E}(t), (\frac{xQ}{2} + x^2Q') \rangle + (\lambda^2(t)\dot{x}(t) - 1) \langle \mathcal{E}(t), xQ' \rangle \\ &\quad + H(\mathcal{E}(t), \mathcal{E}(t)) + R(\mathcal{E}(t)), \end{aligned}$$

where $R(\mathcal{E}(t))$ contains only higher order terms in $\mathcal{E}(t)$ and $H(\mathcal{E}, \mathcal{E}) = \langle (L\mathcal{E})_x, x\mathcal{E} \rangle$ is the bilinear form which already appeared in the proof of Lemma 4.7. The choice of (42) as orthogonality conditions is possible thanks to the implicit function theorem, in view of another non degeneracy property of Q . Despite the “loss of sign” of $\langle L\mathcal{E}, \mathcal{E} \rangle$ with the new orthogonality conditions, one is still able to get smallness bounds on \mathcal{E} and a control on the modulation parameters λ and x of type (24). The variational nature of Q is again used in the smallness estimates on \mathcal{E} .

In order to prove Theorem 7.1, one has to show that $\mathcal{E}(t)$ converges to zero in H^1 weak, as $t \rightarrow T$. One first proves that $\mathcal{E}(t_n)$ converges to zero in H^1 weak, as $n \rightarrow \infty$, for a specific choice of the sequence (t_n) . Namely, t_n is so that $\lambda(t_n) = (1.1)^{-n}$ and $\lambda(t) \leq \lambda(t_n)$ for $t \in [t_n, T[$. The case of an arbitrary sequence (t_n) then can be treated by using the monotonicity of the L^2 -mass.

Let us describe the argument for the specific sequence (t_n) . The proof is by contradiction. We suppose that there exists a subsequence of (t_n) still denoted by (t_n) such that $\mathcal{E}(t_n)$ converges weakly in H^1 to $\tilde{\mathcal{E}}(0)$ which is not zero and we look for a contradiction. Let $\tilde{u}(0) := \pm Q + \tilde{\mathcal{E}}(0)$. We denote by $\tilde{u}(t)$ the local solution of (15) subject to initial data $\tilde{u}(0)$. Let $\tilde{\lambda}(t)$, $\tilde{x}(t)$ and $\tilde{\mathcal{E}}(t)$ be the modulation parameters and the remainder in a decomposition of type (41) applied to $\tilde{u}(t)$. The solution $\tilde{u}(t)$ may develop singularities in finite time and this is the new feature in the analysis.

Notice that $\tilde{\lambda}(0) = 1$. Thanks to the special choice of (t_n) one has $\tilde{\lambda}(t) \leq 1$ and we can define a maximal $\tau \in]0, \infty]$ such that $(1.1)^{-1} < \tilde{\lambda}(t) \leq 1$ for every $t \in [0, \tau)$. Two possibilities appear, either $\tau = \infty$ or $\tau < \infty$. In the case $\tau = \infty$, the solution \tilde{u} is global, uniformly bounded in H^1 , and one can show similarly to before that \tilde{u} is L^2 -compact.

Thus Theorem 4.1 applies and gives a contradiction as in Theorem 5.1. In the case $\tau < \infty$ a fairly new argument is needed. Introduce the new time variable s as in (18) with $\tilde{\lambda}(t)$ instead of $\lambda(t)$. Set $\tau_1 := s(\tau)$. The contradiction arises from a lower and upper bound on the quantity Λ , defined as

$$\Lambda := \int_0^{\tau_1} \int_{-\infty}^{\infty} \tilde{\mathcal{E}}^2(s, x) e^{-\frac{|x|}{2}} dx ds .$$

One can prove an exponential decay of $\tilde{\mathcal{E}}$ to the left by the monotonicity properties considered in section 5. This allows to consider the functional

$$(43) \quad J(s) = \int_{-\infty}^{\infty} \tilde{\mathcal{E}}(s, x) \left(\int_x^{\infty} \left(\frac{Q(y)}{2} + yQ'(y) \right) dy \right) dx - \frac{1}{4} \|Q\|_{L^2}^2 .$$

The second term in (43) is of course not essential. Using direct computations and the basic properties of the decomposition (41), one can show the bound

$$(44) \quad \left| J'(s) + \frac{\lambda_s}{2\lambda} J(s) + 2\langle \tilde{\mathcal{E}}(s), Q \rangle \right| \leq C \int_{-\infty}^{\infty} \tilde{\mathcal{E}}^2(s, x) e^{-\frac{|x|}{2}} dx .$$

Estimate (44) is now the key for the proof of the lower bound

$$(45) \quad \Lambda \geq C\Lambda_1 ,$$

where $C > 0$ is independent of α ,

$$\Lambda_1 := 1 + \int_0^{\tau_1} \int_{-\infty}^{\infty} \tilde{\mathcal{E}}_x^2(s, x) dx ds + |E| \int_0^{\tau_1} (\tilde{\lambda}(s))^2 ds$$

and E is the energy of \tilde{u} .

For the upper bound on Λ , a localized⁹ viriel type functional is used. For $A > 0$, we consider the function $\psi_A(x) = A\psi(A^{-1}x)$, where $\psi(x)$ is a smooth odd function such that $\psi(x) = x$ for $|x| < 1$, $\psi'(x) = e^{-x}$ for $x > 2$, and for $x \in [1, 2]$, ψ is increasing and concave. We consider the following localized viriel functional

$$I_A(s) := \int_{-\infty}^{\infty} \psi_A(x) \tilde{\mathcal{E}}^2(s, x) dx .$$

Thanks to the new orthogonality conditions one obtains that there exist $A > 2$, $\gamma > 0$ such that

$$(46) \quad I'_A(s) \leq -\gamma \int_{-\infty}^{\infty} e^{-\frac{|x|}{A}} (\tilde{\mathcal{E}}^2(s, x) + \tilde{\mathcal{E}}_x^2(s, x)) dx + \frac{1}{\gamma} \langle \tilde{\mathcal{E}}(s), Q \rangle^2 ,$$

provided u_0 is close enough to Q . The bilinear form H is naturally involved in the proof of (46). More precisely, it turns out that

$$-H(\mathcal{E}(s), \mathcal{E}(s)) \geq C \|\mathcal{E}(s)\|_{H^1}^2 ,$$

if $\langle \mathcal{E}(s), Q \rangle = \langle \mathcal{E}(s), (\frac{xQ}{2} + x^2Q') \rangle = 0$. We have the second term in the right-hand side in (46) because the orthogonality with respect to Q is “forbidden” for $\tilde{\mathcal{E}}$ (see (14)).

⁹This localization is needed because we do not have available a decay of $\tilde{\mathcal{E}}(s, x)$ for $x \rightarrow +\infty$.

Using (46) one can get the upper bound

$$(47) \quad \Lambda \leq C_{u_0} \Lambda_1$$

where the constant C_{u_0} is tending to zero, if $\|u_0\|_{L^2}$ is tending to $\|Q\|_{L^2}$. In view of (45) and (47), we get a contradiction for $\|u_0\|_{L^2}$ close enough to $\|Q\|_{L^2}$.

REMARK 7.2. — A corollary of Theorem 7.1 is a lower bound on the blow-up rate which excludes the existence of self-similar blow-up solutions for data in H^1 . However, in [9], Bona-Weissler construct solutions of (7) with self-similar blow-up which are missing the space H^1 .

8. BLOW-UP IN FINITE TIME

In this section, we present a result showing that under an additional assumption on the initial data, the blow-up solutions of Theorem 6.1 develop their singularities in *finite time*.

THEOREM 8.1 ([37]). — *There exists $\alpha > 0$ such that, if $u_0 \in H^1(\mathbb{R})$ satisfies*

$$\|u_0\|_{L^2} \leq \|Q\|_{L^2} + \alpha, \quad E(u_0) < 0, \quad \forall x_0 > 0, \quad \|u_0(x)\|_{L^2(x \geq x_0)} \leq C|x_0|^{-3},$$

then the solution of (15) blows up in finite time, i.e. there exists $T \in]0, \infty[$ such that $\lim_{t \rightarrow T} \|u(t, \cdot)\|_{H^1} = \infty$.

A detailed presentation of all steps in the proof of Theorem 8.1, as written is [37], would be quite technical. For that reason, we will only give an informal presentation of the main idea.

Consider the usual decomposition

$$(48) \quad \lambda^{\frac{1}{2}}(t)u(t, \lambda(t)x + x(t)) = \pm Q(x) + \mathcal{E}(t, x),$$

for a blow-up solution u with t near the blow-up time T . We suppose that the sign in front of Q in the right hand side of (48) is plus. In order to prove that T is finite, the *ideal situation* would be to have the bound

$$(49) \quad \dot{\lambda}(t) \leq -C < 0.$$

Estimate (49) is not known to hold in the context of (15) but one is able to prove a weaker version of (49) as we explain below.

Substituting (48) in (15), we obtain that the equation solved by $\mathcal{E}(t, x)$ is

$$(50) \quad \begin{aligned} \lambda^3(t)\mathcal{E}_t - (L\mathcal{E})_x - \lambda^2(t)\dot{\lambda}(t)\left(\frac{Q}{2} + xQ'\right) - (\lambda^2(t)\dot{x}(t) - 1)Q' = \\ = \lambda^2(t)\dot{\lambda}(t)\left(\frac{\mathcal{E}}{2} + x\mathcal{E}_x\right) + (\lambda^2(t)\dot{x}(t) - 1)\mathcal{E}_x - (10Q^3\mathcal{E}^2 + 10Q^2\mathcal{E}^3 + 5Q\mathcal{E}^4 + \mathcal{E}^5)_x. \end{aligned}$$

Let us impose that \mathcal{E} satisfies the orthogonality conditions

$$(51) \quad \int_{-\infty}^{\infty} \mathcal{E}(t, x)w(x)dx = \int_{-\infty}^{\infty} \mathcal{E}(t, x)xw'(x)dx = 0,$$

where

$$w(x) := \int_{-\infty}^x \left(\frac{Q(y)}{2} + yQ'(y) \right) dy.$$

The choice (51) is formally possible, again due to the implicit function theorem via an explicit calculation on Q . Let us notice that $w(x)$ does not tend to zero as $x \rightarrow \infty$. Hence one needs to ensure that $\mathcal{E}(t, x)$ decays sufficiently fast to the right and this is one of the major analytical problems in the proof of Theorem 8.1.

Imposing (51) as orthogonality conditions is natural, in view of the equation (50) and the identities satisfied by w

$$(52) \quad Lw' = -2Q, \quad \langle w', Q \rangle = 0.$$

The verification of (52) is straightforward. Under the orthogonality conditions (51), using (52) and integration by parts, we easily get the following (at least formal) identities

$$\langle \lambda^3(t)\mathcal{E}_t, w \rangle = 0, \quad \langle (L\mathcal{E})_x, w \rangle = 2\langle \mathcal{E}, Q \rangle, \quad \left\langle \frac{Q}{2} + xQ', w \right\rangle = c_0, \quad \langle Q', w \rangle = \langle x\mathcal{E}_x, w \rangle = 0,$$

where c_0 is a *positive* constant which can be easily written explicitly in terms of Q . Therefore, multiplying (50) with w , gives the identity

$$(53) \quad \lambda^2(t)\dot{\lambda}(t) = -c_1\langle \mathcal{E}(t), Q \rangle + R_1(\mathcal{E}(t)),$$

where $c_1 := 2/c_0$ and $R_1(\mathcal{E}(t))$ is an explicit expression containing terms of quadratic and higher order with respect to $\mathcal{E}(t)$.

On the other hand, if we denote by $E < 0$ the energy of u , then we get easily from (48), the identity

$$(54) \quad E = E(u(t)) = \lambda^{-2}(t)E(Q + \mathcal{E}(t)).$$

Substituting $Q + \mathcal{E}(t)$ in the energy functional yields

$$(55) \quad E(Q + \mathcal{E}(t)) = -\langle \mathcal{E}(t), Q \rangle + \frac{1}{2}\|\mathcal{E}_x(t)\|_{L^2}^2 + R_2(\mathcal{E}(t)),$$

where $R_2(\mathcal{E}(t))$ is another explicit expression of $\mathcal{E}(t)$ containing only quadratic and higher order terms.

Using (53), (54) and (55), we directly obtain

$$(56) \quad \lambda^2(t)E = -\langle \mathcal{E}(t), Q \rangle + \frac{1}{2}\|\mathcal{E}_x(t)\|_{L^2}^2 + R_2(\mathcal{E}(t)),$$

and

$$(57) \quad \dot{\lambda}(t) = -c_2 - \frac{c_1\lambda^{-2}(t)}{2}\|\mathcal{E}_x(t)\|_{L^2}^2 + \lambda^{-2}(t)R(\mathcal{E}(t)),$$

where $c_2 := -c_1E > 0$ and $R(\mathcal{E}(t)) = R_1(\mathcal{E}(t)) - c_1R_2(\mathcal{E}(t))$.

Notice that if we neglect the third term in the right-hand side of (57), we get an estimate of type (49). Thus one needs to bound the third term in the right-hand side of (57). This can be achieved, with a viriel inequality of type (46). Recall that $ds = \lambda^{-3}(t)dt$. Thus an estimate of type (46) for $\mathcal{E}(t)$ together with (56) provides a bound for $\lambda^3(t)R(\mathcal{E}(t))$ in terms of $I'_A(t)$, i.e. it is realistic to expect that the singularity of $\lambda^{-2}(t)$ can be compensated by the smallness of $R(\mathcal{E}(t))$ (quadratic in $\mathcal{E}(t)$). Since the bound is in terms of $I'_A(t)$, the relevant estimates one can get are only for averages of $R(\mathcal{E}(t))$ on time intervals where $\lambda(t)$ does not vary much.

In [37], Martel-Merle are able to make the previous formal discussion rigorous. More precisely, let us define a sequence (t_n) such that $t_n \rightarrow T$ and such that

$$\|u_x(t_n, \cdot)\|_{L^2} = 2^n \|Q'\|_{L^2}$$

and, for $t \in]t_n, T[$, one has $\|u_x(t, \cdot)\|_{L^2} > 2^n \|Q'\|_{L^2}$. Notice that the existence of (t_n) follows from Theorem 6.1. It turns out that for $n \gg 1$,

$$(58) \quad t_{n+1} - t_n \leq C(\lambda(t_n) - \lambda(t_{n+1}))$$

which is an integrated form of an estimate of type (49). We can deduce directly from (58) that $T < \infty$.

The main point in the proof of (58) is of course the estimate of $\int_{t_n}^{t_{n+1}} |R(\mathcal{E}(t))| dt$. For that purpose, two modulations of the solutions with different orthogonality conditions are used. The first one is very similar to the one considered in the previous section and enjoys the viriel type estimates of Theorem 7.1. One uses the assumption

$$\int_{x_0}^{\infty} u_0^2(x) dx \leq C |x_0|^{-6}, \quad x_0 > 0$$

to get a decay to the right of the solution. This decay allows one to use a second decomposition with the orthogonality conditions (51). One then needs to compare the remainders of the two decompositions. It turns out that one gets cancelations up to second order which is the crucial point in the comparison between the two key quantities $\lambda^2(t)\dot{\lambda}(t)$ and $\langle \mathcal{E}(t), Q \rangle$ which in turn provides the key estimate (58).

REMARK 8.2. — In Theorem 8.1 the L^2 -mass accumulated in the blow-up time is $\|Q\|_{L^2}^2$ (see also [28]). A natural question is whether one may construct blow-up solutions that do not disperse any mass¹⁰ at the blow-up time, i.e. such that $\|u_0\|_{L^2} = \|Q\|_{L^2}$. It turns out that, due to the result in [38], the answer of that question is negative. Therefore the blow-up solutions of Theorem 8.1 necessarily lose some mass on the left of the “main core” during the time evolution.

¹⁰Notice that such solutions exist in the case of the L^2 -critical NLS (see [59]). Moreover they are completely classified (see [41]).

9. FINAL REMARKS

The work of Martel-Merle has already been quite influential. In a remarkable series of recent papers, using many of Martel-Merle ideas, Merle-Raphaël [45, 46, 47, 48, 49, 55] obtained a number of new results on the understanding of the blow-up phenomena for the L^2 -critical nonlinear Schrödinger equations (NLS). The literature on blow-up for NLS is enormous and we refer to the recent books [11, 13, 59] for an introduction to that domain. In [17, 18], the ideas of [34] are successfully used to get the asymptotic stability for the family of solitary waves for the BBM equation which is an alternative to the KdV model in the theory of water waves (see [5]).

Let us point out that the existence of blow-up solutions in the case $p > 5$ remains an open problem. It seems that the approach of Martel-Merle meets serious difficulties in this case. Notice that for $p = 5$, the solution $Q(x - t)$ is *spectrally stable*, i.e. there is no eigenvalue of $\partial_x L$ with positive real part. Therefore the dynamics for solutions with data close to Q can be successfully parameterized by the modulation parameters $\lambda(t)$ and $x(t)$. It seems that in the case $p > 5$, the eigenfunction of $\partial_x L$ with positive real part is also involved in the long time dynamics, even for data close to Q .

Let us finally notice that it would be interesting to extend the asymptotic stability analysis for (4) to the generalized Benjamin-Ono equation

$$(59) \quad u_t + \partial_x(-Hu_x + u^p) = 0$$

which is another important model in the water waves theory (see [3]). In (59), H is the Hilbert transform and $p \geq 2$ is an integer. The equation (59) has a lower order dispersion compared with the KdV equation and $p = 3$ is the smallest value of p such that one may expect blow-up. In the context of (59), the natural energy space is $H^{\frac{1}{2}}(\mathbb{R})$ and therefore a first difficulty is that at the present moment the space $H^{\frac{1}{2}}(\mathbb{R})$ is not covered by the well-posedness theory for (59) (see [26] and the references therein). However, at least at a formal level, equation (59) shares many of the properties of (4) used in the work of Martel-Merle.

REFERENCES

- [1] M.J. Ablowitz and H. Segur – *Solitons and the inverse scattering transform*. SIAM, Philadelphia (1981).
- [2] S. Alinhac – *Blow-up for nonlinear hyperbolic equations*. Progress in Nonlinear Differential Equations and their Applications, Birkhäuser, Boston (1995).
- [3] T. Benjamin – *Internal waves of permanent form in fluids of great depth*. J. Fluid Mech. 29 (1967) 559-592.

- [4] T. Benjamin – *The stability of solitary waves*. Proc. R. Soc. Lond. A 328 (1972) 153-183.
- [5] T. Benjamin, J. Bona, J. Mahony – *Model equations for long waves in nonlinear dispersive systems*. Phil. Trans. R. Soc. Lond. A 272 (1972) 47-78.
- [6] J. Bona – *The stability of solitary waves*. Proc. R. Soc. Lond. A 344 (1975) 363-374.
- [7] J. Bona, R. Smith – *The initial-value problem for the Korteweg-de Vries equation*. Phil. Trans. R. Soc. Lond. A 278 (1975) 555-601.
- [8] J. Bona, P. Souganidis and W. Strauss – *Stability and instability of solitary waves of Korteweg-de Vries type*. Proc. R. Soc. Lond. A 411 (1987) 395-412.
- [9] J. Bona, F. Weissler – *Similarity solutions of the generalized Korteweg-de Vries equation*. Math. Proc. Camb. Phil. Soc. 127 (1999) 323-351.
- [10] A. de Bouard, Y. Martel – *Non existence of L^2 -compact solutions of the Kadomtsev-Petviashvili II equation*. Math. Ann. 328 (2004) 525-544.
- [11] J. Bourgain – *Global solutions of nonlinear Schrödinger equations*. AMS Colloquium Publications, 46 Providence, R.I. 1999.
- [12] J. Bourgain – *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations II. The KdV equation*. Geom. Funct. Anal. 3 (1993) 209-262.
- [13] T. Cazenave – *Semilinear Schrödinger equations*. Courant lecture notes 10 (2003).
- [14] T. Cazenave, P.L. Lions – *Orbital stability of standing waves for some nonlinear Schrödinger equation*. Comm. Math. Phys. 85 (1982) 549-561.
- [15] J.Y. Chemin – *Explosion géométrique pour certaines équations d'ondes non linéaires (d'après Serge Alinhac)*. Séminaire Bourbaki, Exp. 850, Astérisque 266 (2000) 7-20.
- [16] W. Eckhaus, P. Schuur – *The emergence of solitons of the Korteweg-de Vries equation from arbitrary initial conditions*. Math. Meth. Appl. Sci. 5 (1983) 97-116.
- [17] K. El Dika – *Stabilité asymptotique des ondes solitaires de l'équation de Benjamin-Bona-Mahony*. C. R. Acad. Sci. Paris Ser. I 337 (2003) 649-652.
- [18] K. El Dika – *Asymptotic stability of solitary waves for the Benjamin-Bona-Mahony equation*. Preprint 2003.
- [19] C. Fermanian, F. Merle, H. Zaag – *Stability of the blow-up profile of non-linear heat equations from a dynamical system point of view*. Math. Ann. 317 (2000) 347-387.
- [20] S. Friedlander, W. Strauss, M. Vishik – *Nonlinear instability in an ideal fluid*. Ann. Inst. Henri Poincaré (Analyse non linéaire) 14 (1997) 187-209.
- [21] J. Ginibre, Y. Tsutsumi – *Uniqueness of solutions for the generalized Korteweg-de Vries equation*. SIAM J. Math. Anal. 20 (1989) 1388-1425.
- [22] L. Glangetas, F. Merle – *A geometric approach of existence of blow-up solutions*. Preprint 1995.
- [23] M. Grillakis, J. Shatah and W. Strauss – *Stability theory of solitary waves in the presence of symmetry*. J. Funct. Anal. 74 (1987) 160-197.

- [24] L. Hörmander – *The analysis of linear partial differential operators I*. Springer-Verlag, 1983.
- [25] T. Kato – *On the Cauchy problem for the (generalized) Korteweg-de Vries equation*. Advances in Math. Suppl. Stud. 8 (1983) 93-128.
- [26] C. Kenig, K. Koening – *On the local well-posedness of the Benjamin-Ono and modified Benjamin-Ono equations*. Math. Res. Letters 10 (2003) 879-895.
- [27] C. Kenig, G. Ponce, L. Vega – *Well-posedness and scattering results for the generalized Korteweg-de Vries equation via the contraction principle*. Comm. Pure Appl. Math. 46 (1993) 527-620.
- [28] C. Kenig, G. Ponce, L. Vega – *On the concentration of blow-up solutions for the generalized KdV equation critical in L^2* . Contemp. Math. 263 (2000) 131-156.
- [29] G.L. Lamb Jr. – *Elements of soliton theory*. John Wiley & Sons, New York (1980).
- [30] C. Laurent, Y. Martel – *Smoothness and exponential decay of L^2 -compact solutions of the generalized KdV equation*. Comm. Partial Differential Equations 28 (2003) 2093-2107.
- [31] J. Maddocks, R. Sachs. – *On the stability of KdV multi-solitons*. Comm. Pure Appl. Math. 46 (1993) 867-901.
- [32] Y. Martel – *Multi-soliton-type solutions of the generalized KdV equations*. Amer. J. Math. (to appear).
- [33] Y. Martel and F. Merle – *Instability of solitons for the critical generalized Korteweg-de Vries equation*. Geom. Funct. Anal. 11 (2001) 74-123.
- [34] Y. Martel and F. Merle – *Asymptotic stability of solitons for subcritical generalized KdV equations*. Arch. Ration. Mech. Anal. 157 (2001) 219-254.
- [35] Y. Martel and F. Merle – *A Liouville Theorem for the critical generalized Korteweg-de Vries equation*. J. Math. Pures Appl. 79 (2000) 339-425.
- [36] Y. Martel and F. Merle – *Stability of the blow-up profile and lower bounds on the blow-up rate for the critical generalized Korteweg-de Vries equation*. Ann. Math. 155 (2002) 235-280.
- [37] Y. Martel and F. Merle – *Blow-up in finite time and dynamics of blow-up solutions for the L^2 -critical generalized KdV equation*. J. Amer. Math. Soc. 15 (2002) 617-663.
- [38] Y. Martel and F. Merle – *Nonexistence of blow-up solution with minimal L^2 -mass for the critical GKdV*. Duke Math. J. 115 (2002) 385-408.
- [39] Y. Martel and F. Merle – *Asymptotic stability of solitons for subcritical generalized KdV equations revisited*. Preprint 2004.
- [40] Y. Martel and F. Merle, Tai-Peng Tsai – *Stability and asymptotic stability in the energy space of the sum of N solitons for subcritical gKdV equations*. Comm. Math. Phys. 231 (2002) 347-373.
- [41] F. Merle – *Determination of blow-up solutions with minimal mass for nonlinear Schrödinger equations with critical power*. Duke Math. J. 69 (1993) 427-454.

- [42] F. Merle – *Asymptotics for L^2 -minimal blow-up solutions of critical nonlinear Schrödinger equation*. Ann. Inst. Henri Poincaré (Analyse non linéaire) 13 (1996) 553-565.
- [43] F. Merle – *Blow-up phenomena for critical nonlinear Schrödinger and Zakharov equations*. Proceeding of the International congress of Mathematicians (Berlin 1998), Doc. Math. Extra volume ICM 1998 III (1998) 57-66.
- [44] F. Merle – *Existence of blow-up solutions in the energy space for critical generalized KdV equation*. J. Amer. Math. Soc. 14 (2001) 555-578.
- [45] F. Merle, P. Raphaël – *Blow-up dynamic and upper bound on the blow-up rate for critical nonlinear Schrödinger equation*. Ann. Math. (to appear).
- [46] F. Merle, P. Raphaël – *Sharp upper bound on the blow-up rate for critical nonlinear Schrödinger equation*. Geom. Funct. Anal. 13 (2003) 591-642.
- [47] F. Merle, P. Raphaël – *On universality of blow-up profile for L^2 -critical nonlinear Schrödinger equation*. Invent. Math. 156 (2004) 565-672.
- [48] F. Merle, P. Raphaël – *Sharp lower bound on the blow-up rate for critical nonlinear Schrödinger equation*. Preprint 2004.
- [49] F. Merle, P. Raphaël – *Profiles and quantization of the blow-up mass for the critical nonlinear Schrödinger equation*. Comm. Math. Phys. (to appear).
- [50] F. Merle, L. Vega – *L^2 -stability of solitons for KdV equation*. Int. Math. Res. Notices (2003) 735-753.
- [51] F. Merle, H. Zaag – *A Liouville theorem for a vector valued nonlinear heat equation and applications*. Math. Ann. 316 (2000) 103-137.
- [52] R. Miura – *The Korteweg-de Vries equation : a survey of results*. SIAM Rev. 18 (1976) 412-459.
- [53] R. Pego, M. Weinstein – *Eigenvalues, and instability of solitary waves*. Phil. Trans. R. Soc. Lond. A 340 (1992) 47-94.
- [54] R. Pego, M. Weinstein – *Asymptotic stability of solitary waves*. Comm. Math. Phys. 164 (1994) 305-349.
- [55] P. Raphaël – *Stability of the log log bound for blow-up solutions to the critical nonlinear Schrödinger equation*. Math. Ann. (to appear).
- [56] J.C. Saut – *Sur quelques généralisations de l'équation de Korteweg-de Vries*, J. Math. Pures Appl. 58 (1979) 21-61.
- [57] J.C. Saut – *Remarks on generalized Kadomtsev-Petviashvili equations*. Indiana Univ. Math. J. 42 (1993) 1011-1026.
- [58] P. Schuur – *Asymptotic analysis of soliton problems*. Lecture Notes in Math. 1232 (1986), Springer-Verlag, Berlin.
- [59] S. Sulem, P.L. Sulem – *The nonlinear Schrödinger equation. Self-focusing and wave collapse*, in Applied Mathematical Sciences, 139, Springer-Verlag, New York, 1999.

- [60] M. Weinstein – *Nonlinear Schrödinger equations and sharp interpolation estimates.* Comm. Math. Phys. 87 (1983) 567-576.
- [61] M. Weinstein – *Modulational stability of ground states of nonlinear Schrödinger equations.* SIAM J. Math. Anal. 16 (1985) 472-491.
- [62] M. Weinstein – *Lyapunov stability of ground states of nonlinear dispersive equations.* Comm. Pure Appl. Math. 39 (1986) 51-68.
- [63] M. Weinstein – *On the structure and formation of singularities in solutions to nonlinear dispersive equations.* Comm. Partial Differential Equations 11 (1986) 545-565.

Nickolay TZVETKOV

Université de Paris Sud

Département de Mathématiques

Bâtiment 425

F-91405 Orsay Cedex

E-mail: `nikolay.tzvetkov@math.u-psud.fr`

**MÉTHODES GÉOMÉTRIQUES
DANS L'ÉTUDE DES ÉQUATIONS D'EINSTEIN**
[d'après Christodoulou, Klainerman, Nicolò et Rodnianski]

par Serge ALINHAC

Dans cet exposé, nous présentons les outils développés dans les quinze dernières années par Christodoulou, Klainerman, Nicolò et Rodniansky, outils qui ont permis d'importants progrès dans l'étude des équations hyperboliques non-linéaires, notamment des équations d'Einstein. Il existe bien entendu beaucoup d'autres travaux sur ce sujet dont nous ne pourrions parler, et j'espère que leurs auteurs me pardonneront de m'être volontairement limité au choix retenu par Bourbaki. Pour ce qui est des aspects géométriques et physiques des équations d'Einstein, on consultera avec profit l'exposé de J.-P. Bourguignon [3] dans ce même séminaire. Notre propos est en quelque sorte complémentaire de celui de Bourguignon : il vise à rendre compte de façon un peu technique de méthodes qui se sont clarifiées au fil des ans, et dont le « cœur » apparaît dans de nombreux travaux.

1. LES PROBLÈMES

Nous ferons ici référence à trois groupes de travaux, très liés entre eux :

- i) Les travaux de Klainerman [12] et Klainerman et Rodniansky [14] sur les équations d'ondes quasi-linéaires

$$\partial_t^2 \phi - g^{ij}(\phi) \partial_{ij}^2 \phi = N(\phi, \nabla \phi).$$

- ii) Les travaux de Christodoulou et Klainerman [7], Klainerman et Nicolò [13] sur les équations d'Einstein $R_{\alpha\beta} = 0$.
- iii) Les travaux de Klainerman et Rodniansky [15]-[20] sur les équations d'Einstein.

La discussion des travaux i), outre son intérêt intrinsèque, est utile pour mieux appréhender la démarche des travaux ii) et iii).

Dans tous les cas, il s'agit de problèmes qui s'écrivent dans des coordonnées bien choisies comme des systèmes hyperboliques non-linéaires, dont la partie principale est simplement $L_g \times Id$, L_g étant le d'Alembertien associé à une métrique lorentzienne g . Il n'y a donc qu'une seule géométrie associée à de tels systèmes. Deux types de problèmes se posent alors :

- A) Le problème de l'existence *globale en temps* de solutions C^∞ associées à des données de Cauchy elles-mêmes C^∞ et suffisamment décroissantes lorsque $|x| \rightarrow \infty$.
- B) Le problème de l'existence *locale en temps* de solutions associées à des données de Cauchy peu régulières, par exemple dans un espace de Sobolev $H^s(\mathbf{R}^3)$, avec s aussi petit que possible.

Dans le problème A, comme on peut le voir en consultant Hörmander [9] ou la première partie de l'exposé de Chemin [4], l'enjeu est de prouver la décroissance en temps, uniformément en espace, de la solution et de ses dérivées.

D'autre part, les problèmes A et B sont liés entre eux de deux façons :

- i) Si l'on parvient à abaisser s jusqu'à un niveau contrôlé par une quantité « conservée » (comme l'énergie pour l'équation des ondes, par exemple), on obtient l'existence globale de solutions peu régulières.
- ii) Une procédure d'attaque du problème B due à Bahouri et Chemin [1][2], et reprise dans les références i), est la suivante : pour λ fixé grand, notons g_λ une régularisation de la métrique g ($g_\lambda \rightarrow g$ lorsque $\lambda \rightarrow \infty$) ; on établit alors l'équation (à coefficients régularisés) vérifiée par le « bloc » $\Delta_\lambda u$ (qui est, dans l'écriture de u à l'aide de sa transformée de Fourier $\hat{u}(\xi)$, la partie de u pour laquelle $|\xi|$ est de l'ordre de λ)

$$L_{g_\lambda}(\Delta_\lambda u) = R_\lambda.$$

Après un changement d'échelle convenable, le problème se réduit à étudier la décroissance, sur un intervalle $[0, T]$, $T \leq \lambda^a$, $a > 0$, de la solution v d'une équation d'onde

$$L_{h_\lambda} v = 0,$$

où h_λ est une nouvelle métrique déduite de g . On peut alors « recoller » les informations sur les $\Delta_\lambda u$ pour obtenir l'estimation voulue sur la norme H^s de u .

Ce qui est important est que l'étude du problème A, comme celle du problème B par l'approche ii), se réduit à prouver la décroissance des solutions d'une équation *linéaire* associée à une métrique h déduite de g , et jouissant de propriétés connues (P). Le caractère *non-linéaire* du problème consiste en ceci : la métrique h dépend en fait de la solution ϕ . Il faut donc supposer certaines propriétés (Q) de ϕ sur un intervalle de temps $[0, T[$, qui impliquent les propriétés correspondantes (P) de h sur le même intervalle, lesquelles permettent d'établir que (Q) a lieu en fait sur un intervalle plus grand $[0, T + \epsilon]$. C'est le procédé *d'induction sur le temps*.

Nous consacrerons les sections 2 à 6 à décrire les outils d'étude des équations linéaires, ne retournant aux aspects non-linéaires spécifiques qu'au paragraphe 7. La section 8, enfin, évoquera un travail en cours sur la « conjecture H^2 » pour les équations d'Einstein.

Notons que, dans un article récent [21], Lindblad et Rodniansky ont démontré l'existence globale de solutions régulières des équations d'Einstein écrites en coordonnées harmoniques, sans utiliser les outils géométriques qui font l'objet de cet exposé ; la contrepartie en est qu'ils ne semblent pas obtenir les propriétés asymptotiques fines des

composantes de la courbure (« peeling properties ») telles qu'on peut les trouver dans [13], par exemple.

2. LE CŒUR DU DISPOSITIF : FEUILLETAGES, REPÈRES ET FONCTIONS OPTIQUES

Dans toute la suite de l'exposé, nous nous placerons dans \mathbf{R}^4 , où nous tâcherons d'utiliser le moins possible les coordonnées usuelles

$$x^0 = t, x = (x^1, x^2, x^3).$$

Nous supposerons donnée une métrique lorentzienne g , de signature $(-, +, +, +)$, le plus souvent proche de la métrique « plate » η de l'espace de Minkovski

$$\eta = -dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2.$$

Les composantes de g en coordonnées locales seront $g_{\alpha\beta}$, et $g^{\alpha\beta}$ seront les éléments de la matrice inverse de $g_{\alpha\beta}$. Nous noterons $g(X, Y) = \langle X, Y \rangle$, et D la connexion canonique associée à g . Le gradient et le hessien d'une fonction f seront définis par

$$\langle \nabla f, X \rangle = Xf, \nabla^2 f(X, Y) = XYf - (D_X Y)f,$$

le d'Alembertien associé à g étant l'opérateur

$$L_g f = g^{\alpha\beta} \nabla^2 f_{\alpha\beta}.$$

Nous considérerons dans la suite deux situations géométriques distinctes :

(I) Celle des travaux i) sur les équations d'ondes pour une métrique scindée

$$-dt^2 + g_{ij} dx^i dx^j,$$

dans laquelle le feuilletage par les surfaces $\Sigma_{t_0} = \{t = t_0\}$ joue un rôle essentiel.

(II) Celle des travaux de Klainerman et Nicolò et Klainerman et Rodniansky sur les équations d'Einstein, dans lesquels aucune coordonnée n'apparaît *a priori*.

Pour le lecteur désireux d'approfondir, signalons que nous avons adopté les notations de [14] pour décrire la situation I, tandis que nous adoptons celles de [13] pour décrire la situation II. Enfin, au paragraphe 8, nous gardons les notations de [18].

Dans les deux situations géométriques, on suppose donné un feuilletage par des variétés de dimension deux, chacune homéomorphe à une 2-sphère standard. Ce feuilletage est tel qu'en chaque point $p \in S$, la restriction de g à l'orthogonal $H = (T_p S)^\perp$ est de signature $(-, +)$, et l'on note (e_3, e_4) des vecteurs isotropes de H , pointant vers le futur, de directions respectivement « rentrantes » et « sortantes » (cf. section 2). Si l'on choisit un repère orthonormé (e_1, e_2) sur S , on dispose ainsi d'un *repère isotrope* (e_1, e_2, e_3, e_4) (« null frame »), qui jouera un rôle central dans la suite (tous les tenseurs seront décomposés sur ce repère).

Si la collection des 3-plans engendrés par (e_1, e_2, e_4) est intégrable, on peut voir que les variétés intégrales correspondantes, les « cônes sortants », sont les surfaces de niveaux d'une fonction u vérifiant l'équation eikonale

$$g^{\alpha\beta} \partial_\alpha u \partial_\beta u = 0.$$

Une telle fonction est dite « fonction optique ». On définit également une fonction r , constante sur chaque sphère du feuilletage, par $4\pi r^2 = \text{aire} S$.

Dans la situation I, on définit d'abord une fonction optique u en imposant de plus $u = t - r$ infiniment près de l'axe des temps. Le feuilletage en sphères est alors

$$S_{t_0, u_0} = \{t = t_0, u = u_0\}.$$

En notant

$$L' = -\nabla u, b^{-1} = \partial_t u, N = -b \partial^i u \partial_i,$$

on pose

$$e_4 = L = bL' = \partial_t + N, e_3 = \underline{L} = \partial_t - N,$$

et l'on a les propriétés

$$\langle L, L \rangle = 0, \langle \underline{L}, \underline{L} \rangle = 0, \langle L, \underline{L} \rangle = -2.$$

On notera qu'en général le système des 3-plans (e_1, e_2, e_3) n'est pas intégrable. On complète le dispositif en posant néanmoins $\underline{u} = 2t - u$.

Dans la situation II au contraire, on définit d'abord des fonctions optiques « entrantes » et « sortantes » \underline{u} et u de la façon suivante :

- i) On construit une fonction \underline{u}_0 sur $\Sigma_0 = \{t = 0\}$, dont les surfaces de niveaux jouissent de propriétés que nous expliquerons... à la section 7. On choisit alors pour \underline{u} la solution entrante de l'équation eikonale qui vaut \underline{u}_0 pour $t = 0$. On se limite ici à un domaine de Σ_0 délimité par les deux surfaces $\underline{u}_0 = \nu_0$ et $\underline{u}_0 = \underline{u}_*$.
- ii) Sur le cône rentrant $\underline{C}_* = \{\underline{u} = \underline{u}_*\}$ (que les auteurs appellent « the last slice »), on choisit une fonction u_* dont les surfaces de niveaux jouissent aussi de propriétés expliquées à la section 7. On définit alors u comme la solution sortante de l'équation eikonale valant u_* sur \underline{C}_* .

Une fois définies u et \underline{u} , on pose

$$L = -\nabla u, \underline{L} = -\nabla \underline{u}, 2\Omega^2 = -\langle L, \underline{L} \rangle^{-1} = -(g^{\alpha\beta} \partial_\alpha u \partial_\beta \underline{u})^{-1},$$

et finalement on choisit

$$e_3 = 2\Omega \underline{L}, e_4 = 2\Omega L$$

en sorte que $\langle e_3, e_4 \rangle = -2$. Le feuilletage en sphères est finalement défini par

$$S_{\lambda, \nu} = \{u = \lambda, \underline{u} = \nu\}.$$

3. LES OBJETS GÉOMÉTRIQUES DANS L'ESPACE-TEMPS DE MINKOWSKI

Dans l'espace-temps de Minkowski, la situation géométrique de la section 2 est très simple : on choisit

$$u = t - r, \underline{u} = t + r, r = |x|, x = r\omega,$$

et donc

$$e_4 = L = -\nabla u = \partial_t + \partial_r, e_3 = \underline{L} = -\nabla \underline{u} = \partial_t - \partial_r, r\partial_r = x^i \partial_i.$$

Le feuilletage en sphères est celui des sphères standard dans les plans horizontaux, et des champs tangents à ces sphères sont

$$R_i = (x \wedge \partial)_i.$$

Dans la théorie des équations hyperboliques, tous ces objets ont des fonctions multiples plus ou moins « évidentes », que nous allons détailler, car il sera nécessaire de les distinguer plus tard.

1. Nous pouvons définir au moins deux notions d'infini : ce qui se passe lorsque $r \rightarrow \infty$, et ce qui se passe lorsque $\underline{u} \rightarrow \infty$, pour u fixé (« null infinity »). La première notion est claire, et les puissances de r apparaissent partout comme « poids » dans les estimations de décroissance cherchées. Par ailleurs, observons qu'une solution ϕ de l'équation des ondes avec données de Cauchy C_0^∞ peut s'écrire (cf. [9])

$$\phi(x, t) = (1/r)F(r - t, \omega, 1/r).$$

Pour u fixé, $r\phi$ a une limite qui est $F(-u, \omega, 0)$ (que l'on peut calculer explicitement à l'aide des transformées de Radon des données [9]). Dans [13], les auteurs établissent toutes les limites de ce type, qui sont importantes pour l'interprétation physique des résultats.

2. Les cônes sortants et rentrants permettent de préciser les domaines de détermination de divers sous-ensembles, c'est-à-dire de préciser la structure causale de l'univers que l'on décrit.
3. Nous notons que

$$\partial_t = \frac{1}{2}(L + \underline{L}), S = t\partial_t + x^i \partial_i = \frac{1}{2}(ue_3 + \underline{u}e_4),$$

$$K_0 = (t^2 + x^2)\partial_t + 2tx^i \partial_i = \frac{1}{2}(u^2 e_3 + \underline{u}^2 e_4).$$

Les champs ∂_t et K_0 sont utiles (entre autres choses) comme *multiplicateurs* permettant d'obtenir des inégalités d'énergie pour l'équation des ondes. Cela signifie que pour $X = \partial_t$ ou $X = K_0$, l'on calcule en intégrant par parties

$$\int_{[0, T] \times \mathbb{R}^3} L_\eta \phi X \phi dx dt,$$

obtenant ainsi un contrôle au temps T d'une énergie positive $E(\phi)(T)$ de ϕ . Pour $X = \partial_t$, on obtient simplement l'énergie standard $E(\phi)(T) = \|\nabla\phi(\cdot, T)\|_{L^2}^2$, tandis que pour $X = K_0$ (« inégalité conforme »), on obtient une énergie équivalente à

$$(3.1) \quad \|\phi\|^2 + \Sigma\|R_i\phi\|^2 + \|S\phi\|^2 + \Sigma\|H_i\phi\|^2,$$

la norme $\|\cdot\|$ étant la norme L^2 à T fixé, les $H_i = t\partial_i + x^i\partial_t$ étant les champs de rotations hyperboliques. Nous expliquerons à la section 4 la généralisation de ces procédures.

4. Les champs considérés en 3 ont aussi de bonnes propriétés de commutation avec L_η . En effet,

$$[L_\eta, R_i] = 0, [L_\eta, H_i] = 0, [L_\eta, S] = 2L_\eta.$$

Le point 5 sera consacré à ce problème de commutation dans le cas non plat.

5. Enfin, en revenant à l'écriture de 1 d'une solution régulière

$$\phi(x, t) = (1/r)F(r - t, \omega, 1/r),$$

on observe que les différentes composantes de $\nabla\phi$ ont des comportements différents. Plus précisément,

$$L\phi \sim 1/r^2, (R_i/r)(\phi) \sim 1/r^2, \underline{L}\phi \sim 1/r.$$

Plus tard, il sera essentiel d'étudier séparément les diverses composantes des tenseurs que nous voulons estimer dans des repères isotropes analogues à (e_1, e_2, e_3, e_4) .

4. LES INÉGALITÉS D'ÉNERGIE ET LE TENSEUR D'ÉNERGIE-MOMENT

Pour comprendre, dans le cas plat, le formalisme que nous allons expliquer, on pourra consulter [6].

4.1. Le cas des équations d'ondes

Le tenseur d'énergie-moment Q associé à une fonction ϕ est défini par

$$Q_{\alpha\beta} = \partial_\alpha\phi\partial_\beta\phi - (1/2)g_{\alpha\beta}(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi).$$

Ses propriétés principales sont les suivantes :

- i) Si $L_g\phi = F$, alors

$$D^\alpha Q_{\alpha\beta} = F\partial_\alpha\phi.$$

- ii) Si X et Y sont de type temps et orientés vers l'avenir, $Q(X, Y) \geq 0$.

Dans la situation I, la structure des composantes de Q dans notre repère habituel est la suivante :

$$Q_{LL} = |L\phi|^2, Q_{\underline{L}\underline{L}} = |\underline{L}\phi|^2, Q_{L\underline{L}} = \Sigma_{a=1,2}|e_a\phi|^2,$$

$$Q_{Le_a} = (L\phi)(e_a\phi), Q_{\underline{L}e_a} = (\underline{L}\phi)(e_a\phi),$$

$$Q_{e_ae_b} = (e_a\phi)(e_b\phi) - (1/2)\delta_{ab}[-(L\phi)(\underline{L}\phi) + \Sigma_{a=1,2}|e_a\phi|^2].$$

Choisissons maintenant un champ X , et posons $P_\alpha = Q_{\alpha\beta}X^\beta$. Il vient

$$D^\alpha P_\alpha = (1/2)Q^{\alpha\beta}\pi_{\alpha\beta} + FX(\phi).$$

Autrement dit, on écrit le produit $(L_g\phi)(X\phi)$ sous la forme d'une divergence plus une forme quadratique en les dérivées premières de ϕ . Le tenseur π est le tenseur de déformation de X défini par

$$\pi_{\alpha\beta} = D_\alpha X_\beta + D_\beta X_\alpha.$$

On voit que c'est aussi la dérivée de Lie de g . Selon la formule ci-dessus, on a tout intérêt à choisir un champ X (de type temps) pour lequel π serait le plus petit possible. Le cas $\pi = 0$ est celui des champs de Killing (dont le flot laisse g invariante). Malheureusement, de tels champs n'existent pas en général. Les deux possibilités retenues par Klainerman sont d'une part $X = \frac{1}{2}(L + \underline{L}) = \partial_t$, généralisant ainsi l'inégalité d'énergie standard, d'autre part $X = \frac{1}{2}(u^2\underline{L} + \underline{u}^2L)$, généralisant l'inégalité conforme du cas plat.

Il semble que ces choix soient guidés par deux considérations, outre le fait que ce sont les « bons choix » dans le cas plat :

- a) D'abord, ces champs s'expriment simplement en termes de $u, \underline{u}, L, \underline{L}$, ce qui permet de calculer leurs tenseurs π en fonction des « coefficients du repère » (que nous discuterons au point 6).
- b) Mais la vraie raison est la suivante : si nous décomposons la forme quadratique $Q^{\alpha\beta}\pi_{\alpha\beta}$ dans notre repère, tous les termes font intervenir au moins une « bonne » dérivée de ϕ (c'est-à-dire $e_i\phi$, $i = 1, 2$ ou $i = 4$), à l'exception du terme $\pi_{LL}Q_{\underline{L}\underline{L}}$ qui contient $|\underline{L}\phi|^2$. Mais justement, on peut calculer que

$${}^{(T)}\pi_{LL} = 0, {}^{(K_0)}\pi_{LL} = 0.$$

Avec le choix du multiplicateur K_0 , par exemple, on obtient le contrôle à l'instant t d'une énergie équivalant à

$$(4.1) \quad E(\phi)(t) = \int_{\Sigma_t} [|\phi|^2 + t^2|L\phi|^2 + t^2\Sigma_{a=1,2}|e_a\phi|^2 + u^2|\underline{L}\phi|^2] dv,$$

dv étant la mesure de volume induite par g sur Σ_t . Il est intéressant de comparer cette formule avec (3.1), sachant que, dans le cas plat, on a les relations

$$S + \Sigma\omega_i H_i = \underline{u}L, S - \Sigma\omega_i H_i = u\underline{L}.$$

4.2. Le cas relativiste

Dans la situation II de l'étude des équations d'Einstein, les auteurs utilisent des estimations d'énergie du tenseur de courbure R . On sait que R satisfait la deuxième identité de Bianchi (« Bianchi equations »)

$$D_{[\epsilon} R_{\gamma\delta]\alpha\beta} = 0,$$

où le crochet sur les indices signifie la somme alternée (comme pour la dérivée extérieure d'une 2-forme). En fait, puisque l'on étudie les équations d'Einstein dans le vide, qui s'écrivent $R_{\alpha\beta} = 0$, le tenseur de Weyl W est identique à R , satisfait les mêmes équations de Bianchi, les mêmes symétries, avec en plus $W^\alpha_{\beta\alpha\gamma} = 0$. On peut établir une analogie entre ces équations et les équations de Maxwell. Le fait est qu'il existe une « machinerie » d'inégalités d'énergie analogue à celle que nous avons expliquée pour l'équation des ondes. On définit un tenseur d'énergie-moment (dit « de Bel-Robinson ») par

$$Q_{\alpha\beta\gamma\delta} = W_{\alpha\rho\gamma\sigma} W_{\beta\delta}^{\rho\sigma} + {}^*W_{\alpha\rho\gamma\sigma} {}^*W_{\beta\delta}^{\rho\sigma},$$

où ${}^*W_{\alpha\beta\gamma\delta} = 1/2\epsilon_{\alpha\beta\mu\nu} W^{\mu\nu}_{\gamma\delta}$ est le dual de Hodge de W , défini à l'aide de la forme volume ϵ . Le tenseur Q jouit des propriétés suivantes :

- i) Q est symétrique et de trace nulle par rapport à tout couple d'indices.
- ii) $Q(X, Y, Z, U)$ est positif si les quatre champs X, Y, Z, U sont de type temps orientés vers le futur.
- iii) Si W est solution des équations de Bianchi,

$$D^\alpha Q_{\alpha\beta\gamma\delta} = 0.$$

Comme en 4.1, on en déduit qu'en posant, pour trois « multiplicateurs » donnés X, Y, Z ,

$$P_\alpha = Q_{\alpha\beta\gamma\delta} X^\beta Y^\gamma Z^\delta,$$

on trouve, si W satisfait les équations de Bianchi,

$$D^\alpha P_\alpha = 1/2 Q_{\alpha\beta\gamma\delta} [{}^{(X)}\pi^{\alpha\beta} Y^\gamma Z^\delta + {}^{(Y)}\pi^{\alpha\gamma} X^\beta Z^\delta + {}^{(Z)}\pi^{\alpha\delta} X^\beta Y^\gamma].$$

On voit donc que le choix des multiplicateurs X, Y, Z obéit exactement aux mêmes contraintes que pour l'équation des ondes : on prendra comme avant soit T , soit K_0 , ce qui, compte tenu des symétries, ne laisse que quatre possibilités. Dans [13], les auteurs ne retiennent en fait que les choix (K_0, K_0, K_0) ou (K_0, K_0, T) , ce qui correspond à l'inégalité conforme pour l'équation des ondes.

La nouveauté ici est qu'il n'y a pas de coordonnée t . On va donc intégrer $D^\alpha P_\alpha$ dans un domaine de détermination $K_{\lambda\nu}$ bordé par deux surfaces de niveau $u = \lambda$, $\underline{u} = \nu$ et une portion de $\Sigma_0 = \{t = 0\}$: au lieu d'obtenir une énergie qui s'exprime comme une intégrale sur un plan horizontal, comme dans le cas des ondes, on obtient deux intégrales sur le bord latéral de $K_{\lambda\nu}$, qui sont en fait, de façon analogue à (4.1), des intégrales des carrés de diverses composantes de R , agrémentées de divers poids qui sont des puissances de u et \underline{u} (voir [13]).

5. COMMUTATIONS

Il s'avère qu'en général, pour obtenir des estimations ponctuelles à poids (qui sont ici r et $\langle u \rangle = (1 + |u|^2)^{1/2}$) sur les solutions d'une équation hyperbolique, le contrôle de l'énergie de cette solution ne suffit pas. Dans l'esprit des inégalités de Sobolev, il faut aussi contrôler l'énergie d'un certain nombre de dérivées de la solution, résultat que l'on n'atteint qu'en établissant des équations vérifiées par ces dérivées, c'est-à-dire en les « commutant » à l'opérateur.

5.1. Le cas des ondes

Dans le cas plat (voir [9] ou l'exposé de Chemin [4]), on commute avec l'équation des produits Z^α , où les Z sont pris parmi les « champs de Klainerman »

$$\partial_\alpha, R_i, S, H_i.$$

Dans le cas général, on utilise la formule suivante (cf. [12]) :

$$[L_g, X]\phi = \pi^{\alpha\beta} \nabla^2 \phi_{\alpha\beta} + D_\alpha \pi^{\alpha\beta} \partial_\beta \phi - (1/2) \partial^\beta (tr\pi) \partial_\beta \phi.$$

On notera que le commutateur ne dépend que du tenseur de déformation $\pi = {}^{(X)}\pi$ de X , le même qui intervenait déjà dans les inégalités d'énergie. On remarquera à ce propos que, contrairement aux apparences, π contient des dérivées $X(g^{\alpha\beta})$, comme il est normal pour un commutateur, car

$$\pi^{\alpha\beta} = \partial^\alpha (X^\beta) + \partial^\beta (X^\alpha) - X(g^{\alpha\beta}).$$

Il faut donc commuter avec L_g des champs pour lesquels π soit le plus petit possible, et qui apportent l'information souhaitée. Dans le cas de l'équation des ondes (situation I), pour gagner le contrôle des dérivées de ϕ d'ordre supérieur à un, le plus simple est de commuter le champ $T = (1/2)(L + \underline{L})$, dont le tenseur de déformation a de bonnes propriétés. Utilisant alors l'équation, on obtient un contrôle de $\Delta_g \phi$ qui, via la théorie elliptique, donne finalement le contrôle de toutes les dérivées.

Il est cependant des cas où un contrôle de l'analogue des « champs de Klainerman » appliqués à ϕ est souhaitable, pour les mêmes raisons que dans le cas plat (par exemple, pour pouvoir ensuite appliquer l'inégalité de Klainerman [9] [4]). Dans ce cas, on dispose déjà d'un analogue du champ de scaling S défini par

$$S = (1/2)(u\underline{L} + \underline{u}L).$$

Pour ce qui est des rotations analogues aux R_i , on note que les cônes sortants sont des surfaces caractéristiques pour L_g ; il est donc raisonnable de choisir des champs iO tangents aux sphères du feuilletage, c'est-à-dire ici aux sphères S_{t_0, u_0} . Le choix des auteurs est de préserver autant que faire se peut les relations existantes dans le cas plat

$$[R_i, R_j] = \epsilon_{ijk} R^k, [R_i, \partial_r] = 0, \langle e_i, R_j \rangle = 0, i = 3, 4.$$

Pour ce faire (toujours dans la situation I), on considère, dans Σ_0 , le feuilletage en sphères $u = u_0$, et son champ unitaire normal (sortant) N . Le flot de N permet de

ramener sur une sphère donnée les champs (standard!) R_i qui vivent sur la sphère à l'infini : cela est dû au fait que le feuilletage ressemble, à l'infini, au feuilletage standard, et que les champs R_i sont homogènes d'ordre zéro. Pour obtenir les champs tangents à la sphère S_{t_0, u_0} , on prend l'image des champs tangents à la sphère S_{0, u_0} par le flot de L , pendant le temps t_0 . On assure ainsi les relations

$$[{}^i O, {}^j O] = \epsilon_{ijk}^k O, [L, {}^i O] = 0.$$

Il y a un certain arbitraire dans cette définition, car on aurait par exemple obtenu d'autres champs en transportant les champs standard de l'infini directement dans le plan Σ_{t_0} .

Finalement, notons qu'il ne semble pas qu'on dispose d'analogues aux rotations hyperboliques H_i du cas plat, jouissant de bonnes propriétés de commutation.

5.2. Le cas relativiste

On a vu que, dans ce cas, les auteurs utilisent les équations de Bianchi pour contrôler la courbure à l'aide d'inégalités d'énergie. Les champs que l'on souhaite commuter aux équations de Bianchi sont les mêmes que plus haut : $T, S, {}^{(i)}O$. Notons ici que les rotations ${}^{(i)}O$ sont définies à partir des rotations sur la surface $\{\underline{u}_0 = \underline{u}_*\}$ de Σ_0 à l'aide des flots de deux champs « équivariants » (c'est-à-dire que leurs flots appliquent les sphères du feuilletage les unes dans les autres) : on pousse d'abord les champs le long de \underline{C}_* par le flot de $2\Omega^2 \underline{L}$, puis, de la sphère à laquelle on a abouti, le long du cône sortant par le flot de $2\Omega^2 L$ jusqu'à la sphère voulue.

La principale différence est qu'on dérive le tenseur de Weyl, et non pas une fonction, ce qui nécessite quelques précautions. Au lieu de considérer la dérivée de Lie usuelle $\mathcal{L}_X W$, on doit introduire, afin de préserver les symétries de W , la dérivée de Lie modifiée (par des termes linéaires)

$$\hat{\mathcal{L}}_X W = \mathcal{L}_X W - (1/2)^{(X)}[W] + (3/8)tr^{(X)}\pi W,$$

où l'on a posé (avec ${}^{(X)}\pi = \pi$)

$${}^{(X)}[W]_{\alpha\beta\gamma\delta} = \pi_\alpha^\mu W_{\mu\beta\gamma\delta} + \pi_\beta^\mu W_{\alpha\mu\gamma\delta} + \pi_\gamma^\mu W_{\alpha\beta\mu\delta} + \pi_\delta^\mu W_{\alpha\beta\gamma\mu}.$$

Nous renvoyons le lecteur à [3] pour des aperçus géométriques sur cette définition.

6. LES ÉQUATIONS DE STRUCTURE : PROPAGATION ET ELLIPTICITÉ

Nous avons vu plus haut la nécessité d'introduire les champs $T, K_0, S, {}^{(i)}O$, et comment leurs tenseurs de déformation intervenaient, aussi bien dans les inégalités d'énergie que dans les formules de commutation. Il reste à expliquer comment l'on contrôle ces divers tenseurs de déformations.

Bien que les traitements des situations géométriques I (ondes) et II (relativité) soient assez nettement différents, ils ont en commun suffisamment de traits caractéristiques

pour que nous puissions en donner une idée en exposant ici le seul cas I. Nous reviendrons sur le cas II à la section 7.

Nous définissons les coefficients de connexion (qui sont des tenseurs sur les sphères du feuilletage) par les équations

$$\begin{aligned}\chi_{ab} &= \langle D_a L, e_b \rangle, \underline{\chi}_{ab} = \langle D_a \underline{L}, e_b \rangle, \\ 2\xi_a &= \langle D_L L, e_a \rangle = 0, 2\underline{\xi}_a = \langle D_L \underline{L}, e_a \rangle, \\ 2\eta_a &= \langle D_L L, e_a \rangle, 2\underline{\eta}_a = \langle D_L \underline{L}, e_a \rangle.\end{aligned}$$

Il est clair que ces coefficients suffisent à exprimer les tenseurs de déformation des champs T , S , K_0 . Notons que χ et $\underline{\chi}$ sont les secondes formes pour le feuilletage en sphères, et à ce titre, symétriques. Définissons k comme la seconde forme pour les surfaces Σ_t ,

$$k_{ij} = -(1/2)\partial_t g_{ij}.$$

Les formules

$$\begin{aligned}\underline{\chi}_{ab} &= -\chi_{ab} - 2k_{ab}, \underline{\xi}_a = -\eta_a + k_{aN}, \\ \underline{\eta}_a &= -k_{aN}, \eta_a = b^{-1} \neq \nabla_a b + k_{aN}\end{aligned}$$

montrent, en supposant k suffisamment connue, qu'il suffit de contrôler χ, η, b .

Notons que ces coefficients sont à peu de chose près les composantes (non nulles) de $\nabla^2 u$:

$$\begin{aligned}\nabla^2 u_{ab} &= \langle D_a \nabla u, e_b \rangle = -\langle D_a b^{-1} L, e_b \rangle = -b^{-1} \chi_{ab}, \\ \nabla^2 u_{a\underline{L}} &= -\langle D_L b^{-1} L, e_a \rangle = -2b^{-1} \eta_a, \\ \nabla^2 u_{\underline{L}\underline{L}} &= -\langle D_L b^{-1} L, \underline{L} \rangle = 2e_3(b)b^{-2} - 2b^{-1} k_{NN}.\end{aligned}$$

En dérivant deux fois l'équation eikonale, on obtient d'autre part

$$D_L \nabla^2 u_{\alpha\beta} + b \nabla^2 u_\alpha^\mu \nabla^2 u_{\mu\beta} = b^{-1} R_{\alpha L \beta L}.$$

Pour en déduire les équations de transport sur χ, η, b , nous introduisons la partie sans trace de χ définie par

$$\hat{\chi} = \chi - (1/2)g(\text{tr}\chi).$$

Les *équations de transport* sont alors

$$\begin{aligned}L(b) &= -bk_{NN}, \\ L(\text{tr}\chi) + (1/2)(\text{tr}\chi)^2 &= -|\hat{\chi}|^2 - k_{NN}\text{tr}\chi - R_{44}, \\ \mathcal{D}_4 \hat{\chi}_{ab} + (1/2)(\text{tr}\chi)\hat{\chi}_{ab} &= -k_{NN}\hat{\chi}_{ab} - \hat{\alpha}_{ab}, \\ \mathcal{D}_4 \eta_a + (1/2)(\text{tr}\chi)\eta_a &= -(k_{Nb} + \eta_b)\hat{\chi}_{ab} - (1/2)(\text{tr}\chi)k_{aN} - (1/2)\beta_a,\end{aligned}$$

où α, β sont des composantes de R

$$\alpha_{ab} = R_{aLbL}, \text{tr}\alpha = R_{44}, \beta_a = R_{La\underline{L}L},$$

et \mathcal{D}_4 est la projection sur l'espace tangent aux sphères de D_4 . En principe, ces équations suffisent à saisir χ, η, b et leurs dérivées. Cependant, cela conduit à de mauvaises

estimations, à cause de pertes de dérivées. C'est pourquoi l'on introduit l'équation de Codazzi

$$\text{div}\hat{\chi}_a + \hat{\chi}_{ab}k_{bN} = (1/2)(\nabla_a \text{tr}\chi + k_{aN}\text{tr}\chi) - R_{b4ab}.$$

La clé est de penser cette équation comme un *système elliptique* sur les sphères.

La stratégie est alors la suivante :

- i) On remarque que l'équation de transport sur $\text{tr}\chi$ ne fait intervenir que la composante R_{44} du tenseur de Ricci.
- ii) Le point crucial est la structure particulière de $\text{Ric}(L, L)$,

$$\text{Ric}(L, L) = L(m) - (1/2)e_4^\mu e_4^\nu L_g(g_{\mu\nu}) + E.$$

Ici, m ne dépend que des dérivées premières de g , et E est une somme de termes quadratiques en les dérivées premières de g . Dans le contexte du problème non-linéaire de [14], on dispose sur $L_g(g_{\mu\nu})$ d'estimations meilleures que sur les composantes de R en général. En appliquant e_a à l'équation de transport, et en introduisant l'inconnue « normalisée » $y = \text{tr}\chi - 2/(t - u)$, on trouve finalement

$$\begin{aligned} \mathcal{D}_4 \nabla_a(y + m) + (3/2) \nabla_a(y + m) &= -\hat{\chi}_{ab} \nabla_b(y + m) + (y + m) \nabla_a(y + m) \\ &\quad - (y + m) \nabla_a \text{tr}\chi \hat{\chi} - \text{tr}\chi \nabla_a k_{NN} \\ &\quad - k_{NN} \nabla_a \text{tr}\chi + \nabla_a((1/2)e_4^\mu e_4^\nu L_g g_{\mu\nu} - E). \end{aligned}$$

- iii) En combinant cette équation à l'équation de Codazzi, on contrôle finalement les dérivées sur les sphères de $\text{tr}\chi$, $\hat{\chi}$ sans faire intervenir de dérivées de R . On a donc véritablement gagné une dérivée.

Pour contrôler $\nabla_a \eta$ ainsi que la dérivée manquante $\underline{L}\text{tr}\chi$, on ne dérive pas l'équation de transport sur η , car cela introduirait une dérivée de β . On pose

$$\mu = 2\text{div}\eta + 2\text{tr}\chi k_{NN} + 2|\eta|^2 - \text{tr}(\hat{\chi}\cdot\hat{\chi}) + \gamma,$$

où $\gamma = R_{431}^1 + R_{432}^2$. L'équation de transport sur μ s'écrit

$$L\mu + \mu \text{tr}\chi = \dots + \underline{L}(R_{44}),$$

où les points désignent des termes qui ne contiennent pas de dérivées de R . Compte tenu de la structure remarquable de R_{44} , le terme $\underline{L}(R_{44})$ ne pose plus de problème. Une fois μ estimée, on obtient les dérivées de η par le *système elliptique*

$$\begin{aligned} \text{div}\eta &= 1/2\mu - \text{tr}\chi k_{NN} - |\eta|^2 + (1/2)\text{tr}(\hat{\chi}\hat{\chi}) - 1/2\gamma, \\ \text{curl}\eta &= (1/2)(\hat{\chi}_{1c}\hat{\chi}_{2c} - \hat{\chi}_{2c}\hat{\chi}_{1c}) - (1/2)(R_{1432} - R_{2431}). \end{aligned}$$

Finalement,

$$\underline{L}\text{tr}\chi = \mu - k_{NN}\text{tr}\chi + (1/2)(\text{tr}\chi)^2$$

fournit l'information souhaitée.

7. COMPTAGE DES DÉRIVÉES ET ÉQUATIONS EIKONALES EN RELATIVITÉ

Nous n'avons jusqu'ici discuté que de la façon d'obtenir des estimations de solutions d'équations linéaires. Un problème fondamental dans le cas non-linéaire est la mise au point d'une hypothèse d'induction. Une telle hypothèse comporte en général l'indication des comportements d'un certain nombre n de dérivées de la solution. L'étude du problème linéaire dont les propriétés sont déduites de cette hypothèse doit permettre de retrouver (avec une amélioration) ces mêmes propriétés, en particulier de retrouver un contrôle du même nombre n de dérivées.

Nous allons brièvement expliquer ce « comptage des dérivées » dans le cas de la relativité. L'hypothèse d'induction consiste à supposer connues, dans une région \mathcal{K} bordée par une portion de Σ_0 , une portion de cône sortant $u = \nu_0$ et une portion de cône rentrant $\underline{u} = \underline{u}_*$, une métrique, ainsi que deux fonctions optiques u et \underline{u} réalisant un feuilletage dit « canonique », et vérifiant

$$\mathcal{O} \leq \epsilon_0, \mathcal{R} \leq \epsilon_0.$$

En simplifiant outrageusement, on peut dire que la quantité \mathcal{R} décrit les comportements de dérivées d'ordre (au plus) deux des composantes de R , tandis que \mathcal{O} décrit les comportements de dérivées d'ordre (au plus) trois des coefficients de connexion (notons que les dérivées d'ordre trois sont prises sur les sphères). Pour retrouver/améliorer l'hypothèse sur R , on utilise les équations de Bianchi que l'on dérive deux fois. Nous admettrons, ce qui est *très loin* d'être évident, que les informations sur \mathcal{O} suffisent pour maîtriser les commutations et les inégalités d'énergie correspondantes. Il faut ensuite retrouver l'hypothèse sur \mathcal{O} , et c'est ce point que nous allons un peu détailler, pour bien voir comment les fonctions optiques u et \underline{u} sont construites.

Dans la situation géométrique II, on définit des coefficients de connexion, tout à fait comme au point 6 :

$$\begin{aligned} \chi_{ab} &= \langle D_a e_4, e_b \rangle, \underline{\chi}_{ab} = \langle D_a e_3, e_b \rangle, 2\zeta_a = \langle D_a e_4, e_3 \rangle, \\ 2\xi_a &= \langle D_{e_4} e_4, e_a \rangle = 0, \underline{\xi}_a = \langle D_{e_3} e_3, e_a \rangle = 0, \\ \eta_a &= (1/2) \langle D_{e_3} e_4, e_a \rangle = \zeta_a + \nabla_a \log \Omega, \\ \underline{\eta}_a &= (1/2) \langle D_{e_4} e_3, e_a \rangle = -\zeta_a + \nabla_a \log \Omega, \\ 4\omega &= \langle D_{e_4} e_4, e_3 \rangle = -2D_4 \log \Omega, 4\underline{\omega} = \langle D_{e_3} e_3, e_4 \rangle = -2D_3 \log \Omega. \end{aligned}$$

La grande différence est que les composantes de Ric sont nulles. Les *équations de transport naturelles*, où l'on a scindé χ et $\underline{\chi}$, sont comme plus haut

$$\begin{aligned} D_4 tr \chi + (1/2)(tr \chi)^2 - (D_4 \log \Omega) tr \chi + |\hat{\chi}|^2 &= 0, \\ D_4 \hat{\chi} + tr \chi \hat{\chi} - (D_4 \log \Omega) \hat{\chi} &= -\alpha, \\ \not{D}_4 \zeta + 2\chi \zeta + \not{D}_4 \log \Omega &= -\beta, \\ (7.1) \quad D_3 tr \underline{\chi} + (1/2)(tr \underline{\chi})^2 - (D_3 \log \Omega) tr \underline{\chi} + |\hat{\underline{\chi}}|^2 &= 0, \end{aligned}$$

$$D_3 \hat{\chi} + \text{tr} \chi \hat{\chi} - (D_3 \log \Omega) \hat{\chi} = -\alpha.$$

Les équations « anti-naturelles », qui expriment $D_3 \text{tr} \chi$ ou $D_4 \text{tr} \chi$ en fonction de $\text{div} \zeta$, servent en fait à calculer $\text{div} \zeta$, comme on l’a vu plus haut.

Les *équations elliptiques* sont

$$\begin{aligned} \text{div} \hat{\chi} + \hat{\chi} \zeta &= (1/2)(\nabla \text{tr} \chi + \zeta \text{tr} \chi) - \beta, \\ \text{div} \hat{\chi} - \hat{\chi} \zeta &= (1/2)(\nabla \text{tr} \chi - \zeta \text{tr} \chi) + \beta, \\ \text{curl} \zeta &= -(1/2) \hat{\chi} \wedge \hat{\chi} + \sigma, \\ \text{div} \zeta &= \dots, \end{aligned}$$

où les points indiquent des quantités préalablement estimées.

C’est la nécessité de « boucler » l’hypothèse d’induction qui va conduire au concept fondamental de « feuilletage canonique ». Supposons en effet donnée sur Σ_0 une fonction w dont les surfaces de niveau forment un feuilletage en sphères raisonnable de Σ_0 , et qui servira de donnée initiale pour \underline{u} . Le contrôle de trois dérivées de $\underline{\chi}$ via (7.1) nécessite le contrôle de trois dérivées de θ , la seconde forme du feuilletage en sphères sur Σ_0 . En notant N le vecteur unitaire sortant et $a = |\nabla w|^{-1}$, on a

$$\nabla_N \text{tr} \theta + (1/2)(\text{tr} \theta)^2 = -(\Delta \log a + \rho) + [-|\nabla \log a|^2 + |\hat{\theta}|^2 + g(k)],$$

où $g(k)$ est une somme de carrés de composantes de k . Comme on n’a besoin en fait que de trois dérivées ∇ , pour ne pas avoir à dériver ρ , on va demander, et c’est là la définition du feuilletage canonique sur Σ_0 ,

$$\nabla \log a + \rho = \bar{\rho}, \overline{\log a} = 0,$$

où la barre supérieure dénote la moyenne sur la sphère.

Pour choisir la donnée \underline{u}_* de u sur \underline{C}_* , on introduit comme au point 6 la fonction μ

$$\mu = - \text{div} \eta + (1/2) \hat{\chi} \hat{\chi} - \rho.$$

Le point important est que l’équation de transport sur μ est agréable. Pour contrôler les dérivées ∇ , on choisit d’imposer $\mu = \bar{\mu}$ sur \underline{C}_* . Sans entrer dans les détails, disons qu’il est possible de choisir u de cette façon, que les auteurs appellent « canonique ».

8. VERS LA CONJECTURE H^2 ?

Pour terminer ce court exposé, nous voudrions effleurer les travaux en cours iii) de Klainerman et Rodniansky [15]-[20]. Il ne nous est en aucune façon possible d’entrer ici dans les détails. Ce qui importe pour notre propos est d’illustrer l’*unité* des méthodes d’approche.

Il s’agit du problème de type B : résoudre les équations d’Einstein pour des données de Cauchy $g \in H^2, k \in H^1$. Dans le contexte que nous avons expliqué plus haut, un des points clé est de pouvoir contrôler les cônes sortants, surfaces de niveau de la fonction optique u . Rappelons (cf. [13]) que $\int_S \text{tr} \chi$ mesure le taux de variation du volume de S

dans la direction de e_4 : pour éviter l'apparition d'un phénomène de type « caustique », il est donc nécessaire (entre autres choses) de contrôler $tr\chi$ dans L^∞ . En notant ici $L = -\nabla u$ et $\chi(X, Y) = \langle D_X L, Y \rangle$, on a l'équation de transport

$$L(tr\chi) + (1/2)(tr\chi)^2 = -|\hat{\chi}|^2.$$

Il faut donc contrôler $\int_\Gamma |\hat{\chi}|^2$ sur les géodésiques Γ qui tissent un cône sortant C .

On a vu (au paragraphe 7) que la bonne façon d'estimer $\hat{\chi}$ est l'équation de Codazzi

$$div\hat{\chi} = -\beta + (1/2) \nabla tr\chi + (1/2)tr\chi\zeta - \zeta \cdot \hat{\chi}.$$

En supposant que les deux derniers termes ne causent pas de problèmes, et en notant D^{-1} l'opérateur pseudo-différentiel (sur la sphère!) d'ordre -1 qui résout le système ci-dessus, il reste à considérer

$$I_1 = \int_\Gamma |D^{-1}\beta|^2, I_2 = \int_\Gamma |D^{-1} \nabla tr\chi|^2.$$

Nous supposons estimée la quantité R_0 , qui est la somme des normes des « bonnes » composantes de R (dont β) dans $L^2(C)$. Notons que Γ est de codimension deux dans C (qui est de dimension trois), en sorte que le théorème de trace de $H^s(C)$ dans $H^{s-1}(\Gamma)$ qu'on aimerait utiliser... est faux, car ici $s = 1$ et non $s > 1$. De façon tout à fait analogue à la situation de [14] décrite en 6 où la structure particulière de R_{44} permettait de gagner une dérivée, c'est ici la *structure particulière de β* qui va nous aider. Les équations de Bianchi, écrites dans le repère habituel, donnent en particulier (cf. [13])

$$\not{d}iv\beta = D_4\rho + (3/2)\rho tr\chi + (1/2)\hat{\chi} \cdot \alpha - \zeta\beta - 2\underline{\eta}\beta,$$

$$curl\beta = -D_4\sigma - (3/2)tr\chi\sigma + (1/2)\hat{\chi}^* \alpha - \zeta^*\beta - 2\underline{\eta}^*\beta.$$

En abrégé, nous écrivons $\beta = D^{-1}(L(\rho), L(\sigma))$, soit, en ignorant le commutateur

$$D^{-1}\beta = \nabla_L Q + \dots, Q = D^{-2}(\rho, \sigma).$$

L'intégrale I_1 sera majorée par la norme de $Q|_\Gamma$ dans H^1 , qui est estimée par

$$|Q|_{H^2(C)} \leq C|(\rho, \sigma)|_{L^2(C)} \leq CR_0.$$

Si l'on a une chance de majorer $|tr\chi|_{L^\infty}$ à l'aide de l'équation de transport, c'est en majorant I_2 à l'aide de cette même norme. Mais $D^{-1} \nabla$, opérateur (pseudo-différentiel) d'ordre zéro sur la sphère, n'opère pas dans L^∞ ... Nous arrêtons ici, en plein « suspens », le récit de cette passionnante aventure, en espérant avoir donné au lecteur quelques clés pour aborder la lecture des travaux de Klainerman et Rodniansky [18]-[20].

BIBLIOGRAPHIE

Nous nous sommes limités ici aux travaux ayant un rapport immédiat avec l'approche que nous avons choisie. Une bibliographie plus vaste concernant les équations hyperboliques non-linéaires se trouve dans [9], tandis que les lecteurs plus spécialement intéressés par la relativité consulteront [13].

- [1] Bahouri H. et Chemin J-Y. – *Équations d'ondes quasi-linéaires et estimations de Strichartz*, Amer. J. Math. 121 (1999), 1337-1377.
- [2] Bahouri H. et Chemin J-Y. – *Équations d'ondes quasi-linéaires et effet dispersif*, IMRN 21 (1999), 1141-1178.
- [3] Bourguignon J-P. – *Stabilité par déformation non-linéaire de la métrique de Minkowski*, Sémin. Bourbaki (1990/91), exp. n° 740, Astérisque 201-202-203 (1991), 321-358.
- [4] Chemin J-Y. – *Explosion géométrique pour certaines équations d'ondes non linéaires*, Sémin. Bourbaki (1998/99), exp. n° 850, Astérisque 266 (2000), 7-20.
- [5] Christodoulou D. – *The Stability of Minkowski Spacetime*, Proc. Intern. Cong. Math. Kyoto 1990 (1990), 1114-1121.
- [6] Christodoulou D. et Klainerman S. – *Asymptotic properties of linear field equations in Minkowski space*, Comm. Pure Appl. Math. XLIII (1990), 137-199.
- [7] Christodoulou D. et Klainerman S. – *The global nonlinear stability of the Minkowski space*, Princeton Math. Series 41 (1993).
- [8] Hawking S. W. et Ellis G. F. R. – *The Large Scale Structure of Spacetime*, Cambridge Mon. Math. Physics (1973).
- [9] Hörmander L. – *Lectures on Nonlinear Hyperbolic Equations*, Math. Appl. 26, Springer-Verlag (1987).
- [10] Klainerman S. – *Remarks on the global Sobolev inequalities in Minkowski Space*, Comm. Pure Appl. Math. 40 (1987), 111-117.
- [11] Klainerman S. – *Uniform decay estimates and the Lorentz invariance of the classical wave equation*, Comm. Pure Appl. Math. 38 (1985), 321-332.
- [12] Klainerman S. – *A commuting vector field approach to Strichartz type inequalities and applications to quasilinear wave equations*, IMRN 5 (2001), 221-274.
- [13] Klainerman S. et Nicolò F. – *The evolution problem in general relativity*, Prog. Math. Physics 25, Birkhäuser (2003).
- [14] Klainerman S. et Rodniansky I. – *Improved local well posedness for quasilinear wave equations in dimension three*, to appear in Duke Math. J. (2002).
- [15] Klainerman S. et Rodniansky I. – *Rough solutions of the Einstein vacuum equations*, Preprint (2003).
- [16] Klainerman S. et Rodniansky I. – *The causal structure of microlocalized, rough, Einstein metrics*, Preprint (2003).
- [17] Klainerman S. et Rodniansky I. – *Ricci defects of microlocalized Einstein metrics*, Preprint (2003).

- [18] Klainerman S. et Rodniansky I. – *Causal geometry of Einstein-vacuum spacetimes with finite curvature flux*, Preprint (2003).
- [19] Klainerman S. et Rodniansky I. – *A geometric theory of Littlewood-Paley theory*, preprint (2003).
- [20] Klainerman S. et Rodniansky I. – *Sharp trace theorems for null hypersurfaces on Einstein metrics with finite curvature flux*, Preprint (2003).
- [21] Lindblad H. et Rodniansky I. – *Global Existence for the Einstein vacuum equations in wave coordinates*, Preprint (2003).

Serge ALINHAC

Université Paris XI

Département de Mathématiques

Bâtiment 425

F-91405 ORSAY Cedex

E-mail : `Serge.Alinhac@math.u-psud.fr`

PARAMÉTRISATION DE STRUCTURES ALGÈBRIQUES
ET DENSITÉ DE DISCRIMINANTS

[d'après M. Bhargava]

par Karim BELABAS

Gauss publie ses *Disquisitiones Arithmeticae* en 1801. La moitié du traité est consacrée aux formes quadratiques binaires $f(x, y) = ax^2 + bxy + cy^2$, $a, b, c \in \mathbb{Z}$, notées (a, b, c) , de discriminant $D = b^2 - 4ac$ ⁽¹⁾. Intéressé par les valeurs représentées par ces formes, c'est-à-dire par $\{f(x, y) : x, y \in \mathbb{Z}\}$, Gauss constate que l'action du groupe linéaire $SL_2(\mathbb{Z})$ par changement de variables

$$(1) \quad (\gamma \cdot f)(x, y) = f((x, y)\gamma),$$

permet de ranger les formes en classes, les formes d'une orbite représentant les mêmes entiers. Le discriminant est constant sur une orbite et le nombre d'orbites de discriminant fixé est fini. Enfin,

« *sujet très important et dont personne ne s'est encore occupé* » [§234],

il munit les orbites primitives, telles que $\text{pgcd}(a, b, c) = 1$, d'une structure de groupe, compatible avec les valeurs représentées. L'idée est de généraliser l'identité de Brahmagupta

$$(x^2 + Dy^2)(z^2 + Dt^2) = X^2 + DY^2, \quad \text{pour } X = xz + Dyt, Y = xt - yz,$$

qu'on n'expliquait pas encore par la multiplicativité de la norme dans $\mathbb{Z}[\sqrt{D}]$. Gauss écrit en complète généralité

$$(a_1x^2 + b_1xy + c_1y^2)(a_2z^2 + b_2zt + c_2t^2) = AX^2 + BXY + CY^2$$

dans $\mathbb{Z}[x, y, z, t]$, où X et Y sont des fonctions linéaires de (xz, xt, yz, yt) données par une transformation primitive (les mineurs maximaux de la matrice 2×4 associée sont premiers entre eux) et où tous les coefficients sont indéterminés et entiers. Puis il résout tranquillement le système. Il découvre ainsi toutes les lois de composition possibles : il n'y en a essentiellement qu'une ⁽²⁾, qui s'exprime plus agréablement pour les formes

⁽¹⁾Gauss considère les formes dont la forme polaire bilinéaire est à *valeurs* entières, et le coefficient de xy est toujours pair. Il utilise donc le symbole (a, b, c) là où nous écrivons $(a, 2b, c)$ et définit son discriminant par $b^2 - ac$. Nous traduisons dans les notations modernes.

⁽²⁾Gauss exclut les formes de discriminant nul et impose une transformation primitive, ainsi qu'un choix de signe (distinguant ainsi composition directe et indirecte).

primitives de même discriminant. Voici la formulation qu'en donne Dirichlet⁽³⁾ : pour deux formes primitives de discriminant $D \neq 0$, vérifiant $a_1 a_2 \neq 0$, on pose

$$(2) \quad (a_1, b_1, *) \times (a_2, b_2, *) = (A, B, *),$$

où $n = \text{pgcd}(a_1, a_2, (b_1 + b_2)/2)$, $A = a_1 a_2 / n^2$, B est solution du système de congruences

$$\begin{aligned} B &\equiv b_1 \pmod{2a_1/n} \\ B &\equiv b_2 \pmod{2a_2/n} \\ B^2 &\equiv D \pmod{4a_1 a_2 / n}, \end{aligned}$$

et le troisième coefficient, déterminé par les deux premiers et le discriminant, est omis. Conscientieux, Gauss vérifie que l'opération passe au quotient et qu'elle est associative. En langage moderne, il définit la multiplication des idéaux dans un anneau quadratique S et identifie le groupe des classes de S -idéaux projectifs (*i.e.* inversibles). Cette caractérisation est toujours algorithmiquement utile et permet d'autre part d'estimer de nombreuses densités liées à ces groupes de classes quand le discriminant varie.

Dans sa thèse, Bhargava entreprend une vaste recherche de « lois de composition » arithmétiques, guidé par une série d'heuristiques et la classification des espaces vectoriels préhomogènes (voir §3). Il considère un groupe algébrique G , une représentation naturelle V , choisit tels que l'action de $G_{\mathbb{Z}}$ sur $V_{\mathbb{Z}}$ n'ait qu'un seul invariant, baptisé discriminant, puis montre que les orbites $V_{\mathbb{Z}}/G_{\mathbb{Z}}$ paramètrent les paires $(R, *)$, où R est une classe d'isomorphisme d'anneaux de nombres de petit degré (voir §1.1) et $*$ désigne des structures supplémentaires, en général des R -modules. Ce sont les structures algébriques du titre de l'exposé. Le discriminant usuel de R coïncide avec celui de l'orbite de $V_{\mathbb{Z}}/G_{\mathbb{Z}}$ associée. Bhargava obtient une dizaine de tels exemples, très explicites, et d'autres encore conjecturaux.

D'une part, il munit un sous-ensemble « projectif » de $V_{\mathbb{Z}}/G_{\mathbb{Z}}$ d'une loi de groupe intrinsèque et élégante, qui se réinterprète en termes du groupe des classes $\text{Cl}(R)$. D'autre part, il peut énumérer les orbites par discriminant croissant, algorithmiquement ou asymptotiquement quand le discriminant tend vers l'infini. En particulier, en oubliant les structures $*$ et en se restreignant aux anneaux R intègres maximaux, Bhargava obtient de nouveaux résultats sur les densités de discriminants de corps de nombres quartiques et quintiques. Il convient toutefois de rester prudent pour cette dernière application : seul le cas des corps quartiques totalement réels est complètement rédigé à ce jour. Ces résultats restent mystérieux : la vision est unifiée et élégante, mais chaque démonstration est unique quoique suivant un motif commun dans l'esprit de la théorie des invariants classique, et laisse une part décisive au calcul formel explicite. Tout comme la démonstration de Gauss.

Après quelques définitions, nous détaillons sur l'exemple de Gauss la technique de comptage employée par Bhargava en dimension supérieure. Nous décrivons ensuite les techniques alternatives utilisant les fonctions zêta de Sato-Shintani, plus générales mais

⁽³⁾Pour l'essentiel. Dirichlet-Dedekind se restreint au cas $n = 1$ des formes « unifiées » [23, Supp. X].

aussi plus sophistiquées, qui fournissent de nombreux résultats de densités, en particulier pour les discriminants des corps quadratiques et cubiques sur une base arbitraire, et proposent un vaste programme susceptible d’aboutir à d’autres résultats de ce type, mais sans être pour l’instant en mesure de fournir les résultats annoncés par Bhargava sur \mathbb{Q} . Elles inspirent néanmoins ses paramétrisations et lois de composition que nous présentons ensuite. Nous énonçons finalement les résultats de densité obtenus ainsi que les conjectures qu’ils corroborent.

Je voudrais remercier A. Chambert-Loir, H. Cohen, O. Gabber, H. Gangl, J. Klüners, B. Perrin-Riou et J.-P. Serre pour leurs suggestions.

1. DÉFINITIONS

1.1. Anneaux de nombres

On appelle *anneau de nombres de degré n* un anneau R (commutatif, associatif, unitaire) qui est un \mathbb{Z} -module libre de rang n . On dit que R est un *ordre* s’il est intègre, auquel cas son corps des fractions est un corps de nombres. Dans cet exposé, $2 \leq n \leq 5$; conformément à une respectable tradition, nous parlerons d’anneaux quadratiques, cubiques, quartiques et quintiques pour $n = 2, 3, 4, 5$ respectivement. La trace $\text{Tr} : R \rightarrow \mathbb{Z}$ assigne à $\alpha \in R$ la trace de la multiplication par α . Elle permet de définir le *discriminant* $\text{Disc}(R)$ comme $\det(\text{Tr}(\alpha_i \alpha_j))$, où $(\alpha_i)_{1 \leq i \leq n}$ est une \mathbb{Z} -base arbitraire de R . C’est un entier relatif congru à 0 ou 1 modulo 4.

Un anneau de nombres est dit maximal s’il n’est pas strictement inclus dans un anneau de même degré. En particulier un ordre maximal est l’anneau des entiers de son corps des fractions, *i.e.* il est intégralement clos. La maximalité est une propriété locale qui se voit sur les $R \otimes_{\mathbb{Z}} \mathbb{Z}_p$.

1.2. Anneaux quadratiques

Soit $D \equiv 0, 1 \pmod{4}$ un entier relatif. À isomorphisme près, il existe un unique anneau quadratique de discriminant D , à savoir $S(D) := \mathbb{Z}[X]/(X^2 - DX + (D^2 - D)/4)$. Une *orientation* sur $S = S(D)$ est un choix d’isomorphisme $\pi : S/\mathbb{Z} \rightarrow \mathbb{Z}$, ce qui revient à choisir une racine carrée de D , ou encore une \mathbb{Z} -base $\alpha \wedge \beta$ de $\Lambda^2 S \simeq \mathbb{Z}$. Une base $\langle x, y \rangle$ d’un sous-module de rang 2 de l’algèbre $K = S \otimes_{\mathbb{Z}} \mathbb{Q}$ est *orientée positivement* si et seulement si $x \wedge y = c \cdot \alpha \wedge \beta$, avec $c > 0$.

Un anneau quadratique orienté n’ayant pas d’automorphismes non triviaux, deux tels anneaux de même discriminant sont canoniquement isomorphes. Ainsi, l’ensemble des entiers $D \equiv 0, 1 \pmod{4}$ paramètre les classes d’isomorphismes d’anneaux quadratiques orientés. Un *idéal orienté* de S est un couple (I, ε) , où $I \subset K$ est un idéal fractionnaire de S et $\varepsilon \in \{\pm 1\}$. (Alternativement, on peut définir (I, ε) par une \mathbb{Z} -base de I d’orientation donnée par le signe de ε .) La norme d’un idéal orienté (I, ε) est

$\varepsilon |L/S| / |L/I| \in \mathbb{Z}$, où L est un sous- \mathbb{Z} -module de rang 2 de K arbitraire contenant S et I .

Les idéaux orientés forment un monoïde pour la multiplication composante par composante et tout $\kappa \in K^*$ définit un idéal orienté principal $((\kappa), \text{sgn}(N_{K/\mathbb{Q}}\kappa))$. Les idéaux orientés inversibles forment un groupe, dont les idéaux principaux inversibles forment un sous-groupe. Le quotient, noté $\text{Cl}(D)^+$, est le *groupe des classes orientées*, de discriminant D . Si $D > 0$, c'est le groupe des classes au sens restreint. Si $D < 0$, $\text{Cl}^+(D) = \{\pm 1\} \times \text{Cl}(D)$, où $\text{Cl}(D)$ est le groupe des classes usuel.

1.3. Formes

Une forme k -ique n -aire est un polynôme homogène de degré k en n variables ou, par abus de langage, le polynôme nul. Par exemple, une forme quadratique binaire est un polynôme $f(x, y) = ax^2 + bxy + cy^2$, pour certains coefficients a, b, c , éventuellement tous nuls. On notera (a_0, a_1, \dots, a_n) la forme binaire $\sum_i a_i x^{n-i} y^i$ de degré n , quand le contexte ne portera pas à confusion. On note $\text{Sym}^k \mathbb{Z}^n$ l'ensemble des formes $f : \mathbb{Z}^n \rightarrow \mathbb{Z}$ satisfaisant $f(x) = F(x, \dots, x)$ pour une forme polaire F k -linéaire symétrique de $(\mathbb{Z}^n)^k \rightarrow \mathbb{Z}$, et $(\text{Sym}^k \mathbb{Z}^n)^*$ l'ensemble des formes k -iques n -aires. Par exemple $(\text{Sym}^2 \mathbb{Z}^2)^*$ est l'ensemble des f comme ci-dessus, avec $(a, b, c) \in \mathbb{Z}^3$. On a $f \in \text{Sym}^2 \mathbb{Z}^2$ si et seulement si b est pair ; plus généralement, les monômes de $\text{Sym}^n \mathbb{Z}^k$ sont pondérés de coefficients multinomiaux. Finalement, soit $\Lambda^k \mathbb{Z}^n$ l'espace des fonctions multilinéaires $(\mathbb{Z}^n)^k \rightarrow \mathbb{Z}$ alternées.

2. DOMAINES FONDAMENTAUX : UN EXEMPLE CLASSIQUE

2.1. Paramétrisation

Le prototype des résultats que l'on veut obtenir remonte à Gauss, au langage près.

THÉORÈME 2.1. — *Il existe une bijection canonique entre les deux ensembles suivants :*

- les classes d'isomorphismes de paires (S, I) , où S est un anneau quadratique orienté de discriminant non nul, et I une classe d'idéaux orientés de S ,
- les classes de formes quadratiques binaires entières, modulo l'action de $\text{SL}_2(\mathbb{Z})$.

Cette bijection préserve le discriminant et associe une classe de formes quadratiques primitives à une classe de S -idéaux inversibles. Muni de la composition des formes quadratiques, l'ensemble des classes de formes primitives de discriminant $D \neq 0$ est un groupe, isomorphe au groupe des classes orientées $\text{Cl}^+(D)$.

Dans ce théorème, les formes quadratiques entières sont les $(a, b, c) \in (\text{Sym}^2 \mathbb{Z}^2)^* =: V_{\mathbb{Z}}$, une forme est primitive si le pgcd de ses coefficients est 1, et l'action (à droite) de SL_2 est donnée par le changement de variable $(g \cdot F)(x, y) = F((x, y)g)$. Le discriminant $\text{Disc}(F) = b^2 - 4ac$ est un invariant de cette action, et il engendre l'algèbre des invariants sur \mathbb{C} . Par abus de langage, on dira que l'action a un unique invariant.

2.2. Domaine fondamental

Les orbites sous $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ ont donc une signification arithmétique et les représentants des classes sont les points entiers $V_{\mathbb{Z}}$ de l'espace affine $V = \mathbb{A}^3$, et non pas un ensemble « mince », comme une sous-variété de codimension ≥ 1 par exemple. Cette représentation s'utilise algorithmiquement pour calculer ou manipuler concrètement le groupe des classes d'idéaux de corps quadratiques, dans les méthodes développées par Shanks [48] après Gauss (voir [8, 10] pour les détails algorithmiques), mais elle permet aussi de démontrer des résultats de densité, par exemple

THÉORÈME 2.2 (Lipschitz [36], conjecturé par Gauss). — *Quand $X \rightarrow +\infty$, on a*

$$\sum_{0 < -D < X} |V_{\mathbb{Z}}/\Gamma| \sim \frac{\pi}{9} X^{3/2}.$$

(On peut être plus précis, voir [9].) Le principe est simple : on identifie les orbites de discriminant inférieur à X aux points à coordonnées entières du domaine fondamental de Gauss, dont l'adhérence est $C_X \cup (-C_X)$ où

$$C_X = \{(a, b, c) \in \mathbb{R}^3 : |b| \leq a \leq c, 4ac - b^2 \leq X\},$$

et qui s'obtient en imposant qu'une racine de $ax^2 + bx + c = 0$ soit dans le domaine fondamental standard de l'action de Γ sur le demi-plan supérieur. Leur nombre est approché par le volume de C_X .

THÉORÈME 2.3 (« principe de Lipschitz », Davenport [20]). — *Soit $C \subset \mathbb{R}^n$ un ensemble semi-algébrique compact, de volume $\mathrm{Vol}(C)$, et soit $N(C) = |C \cap \mathbb{Z}^n|$. On note $R(C)$ le maximum des volumes des projections de C sur les variétés linéaires d'équations $\{x_i = 0, i \in I\}$, où I parcourt les sous-ensembles non-vides de $\{1, \dots, n\}$. Alors*

$$N(C) = \mathrm{Vol}(C) + O(1 + R(C)).$$

La constante implicite est effective et ne dépend que de la dimension n , du nombre et du degré des équations définissant C .

Nous avons négligé deux points techniques : d'abord les stabilisateurs

$$\Gamma_x := \{\gamma \in \Gamma, \gamma x = x\}$$

ont pour cardinaux 1, 2, 4, ou 6. Comme dans toute formule de masse, il serait plus habile de compter une classe x avec poids $1/|\Gamma_x|$ plutôt que d'exclure arbitrairement certains points du bord du domaine fondamental. Ici ces derniers sont peu nombreux et absorbés dans le terme d'erreur. Ensuite, nous n'avons pas à séparer le bon grain de l'ivraie : il n'y a pas de points de discriminant nul dans l'intérieur de C_X et, dans le cas de discriminants négatifs, il n'y a pas non plus lieu d'isoler les anneaux isomorphes à \mathbb{Z}^2 , dont le discriminant est un carré parfait.

2.3. Densités locales et crible

Si on se restreint aux classes primitives, la formule d'inversion de Moebius donne :

COROLLAIRE 2.4. — *Quand $X \rightarrow +\infty$, on a*

$$\sum_{0 < -D < X} |\text{Cl}^+(D)| \sim \frac{\pi}{9\zeta(3)} X^{3/2}, \quad \text{soit} \quad \sum_{0 < -D < X} |\text{Cl}(D)| \sim \frac{\pi}{18\zeta(3)} X^{3/2}.$$

La condition imposée est de nature locale (une condition p -adique pour chaque premier p), se traduisant par l'apparition du facteur eulérien $\prod(1 - p^{-3}) = 1/\zeta(3)$, et admet un grand nombre de variantes naturelles. On peut par exemple se limiter aux D qui sont discriminants d'un corps quadratique, ou *discriminants fondamentaux*, ce qui permet d'obtenir une somme sur les corps quadratiques, en fait une somme sur leurs ordres maximaux. Ceci se traduit par l'élimination des points (a, b, c) satisfaisant l'une des congruences

$$(*_p) \quad \begin{cases} \text{Disc}(a, b, c) \equiv 0 \pmod{p^2} & \text{pour } p \text{ premier impair,} \\ \text{Disc}(a, b, c) \equiv 0, 4 \pmod{2^4} & \text{pour } p = 2. \end{cases}$$

(Cette condition ne dépend que de (a, b, c) modulo p^2 , y compris quand $p = 2$.) La formule de Moebius prend alors la forme du crible d'inclusion-exclusion et on retrouve un cas particulier d'un résultat de Goldfeld-Hoffstein [27], qu'ils obtenaient en utilisant des séries d'Eisenstein de poids demi-entier.

THÉORÈME 2.5. — *Si k parcourt les corps quadratiques, on a*

$$\sum_{0 < -\text{Disc } k < X} |\text{Cl}(\text{Disc } k)| \sim \prod_p (1 - p^{-2} - p^{-3} + p^{-4}) \cdot \frac{\pi}{18} X^{3/2}.$$

Preuve. — Indiquons les deux ingrédients nécessaires pour résoudre l'exercice : pour un entier sans facteur carré q , soit $N_q(C_X)$ le nombre des points de C_X vérifiant $(*_p)$ pour tout $p \mid q$. Alors

$$(3) \quad N_q(C_X) = \nu(q)N(C_X) + R_q(C_X)$$

pour un reste R_q effectif venant du principe de Lipschitz et une fonction (de densité) multiplicative

$$\begin{aligned} \nu(q) &= \frac{1}{(q^2)^3} \# \{ (a, b, c) \in (\mathbb{Z}/q^2\mathbb{Z})^3 \text{ satisfait } (*_p) \text{ pour tout } p \mid q \} \\ &= \prod_{p \mid q} (p^{-2} + p^{-3} - p^{-4}) = O(q^{-2} \prod_{p \mid q} (1 + 1/p)) \end{aligned}$$

obtenue par un dénombrement élémentaire (voir par exemple [1, §4]). Si q est grand, le reste $R_q(X)$ domine le terme principal et on remplace (3) par une majoration uniforme

$$(4) \quad N_q(C_X) = O(N(C_X)q^{-2} \prod_{p \mid q} (1 + 1/p)),$$

obtenue en majorant

$$|\text{Cl}(q^2D)| / |\text{Cl}(D)| = O\left(q \prod_{p|q} (1 + 1/p)\right),$$

voir par exemple [16, §7.D], avec une complication technique pour gérer les formes non primitives. Pour tout paramètre $Q > 0$, la somme restreinte aux D fondamentaux vaut

$$\begin{aligned} \sum_{q \geq 1} \mu(q) N_q(C_X) &= \sum_{q < Q} \mu(q) (\nu(q) N(C_X) + R_q) + \sum_{q \geq Q} O(N_q(C_X)) \\ &= N(C_X) \prod_p (1 - \nu(p)) + O\left(\sum_{q < Q} R_q(C) + \sum_{q \geq Q} \nu(q) N(C_X) + N_q(C_X)\right), \end{aligned}$$

où μ est la fonction de Moebius. Il ne reste plus qu'à optimiser Q en fonction de X pour minimiser les termes d'erreur. Les conditions locales $(*_p)$ ne raréfient pas trop l'ensemble des points, ce qui se traduit par la convergence de $\sum_q N_q(C_X)$, $\sum_q \nu(q)$ et $\prod_p (1 - \nu(p))$. \square

Remarque 2.6. — Par cette méthode du « domaine fondamental », on obtient naturellement des termes d'erreur, que nous avons omis ci-dessus. Sous la forme générale du principe de Lipschitz, ils sont en effet loin d'être optimaux. Pour s'en convaincre, considérons le problème des points entiers du disque : on obtient comme Gauss

$$\#\{(a, b) \in \mathbb{Z}^2, a^2 + b^2 \leq X\} = \pi X^2 + O(X).$$

L'argument revient à considérer la formule de Poisson

$$\sum_{n \in \mathbb{Z}} f(n) = \sum_{n \in \mathbb{Z}} \hat{f}(n),$$

avec $f(n) = \sqrt{X^2 - n^2} \times \mathbf{1}_{|n| \leq X}$, pour ne retenir du membre de droite que le terme $\hat{f}(0)$. Il est naturel qu'un lissage convenable et une analyse harmonique plus fine faisant intervenir des majorations de sommes d'exponentielles permettent des progrès (reste en $O(X^{131/208})$ actuellement pour le problème du cercle, voir le survol de Huxley [29]).

Remarque 2.7. — Cet exemple est pédagogique. Tous ces résultats s'obtiennent en quelques lignes avec de meilleurs termes d'erreur par un argument de Siegel [51] fondé sur la formule du nombre de classes de Dirichlet.

$$(5) \quad L(1, \chi_D) = \sum_{n \geq 1} \frac{\chi_D(n)}{n} = \frac{|\text{Cl}(D)|}{w(D) \sqrt{|D|}} \times \begin{cases} 2\pi & \text{si } D < 0 \\ 4 \log \varepsilon(D) & \text{si } D > 0 \text{ non carré,} \end{cases}$$

où χ_D est le caractère de Kronecker modulo $D \equiv 0, 1 \pmod{4}$, où $\varepsilon(D) > 1$ est l'unité fondamentale de l'ordre quadratique de discriminant D et $w(D)$ le nombre de racines de l'unité, soit $w(D) = 2$ pour $D \neq -4, -6$. En sommant les $L(1, \chi_D)$ au lieu des $|\text{Cl}(D)|$, on introduit essentiellement un poids $|D|^{-1/2}$, que l'on supprime par intégration par partie. Il suffit d'invertir les sommations et de majorer non trivialement $\sum_{D < X} \chi_D(n)$

pour n non carré, par réciprocité quadratique et Pólya-Vinogradov par exemple. Comme au Corollaire 2.4, mais à l'envers, la formule de Moebius donne les formes non primitives.

Avantage supplémentaire, on obtient des estimations analogues pour les discriminants D positifs, non carrés parfaits. Suite à la modification de (5), on remplace $|\text{Cl}(D)|$ par $|\text{Cl}(D)| \log \varepsilon(D)$ dans la somme, et l'équivalent est multiplié par $\pi/2$ (on élimine les carrés dans le terme d'erreur). Siegel [51] en donne une interprétation en termes de domaine fondamental mais, suite à la présence d'un nombre infini d'unités, les stabilisateurs ne sont plus finis : il faut introduire une densité analogue aux $\mu(x)$ du §3.2, formule (9).

3. ESPACES VECTORIELS PRÉHOMOGÈNES

3.1. Motivation

Le Théorème 2.2 estime une formule de masse asymptotique de type

$$(6) \quad \sum_{\substack{x \in L/\Gamma \\ 0 < |P(x)| < X}} \mu(x)$$

où L est un réseau sur lequel agit un groupe linéaire discret Γ , dont les stabilisateurs Γ_x sont supposés finis, $\mu(x) = |\Gamma_x|^{-1}$, et P est un polynôme Γ -invariant. Nous avons évoqué la « méthode du domaine fondamental ». Une autre méthode classique étudie les propriétés analytiques (prolongement analytique, pôles et résidus, croissance dans les bandes verticales) de la série de Dirichlet associée

$$(7) \quad \sum_{x \in L'/\Gamma} \mu(x) |P(x)|^{-s}, \quad \text{où } L' = \{x \in L : P(x) \neq 0\}.$$

Les espaces vectoriels préhomogènes introduits par Sato [44, 46] systématisent cette étude.

DÉFINITION 3.1. — *Soit k un corps. Un espace vectoriel préhomogène sur k est une représentation (G, V) d'un groupe linéaire algébrique connexe G défini sur k sur un espace affine de dimension finie $V = \mathbb{A}_k^n$, possédant une G -orbite Zariski-dense.*

On notera S le fermé complémentaire de cette orbite dense. Un espace préhomogène est donc un espace « presque homogène », clôture de l'espace homogène Gx pour $x \notin S$.

DÉFINITION 3.2. — *On dit que (G, V) préhomogène est réductif régulier si G est réductif et S est une hypersurface irréductible de V . Un polynôme irréductible définissant S sera appelé invariant relatif de (G, V) .*

PROPOSITION 3.3 (Sato). — Si (G, V) est préhomogène réductif régulier, un invariant relatif P est un polynôme homogène, unique modulo k^* . Il existe un caractère rationnel χ de G tel que $P(g \cdot x) = \chi(g)P(x)$ pour tout $g \in G$ et

$$(8) \quad (\det \rho(g))^2 = \chi(g)^{2\kappa}, \quad \text{où } \kappa := \frac{\dim V}{\deg P} \in \frac{1}{2}\mathbb{Z},$$

et $\rho : G \rightarrow \mathrm{GL}(V)$ est la représentation de G sur V .

On verra au paragraphe suivant le lien entre ces définitions et la série (en fait, les séries) de Dirichlet associée à (6). Pour les applications, il est crucial que (G, V) soit défini sur \mathbb{R} , et utile qu'il soit régulier (sinon les fonctions obtenues ont une variable par composante irréductible de S , ce qui complique les choses).

On dispose d'une construction abstraite assez générale d'espaces préhomogènes : si G est un groupe réductif, $P = LU$ un sous-groupe parabolique maximal, la composante de Levi L agit par conjugaison sur le radical unipotent U , donc sur l'abélianisé $V = U/[U, U]$. Vinberg [52] démontre que (L, V) est préhomogène.

DÉFINITION 3.4. — Un tel espace préhomogène est dit de type parabolique.

Sur \mathbb{C} , les espaces préhomogènes irréductibles, modulo une relation d'équivalence naturelle (*roques* ou *castling transforms*), sont classifiés par Kimura et Sato [44] : il y a 29 types réguliers, dont 5 séries infinies : matrices $m \times n$, formes quadratiques en n variables, etc. Les formes réelles de ces espaces qui sont paraboliques sont classifiées par Rubenthaler [41] : presque tous les types de Kimura-Sato sont paraboliques, il y a 6 exceptions ne comportant aucune série infinie. Il existe aussi une classification sur un corps local ou un corps de nombres, due à Saito [42].

Inspiré par Wright et Yukié [56] (qui s'intéressent aux orbites rationnelles, cf. §3.3), Bhargava [3, 4, 5, 6, 7] a recherché systématiquement comment paramétrer des situations liées aux anneaux de nombres par les orbites *entières* d'espaces préhomogènes. Il se trouve que tous les exemples obtenus sont de type parabolique (réductif) régulier, associés aux groupes de Lie simples $G = B_2, G_2, B_3, D_4, D_5, E_6$ (corps quadratiques), G_2, F_4, E_6, E_7 (cubiques), F_4 (quartiques), E_8 (quintiques), pour un parabolique LU convenable. Guidé par la classification de Kimura-Sato et des considérations heuristiques sur les diagrammes de Dynkin et leurs symétries, Bhargava construit pour beaucoup d'entre eux des lois de composition que nous décrirons aux §§4 et 5. Les structures de groupe obtenues ont peu d'intérêt intrinsèque puisqu'on les obtient aussi via les groupes des classes des anneaux sous-jacents, du moins pour l'instant. Mais elles en donnent des descriptions explicites, et sont justiciables du même type de traitement qu'au §2.

3.2. La théorie de Sato-Shintani

Soit (G, V) réductif régulier associé à P, χ comme ci-dessus. On note H le noyau de χ . Soit $G_{\mathbb{R}}^+$ la composante connexe du neutre dans le groupe de Lie $G_{\mathbb{R}}$, $H_{\mathbb{R}}^+ = H_{\mathbb{R}} \cap G_{\mathbb{R}}^+$.

On fixe des mesures de Haar dg sur $G_{\mathbb{R}}^+$, d^1h sur $H_{\mathbb{R}}^+$, $d\nu_x$ sur $(H_{\mathbb{R}}^+)_x$ pour $x \in (V - S)$ avec

$$\begin{aligned} \int_{G_{\mathbb{R}}^+} \phi(g)dg &= \int_{G_{\mathbb{R}}^+/H_{\mathbb{R}}^+} \frac{d\chi(g)}{|\chi(g)|} \int_{H_{\mathbb{R}}^+} \phi(gh)d^1h \\ &= \int_{G_{\mathbb{R}}^+/(H_{\mathbb{R}}^+)_x} |P(g \cdot x)|^{-\kappa} d(g \cdot x) \int_{(H_{\mathbb{R}}^+)_x} \phi(gh)d\nu_x(h), \quad \forall \phi \in L^1(G_{\mathbb{R}}^+). \end{aligned}$$

(D'après (8), $|P(y)|^{-\kappa} dy$ est une mesure $G_{\mathbb{R}}^+$ -invariante sur $V - S$.) Soit $\Gamma = G_{\mathbb{Z}} \cap H_{\mathbb{R}}^+$. On fixe un réseau Γ -stable L , et on pose $L' = L - (L \cap S)$. On fait les hypothèses simplificatrices suivantes :

- (H1) (G, V) est *réductif régulier* défini sur \mathbb{Q} , tel que $G_{\mathbb{Z}} \cdot V_{\mathbb{Z}} \subset V_{\mathbb{Z}}$,
- (H2) $S \cap V_{\mathbb{R}}$ se décompose en un nombre fini de $H_{\mathbb{R}}^+$ -orbites,
- (H3) une hypothèse technique destinée à justifier la convergence des identités formelles, en particulier (11) : pour toute fonction ϕ dans la classe de Schwartz sur $V_{\mathbb{R}}$, l'intégrale

$$I(\phi) = \int_{H_{\mathbb{R}}/H_{\mathbb{Z}}} \sum_{x \in L} \phi(h \cdot x)d^1h$$

converge absolument et définit une distribution tempérée sur $V_{\mathbb{R}}$.

La dernière hypothèse est restrictive, elle ne couvre pas le cas des formes quadratiques $\text{Sym}^2 \mathbb{A}^n$ si $n \leq 4$ ou des formes cubiques $\text{Sym}^3 \mathbb{A}^2$ par exemple. Shintani [49, 50] règle le problème pour ces deux exemples en renormalisant les fonctions zêta.

Soient V_1, \dots, V_l les composantes connexes de $(V - S)_{\mathbb{R}}$, qui sont des $G_{\mathbb{R}}^+$ -orbites. Pour tout $x \in (V - S)_{\mathbb{Q}}$, on définit la fonction Γ -invariante

$$(9) \quad \mu(x) = \int_{(H_{\mathbb{R}}^+)_x/\Gamma_x} d\nu_x < \infty.$$

L'hypothèse (H3) implique la finitude de $\mu(x)$, en fait une majoration polynomiale en X de la somme des $\mu(x)$ sur $\{x \in L'/\Gamma, |P(x)| < X\}$. On définit les séries de Dirichlet

$$(10) \quad \xi_i(s, L) = \sum_{x \in (L \cap V_i)/\Gamma} \mu(x) |P(x)|^{-s}, \text{ associées à } \sum_{\substack{x \in (L \cap V_i)/\Gamma \\ 0 < |P(x)| < X}} \mu(x) \quad (1 \leq i \leq l),$$

qui convergent donc pour $\Re(s) \gg 1$. (Ceci est démontré par Saito [43] sous des hypothèses bien moins restrictives que (H3).) Si Γ_x est fini pour tout x , on obtient $\mu(x) = c |\Gamma_x|^{-1}$ pour une constante c indépendante de x . Il s'agit donc bien d'un raffinement de (7), qu'on retrouve par sommation sur les composantes connexes.

On obtient des notions analogues pour la représentation contragrédiente sur le dual (G, V^*) qui est aussi préhomogène : S^*, P^* de même degré que P , $\chi^* = \chi^{-1}$, $V_1^* \cup \dots \cup V_l^*$ (pour le même l), $d\nu^*(x)$, $\xi_i^*(s, L^*)$, etc. Sous nos hypothèses, en considérant

$$(11) \quad Z(\phi, s) := \int_{G_{\mathbb{R}}^+/\Gamma} |\chi(g)|^s \sum_{x \in L'} \phi(g \cdot x)dg = \sum_{i=1}^l \xi_i(s) \int_{V_i} \phi(y) |P(y)|^{s-\kappa} dy$$

et la distribution duale Z^* , Sato et Shintani [45] démontrent que les séries $\xi_i(s, L)$, $\xi_i^*(s, L^*)$ admettent un prolongement analytique méromorphe sur \mathbb{C} et satisfont une équation fonctionnelle matricielle $l \times l$ reliant

$$(\xi_i(s, L): 1 \leq i \leq l) \quad \text{et} \quad (\xi_i^*(\kappa - s, L^*): 1 \leq i \leq l).$$

Le prolongement vient d’une intégration par parties qui fait intervenir des polynômes de Bernstein-Sato ou « fonctions b », définis par

$$P^*(\nabla_x)P(x)^s = b(s)P(x)^{s-1},$$

où $P^*(\nabla_x)$ est l’opérateur différentiel à coefficients constants sur $V_{\mathbb{R}}$ satisfaisant

$$P^*(\nabla_x) \exp(\langle x, y \rangle) = P^*(y) \exp(\langle x, y \rangle), \quad \forall y \in V_{\mathbb{R}}^*.$$

On a $\deg b = \deg P$ et, sous nos hypothèses, les zéros de $b(s - \kappa)$ sont les pôles (simples) des ξ_i et ξ_i^* . Les fonctions b sont connues explicitement pour les 29 types réguliers de Kimura-Sato, ainsi que leur comportement par roque (*cf.* Kimura [33]) et leurs zéros sont rationnels (voir aussi Kashiwara [32]). Dans les autres cas où les ξ_i , ξ_i^* sont connus, la multiplicité d’un zéro de $b(s - \kappa)$ borne celle du pôle, mais il arrive qu’on n’ait pas égalité. L’équation fonctionnelle suit de la formule sommatoire de Poisson appliquée à la série théta intervenant dans $Z(\phi, s)$, en isolant la partie polaire ($x \in L - L'$, $x^* \in L^* - L'^*$). Par transformée de Mellin et déplacement de contour, on en tire un développement asymptotique des sommes partielles (10) quand $X \rightarrow +\infty$.

Exemple 3.5. — Pour $(G, V) = (\mathrm{GL}_1, \mathbb{A}^1)$, le plus simple des espaces préhomogènes, on trouve $P(x) = x$, $\kappa = 1$, $b(s) = s$, $(V - S)_{\mathbb{R}} = V_1 \cup V_2 = \mathbb{R}_+^* \cup \mathbb{R}_-^*$ et $G_x = \{1\}$ pour $x \in (V - S)_{\mathbb{R}}$, soit $\mu(x) = 1$ avec la normalisation naturelle. Ainsi $\xi_i(s) = \xi_i^*(s) = \zeta(s)$ pour $i = 1, 2$ et on retrouve les résultats de Riemann, essentiellement par la même méthode.

3.3. Adélisation et G_k -orbites

Si \mathbb{A} désigne les adèles du corps global k , on remplace le sous-groupe discret $G_{\mathbb{Z}} \subset G_{\mathbb{R}}$ par $G_k \subset G_{\mathbb{A}}$. Pour $\phi = \otimes_v \phi_v$ une fonction de Schwartz-Bruhat sur $V_{\mathbb{A}}$, on pose

$$(12) \quad Z(\phi, s) = \int_{G_{\mathbb{A}}/G_k} |\chi(g)|_{\mathbb{A}}^s \sum_{x \in L'} \phi(g \cdot x) dg,$$

et le formalisme est proche de celui du paragraphe précédent, en distinguant une place infinie de k . Les séries de Dirichlet associées se décomposent en somme sur les orbites L'/G_k de termes faisant apparaître des produits eulériens sur v de fonctions zêta locales

$$(13) \quad \int_{G_{k_v} \cdot x} \phi_v(y) |P(y)|_v^{s-\kappa} dy,$$

pondérés par $\mu(x)/o(x)$, où $o(x) = [(H_x)_k : (H_x^0)_k]$ et $\mu(x)$ est le volume de $(H_x^0)_{\mathbb{A}} / (H_x^0)_k$ pour une mesure convenable (voir [57]). Pour v infinie, les b -fonctions permettent comme précédemment le prolongement méromorphe de (13). Si v est finie et ϕ_v est la fonction caractéristique du disque unité \mathcal{O}_{k_v} , on obtient une fonction locale d’Igusa, d’expression

élémentaire connue pour presque tous les types de Kimura-Sato (voir [30]). Hors d'un ensemble fini de places T , ϕ_v est de cette forme, et la fonction zêta pour T fixée est contrôlée. Par contre, pour énumérer les G_k -orbites, il faut un passage à la limite sur T , axiomatisé par Wright (cf [57, §0.5]), d'esprit analogue à celui du §2.3, mais bien plus contraignant. Ceci étant dit, que représentent les G_k -orbites ?

Exemple 3.6. — Reconsidérons les formes quadratiques binaires du §2 ; la $\mathrm{SL}_2(k)$ -orbite d'une forme quadratique irréductible x de $\mathrm{Sym}^2 k^2$ définit un corps quadratique K_x/k : l'orbite des racines de x . Les formes réductibles forment deux orbites, suivant qu'elles ont ou non une racine multiple. En étudiant la fonction zêta adélique associée, Datskovsky [19] obtient une version relative du Théorème 2.5 sur un corps de nombres k : dans ce cas, $\mu(x)$ est essentiellement $|\mathrm{Disc} K_x|^{1/2} \mathrm{Res}(\zeta_{K_x}, s = 1)$.

Exemple 3.7. — Shintani [50, 49], dans les premières études sur les séries de Dirichlet provenant de la théorie du §3.2, obtient d'excellents termes d'erreur pour le nombre de classes de formes quadratiques ou cubiques binaires avec les méthodes du §3.2. À nouveau, l'ensemble naturel des orbites de formes cubiques est stratifié suivant le type de décomposition de l'équation associée. En traitant les formes quadratiques et cubiques générales, Shintani peut isoler les $G_{\mathbb{Z}}$ -orbites de formes cubiques irréductibles. Nous verrons au §5.1 que ce dernier exemple paramètre les anneaux cubiques. Datskovsky et Wright [54, 17, 18] adélisent le travail de Shintani sur un corps global k de caractéristique différente de 2 ou 3. Comptant les G_k -orbites de formes cubiques irréductibles, ils obtiennent la densité des discriminants des corps cubiques sur k .

Remarque 3.8. — Dans ces deux cas, la fonction zêta pour T fixé admet un autre pôle réel à gauche de 1, mais le passage à la limite sur T empêche de tenir compte de ce pôle secondaire, et même d'obtenir un terme d'erreur. On conjecture que le passage à la limite formel donne le développement asymptotique correct (voir Roberts [40] pour le cas cubique).

Plus généralement, soit $n \geq 1$ et k un corps infini de caractéristique nulle ou strictement supérieure à n . On note $E(k, n)$ l'ensemble des classes d'isomorphismes d'extensions galoisiennes de k qui sont corps de décomposition d'un polynôme séparable de degré inférieur à n . Wright et Yukie [56] (voir aussi [31]) construisent pour six nouveaux types d'espaces préhomogènes paraboliques (G, V) une bijection naturelle entre G_k -orbites de $(V - S)_k$ et classes de conjugaison d'homomorphismes de $\mathrm{Gal}(\bar{k}/k)$ dans le groupe symétrique à n éléments où $n = 2, 3, 4, 5$ suivant les cas. Ceci définit une application

$$\alpha_V : (V - S)_k / G_k \rightarrow E(k, n),$$

où une orbite a pour image le noyau d'un des homomorphismes associés. En inspectant les classes de conjugaison de S_n , on voit que α_V est bijective pour les six exemples associés à $n = 2, 3$. Dans les deux cas restants, $n = 4, 5$ et les fibres sont réduites à un point, sauf dans des cas dégénérés (2 ou 4 points) correspondant à des extensions

de petit degré, inférieur à 12. Quand k est un corps global, en supposant résolus les problèmes de convergence et la détermination des pôles et résidus, on espère pouvoir compter les extensions x de k de degré n munies du poids $\mu(x)/o(x)$, mais ce programme n'est pas achevé. La partie globale du cas quartique est traitée par Yukie [57].

3.4. Comparaison

Pour conclure cette présentation des méthodes en présence⁽⁴⁾, les calculs de densités de discriminants d'extensions de type Wright-Yukie reposent sur l'existence d'une représentation (G, V) préhomogène, telle que V_k/G_k paramètre les extensions du corps k . Ces constructions sont actuellement restreintes aux extensions de degré $n \leq 5$.

Bhargava démontre que dans ces mêmes cas, $V_{\mathbb{Z}}/G_{\mathbb{Z}}$ paramètre essentiellement des anneaux de nombres de degré n sur \mathbb{Z} (et non des corps). En principe, les méthodes élémentaires du §2 permettent de les énumérer, puis de les cribler pour ne retenir que les ordres maximaux, ce qui revient à compter leurs corps de fractions, *i.e.* les $G_{\mathbb{Q}}$ -orbites de $V_{\mathbb{Q}}$. Ces méthodes fournissent des termes d'erreur qui survivent aux cribles, mais restent loin des valeurs conjecturées à la lecture des pôles des séries de Sato-Shintani.

À l'inverse, ces dernières fournissent les termes d'erreur attendus sur les problèmes de comptage de $G_{\mathbb{Z}}$ -orbites et leur adélisation permet le passage aux G_k -orbites sur V_k , et donc aux extensions du corps global k par la théorie de Wright-Yukie, pour $n = 2, 3$. Par contre, elle ne fournit pour l'instant qu'un équivalent pour ces décomptes d'extensions de k , y compris quand $k = \mathbb{Q}$, et se heurte à de redoutables problèmes de convergence d'identités formelles (voir [57, Part IV]).

Ce sont les problèmes de comptage associés aux deux cas $n = 4, 5$ ci-dessus que Bhargava vient, semble-t-il, de résoudre sur \mathbb{Q} , en étudiant directement les domaines fondamentaux associés aux $G_{\mathbb{Z}}$ -orbites des espaces préhomogènes de Wright et Yukie. Nous décrirons une partie de ces travaux à partir du §5.3, mais nous commençons par sa ré-interprétation des cas $n = 2, 3$.

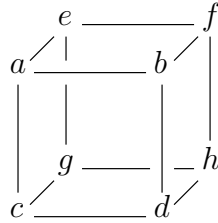
4. PARAMÉTRISATIONS ET COMPOSITIONS QUADRATIQUES

4.1. Cubes

Soit $\mathcal{C}_2 = \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$, dont on peut représenter les éléments par des octuplets $(a, b, c, d, e, f, g, h) \in \mathbb{Z}^8$ ou plus naturellement par un cube de sommets étiquetés par

⁽⁴⁾Pour être complet, il eût fallu inclure les séries génératrices issues de la théorie du corps de classes ou directement de la théorie de Kummer, pour les extensions abéliennes d'une base raisonnable. Elles sont disjointes des travaux de Bhargava que nous présentons et nous renvoyons au survol de Cohen [11].

des entiers :



En notant (α, β) la base canonique de \mathbb{Z}^2 , ce cube remplace avantageusement l'élément

$$a(\alpha \otimes \alpha \otimes \alpha) + b(\alpha \otimes \beta \otimes \alpha) + c(\beta \otimes \alpha \otimes \alpha) + d(\beta \otimes \beta \otimes \alpha) \\ + e(\alpha \otimes \alpha \otimes \beta) + f(\alpha \otimes \beta \otimes \beta) + g(\beta \otimes \alpha \otimes \beta) + h(\beta \otimes \beta \otimes \beta).$$

On peut partitionner un tel cube $A \in \mathcal{C}_2$ en deux matrices 2×2 de trois façons différentes : $M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $N_1 = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$, ou $M_2 = \begin{pmatrix} a & c \\ e & g \end{pmatrix}$, $N_2 = \begin{pmatrix} b & d \\ f & h \end{pmatrix}$, ou encore $M_3 = \begin{pmatrix} a & e \\ b & f \end{pmatrix}$, $N_3 = \begin{pmatrix} c & g \\ d & h \end{pmatrix}$. Soit $\gamma = \begin{pmatrix} r & s \\ t & u \end{pmatrix}$ un élément générique de $\mathrm{SL}_2(\mathbb{Z})$. On définit une action de $\Gamma = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ sur \mathcal{C}_2 , en spécifiant que $(\gamma \times \mathrm{Id} \times \mathrm{Id})$ agit sur le cube A en remplaçant (M_1, N_1) par $(rM_1 + sN_1, tM_1 + uN_1)$, $(\mathrm{Id} \times \gamma \times \mathrm{Id})$ et $(\mathrm{Id} \times \mathrm{Id} \times \gamma)$ agissant de même sur (M_2, N_2) et (M_3, N_3) respectivement. Il s'agit bien d'une action (à gauche!), on vérifie que les actions des trois facteurs $\mathrm{SL}_2(\mathbb{Z})$ dans Γ commutent. C'est l'analogie de l'associativité de la multiplication matricielle : les opérations sur les lignes et colonnes d'une matrice rectangulaire commutent. Par exemple, la transformation de (M_1, N_1) indiquée se traduit par $(M_2, N_2) \rightarrow (\gamma M_2, \gamma N_2)$ et $(M_3, N_3) \rightarrow (M_3^t \gamma, N_3^t \gamma)$. Alternativement, cette action est donnée par

$$(\gamma_1, \gamma_2, \gamma_3) \cdot (x_1 \otimes x_2 \otimes x_3) = \gamma_3 x_1 \otimes \gamma_2 x_2 \otimes \gamma_1 x_3.$$

(Nous conservons les normalisations de Bhargava.)

Étant donné un cube $A \in \mathcal{C}_2$, on construit pour $1 \leq i \leq 3$ une forme quadratique binaire $Q_i = Q_i^A$ par la règle

$$Q_i(x, y) = -\det(xM_i - yN_i).$$

La forme Q_1 est invariante sous l'action de $\{\mathrm{Id}\} \times \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}) \subset \Gamma$, puisque les deux autres facteurs agissent par multiplication à gauche ou à droite sur (M_1, N_1) . Le premier facteur agit de la façon habituelle $(\gamma \cdot Q_1)(x, y) = Q_1((x, y)\gamma)$, et cette action a un unique invariant, à savoir $\mathrm{Disc} Q_1$. Il en est donc de même pour l'action de Γ sur \mathcal{C}_2 . Par symétrie, $\mathrm{Disc} Q_2$ et $\mathrm{Disc} Q_3$ sont aussi des invariants, et on vérifie que $\mathrm{Disc} Q_1 = \mathrm{Disc} Q_2 = \mathrm{Disc} Q_3$. On baptise cette valeur commune $\mathrm{Disc} A$. Explicitement,

$$Q_1 = (bc - ad, -ah + bg + cf - de, fg - eh),$$

$$Q_2 = (ce - ag, -ah - bg + cf + de, df - bh),$$

$$Q_3 = (be - af, -ah + bg - cf + de, dg - ch),$$

$$\mathrm{Disc} A = a^2 h^2 + b^2 g^2 + c^2 f^2 + d^2 e^2 - 2(abgh + acfh + adeh + bcfg + bdeg + cdef) + 4(adfg + bceh).$$

Toute forme quadratique $(d, h, g) \in (\mathrm{Sym}^2 \mathbb{Z}^2)^*$ provient d'un cube : prendre $a = -1$, $b = c = e = 0$, $f = 1$ par exemple.

4.2. Loi du cube et loi de Gauss

DÉFINITION 4.1. — *Un cube $A \in \mathcal{C}_2$ est dit projectif si les trois formes quadratiques associées (Q_1^A, Q_2^A, Q_3^A) sont primitives. On note $\text{Cl}(\mathcal{C}_2; D)$ l'ensemble des classes de cubes projectifs modulo Γ de discriminant D .*

Motivé par la loi de groupe sur une courbe elliptique, Bhargava considère le groupe libre engendré par les formes quadratiques *primitives* de discriminant D , modulo les relations

$$(14) \quad Q_1^A \oplus Q_2^A \oplus Q_3^A = 0,$$

où $A \in \mathcal{C}_2$. En particulier, deux formes $\text{SL}_2(\mathbb{Z})$ -équivalentes sont identifiées puisque, si Q_1 est donnée, alors il existe un cube A comportant Q_1 parmi ses formes associées (Q_1, Q_2, Q_3) . En considérant $(\gamma \times \text{Id} \times \text{Id}) \cdot A$, on en déduit que

$$Q_1 \oplus Q_2 \oplus Q_3 = \gamma Q_1 \oplus Q_2 \oplus Q_3 = 0,$$

soit $Q_1 = \gamma Q_1$ dans le quotient de Bhargava, pour tout $\gamma \in \text{SL}_2(\mathbb{Z})$. De même, on voit facilement que $(a, b, c) \oplus (c, b, a) = 0$.

THÉORÈME 4.2. — *Soit $D \equiv 0, 1 \pmod{4}$ un entier. Fixons un cube projectif A_0 de discriminant D , dont les formes quadratiques associées sont égales ; soit Q^0 cette forme primitive. Il existe une unique loi de groupe additif sur l'ensemble des classes de formes quadratiques binaires $[Q]$ primitives modulo $\text{SL}_2(\mathbb{Z})$, de discriminant D , telle que $[Q^0] = 0$, et $[Q_1^A] + [Q_2^A] + [Q_3^A] = 0$ pour tout cube projectif $A \in \mathcal{C}_2$. Réciproquement, pour tout triplet de classes de formes primitives de somme nulle dans ce groupe, il existe un cube projectif $A \in \mathcal{C}_2$, unique à Γ -équivalence près, dont ce sont les formes associées.*

On démontrera ce résultat en même temps que le Théorème 4.6. Les cubes A_0 satisfaisant la condition du théorème avec $Q^0 \neq (0, 0, 0)$ présentent une triple symétrie :

$$(15) \quad \begin{array}{ccc} & b & \text{---} & g \\ a & \diagdown & & \diagup b \\ & | & & | \\ & g & \text{---} & h \\ b & \diagdown & & \diagup g \end{array}$$

Le choix le plus naturel est $Q^0 = (1, \varepsilon, (\varepsilon - D)/4)$ pour $D \equiv \varepsilon \pmod{4}$, $\varepsilon \in \{0, 1\}$, associée au cube A_0 de discriminant D donné par

$$(16) \quad \begin{array}{ccc} & 1 & \text{---} & \varepsilon \\ 0 & \diagdown & & \diagup 1 \\ & | & & | \\ & \varepsilon & \text{---} & (D + 3\varepsilon)/4 \\ 1 & \diagdown & & \diagup \varepsilon \end{array}$$

THÉORÈME 4.3. — Pour le choix de A_0 donné par (16), la loi de groupe du Théorème 4.2 est la composition de Gauss du Théorème 2.1.

Preuve. — $[Q^0]$ est bien la classe principale de Gauss. Si A est un cube projectif, le pgcd de ses coefficients est 1. À Γ -équivalence près, on peut donc supposer qu'un sommet est 1. Toujours à Γ -équivalence près, on utilise ce 1 pour annuler les trois sommets adjacents (pivot de Gauss tridimensionnel!). On suppose donc que A est de la forme

$$\begin{array}{ccc}
 & 0 & \text{---} & f \\
 1 & \diagdown & & \diagup & 0 \\
 & | & & | & \\
 & 1 & \text{---} & 0 & \\
 & | & & | & \\
 0 & \diagdown & g & \text{---} & d & \diagup & h \\
 & | & & | & \\
 & 0 & \text{---} & d & \\
 & & & & & & h
 \end{array}$$

soit

$$Q_1 = (-d, h, fg), \quad Q_2 = (-g, h, df), \quad Q_3 = (-f, h, dg).$$

De $[Q_1] + [Q_2] = -[Q_3] = (dg, h, -f)$, on tire la composition de Dirichlet (2). □

COROLLAIRE 4.4. — Soit $D \equiv 0, 1 \pmod{4}$ un entier, et soit A_0 le cube (16). Il existe une unique loi de groupe sur $\text{Cl}(\mathcal{C}_2; D)$, de neutre $\Gamma \cdot A_0$, telle que les projections

$$\begin{aligned}
 \phi_i : \text{Cl}(\mathcal{C}_2; D) &\rightarrow \text{Cl}^+(D) \\
 A &\mapsto [Q_i^A]
 \end{aligned}$$

soient des homomorphismes de groupe pour $1 \leq i \leq 3$.

Preuve. — Si A et A' sont deux cubes projectifs, alors

$$\sum_{i=1}^3 ([Q_i^A] + [Q_i^{A'}]) = \sum_{i=1}^3 [Q_i^A] + \sum_{i=1}^3 [Q_i^{A'}] = 0 + 0 = 0.$$

D'après le dernier point du Théorème 4.2, il existe $A'' \in \mathcal{C}_2$ tel que $[Q_i^{A''}] = [Q_i^A] + [Q_i^{A'}]$ pour $i = 1, 2, 3$. Les $Q_i^{A''}$ étant primitives par définition de la loi de groupe, A'' est projectif. On pose $A'' = A' + A$. □

4.3. Paramétrisations

Identifions maintenant ces lois de composition.

DÉFINITION 4.5. — Soit S l'anneau quadratique de discriminant D et $K = S \otimes_{\mathbb{Z}} \mathbb{Q}$ l'algèbre quadratique associée. Un triplet (I_1, I_2, I_3) d'idéaux orientés de S est équilibré si $I_1 I_2 I_3 \subset S$ et $N(I_1) N(I_2) N(I_3) = 1$. Deux tels triplets (I_1, I_2, I_3) et (I'_1, I'_2, I'_3) sont équivalents s'il existe $\kappa_1, \kappa_2, \kappa_3 \in K$ tel que $I_i = \kappa_i I'_i$ pour $1 \leq i \leq 3$.

Si S est un anneau de Dedekind, une classe d'équivalence de triplets équilibrés n'est rien d'autre qu'un triplet de classes d'idéaux restreintes de produit 1.

THÉORÈME 4.6. — Il existe une bijection canonique entre

- l'ensemble des Γ -orbites de discriminant $D \neq 0$ sur l'espace $\mathcal{C}_2 = \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$,
- l'ensemble des classes d'isomorphismes de paires $(S, (I_1, I_2, I_3))$, où S est un anneau quadratique orienté de discriminant D , et (I_1, I_2, I_3) est une classe d'équivalence de triplets équilibrés d'idéaux orientés de S .

Sa restriction aux cubes projectifs induit un isomorphisme de groupes

$$\mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; D) \simeq \mathrm{Cl}^+(D) \times \mathrm{Cl}^+(D).$$

Preuve. — Soit $D \equiv \varepsilon \pmod{4}$, $\varepsilon \in \{0, 1\}$ et soit $\langle 1, \tau \rangle$ une base positive de S , pour une orientation $\pi : S/\mathbb{Z} \rightarrow \mathbb{Z}$, où $\tau^2 - \varepsilon\tau + (\varepsilon - D)/4 = 0$. Soient $\langle \alpha_1, \alpha_2 \rangle$, $\langle \beta_1, \beta_2 \rangle$, $\langle \gamma_1, \gamma_2 \rangle$ des bases de I_1, I_2, I_3 , de même orientation que I_1, I_2, I_3 respectivement. Comme $I_1 I_2 I_3 \subset S$, on a

$$(17) \quad \alpha_i \beta_j \gamma_k = c_{ijk} + a_{ijk} \tau,$$

pour des entiers a_{ijk} et c_{ijk} , $1 \leq i, j, k \leq 2$. Le cube associé est $A = (a_{ijk})$. De façon plus intrinsèque, $A \in \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2$ représente l'application trilinéaire $I_1 \times I_2 \times I_3 \rightarrow \mathbb{Z}$ donnée par la formule $(x, y, z) \mapsto \pi(xyz)$. On vérifie que A est bien défini à Γ -équivalence près.

Réciproquement, pour un cube $A = (a_{ijk})$ fixé, on considère le système (17), qui comporte essentiellement des indéterminées pour l'instant. On cherche $\tau, (\alpha_i), (\beta_j), (\gamma_k)$ le satisfaisant tels que les S, I_1, I_2, I_3 associés vérifient $I_1 I_2 I_3 \subset S$ et $N(I_1)N(I_2)N(I_3) = 1$. Ceci implique

$$\mathrm{Disc} A = N(I_1)^2 N(I_2)^2 N(I_3)^2 \mathrm{Disc} S,$$

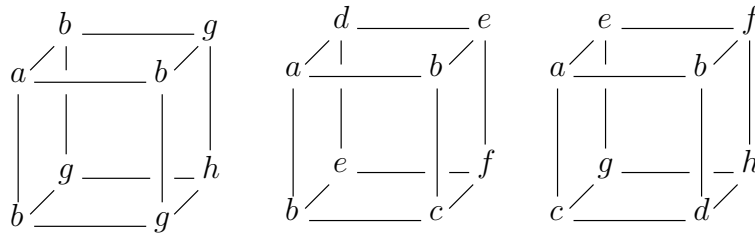
ainsi $\mathrm{Disc} S = \mathrm{Disc} A$ et donc S sont déterminés. Comme dans la preuve du Théorème 4.3, on peut supposer que les trois sommets adjacents à un sommet fixé (qui porte le pgcd des coefficients) sont nuls. Par associativité et commutativité de la multiplication dans S , un calcul explicite montre que les c_{ijk} sont déterminés, et entiers! Grâce à la nullité de trois des a_{ijk} , on déduit de (17) que les $\alpha_i, \beta_j, \gamma_k$ sont inversibles dans K , puis que les quotients $\alpha_1/\alpha_2, \beta_1/\beta_2$ et γ_1/γ_2 sont fixés. Un calcul explicite montre que les \mathbb{Z} -modules obtenus sont bien des idéaux.

Un triplet équilibré (I_1, I_2, I_3) est dit projectif si les I_i sont projectifs (c'est-à-dire inversibles) comme S -modules. Les formes normes associées aux I_i sont exactement les Q_i^A , donc les I_i sont projectifs si et seulement si A l'est. L'ensemble des classes d'équivalence de triplets équilibrés projectifs est muni de la loi de groupe naturelle $(I_1, I_2, I_3) \cdot (I'_1, I'_2, I'_3) = (I_1 I'_1, I_2 I'_2, I_3 I'_3)$, qui le rend isomorphe à $\mathrm{Cl}^+(D) \times \mathrm{Cl}^+(D)$ par la projection $(I_1, I_2, I_3) \mapsto (I_1, I_2)$. On conclut grâce aux Théorèmes 4.3 et 2.1. \square

Bhargava définit de même des flèches naturelles préservant le discriminant entre espaces de formes

$$\begin{array}{ccccc}
 \mathrm{Sym}^3 \mathbb{Z}^2 & \longrightarrow & \mathbb{Z}^2 \otimes \mathrm{Sym}^2 \mathbb{Z}^2 & \longrightarrow & \mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2 \\
 & & \downarrow & & \downarrow \\
 & & (\mathrm{Sym}^2 \mathbb{Z}^2)^* & \longleftarrow & \mathbb{Z}^2 \otimes \Lambda^2 \mathbb{Z}^4 \\
 & & & & \downarrow \\
 & & & & \Lambda^3 \mathbb{Z}^2
 \end{array}$$

munies d’actions de groupes linéaires. Par exemple, la première rangée correspond à l’inclusion des ensembles de cubes présentant une triple symétrie, une double symétrie ou pas de symétrie *a priori* :



À partir de la loi du cube, on obtient des lois de groupe sur les ensembles d’orbites projectives de discriminant D non nul :

$$\begin{array}{ccccc}
 \mathrm{Cl}(\mathrm{Sym}^3 \mathbb{Z}^2; D) & \longrightarrow & \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathrm{Sym}^2 \mathbb{Z}^2; D) & \longrightarrow & \mathrm{Cl}(\mathbb{Z}^2 \otimes \mathbb{Z}^2 \otimes \mathbb{Z}^2; D) \\
 & & \downarrow \simeq & & \downarrow \\
 \mathrm{Cl}^+(D) \simeq \mathrm{Cl}((\mathrm{Sym}^2 \mathbb{Z}^2)^*; D) & \longleftarrow \simeq & \mathrm{Cl}(\mathbb{Z}^2 \otimes \Lambda^2 \mathbb{Z}^4; D) & & \\
 & & & & \downarrow \\
 & & & & \mathrm{Cl}(\Lambda^3 \mathbb{Z}^2; D) = \{0\}
 \end{array}$$

Bhargava identifie ensuite les structures dont ils paramètrent les classes d’équivalence. Dans l’énumération suivante, toutes les formes et les anneaux sont de discriminant non nul, S désigne un anneau quadratique orienté, I (avec ou sans indice) un S -idéal orienté, et un « S -idéal de rang n » est un sous- S -module de $(S \otimes \mathbb{Q})^n$, de rang maximal $2n$ sur \mathbb{Z} . Voir [5] pour les définitions manquantes et les démonstrations.

- $\mathrm{Sym}^3 \mathbb{Z}^2$, $\mathrm{SL}_2(\mathbb{Z})$: formes cubiques binaires et triplets (S, I, δ) , où $\delta \in S \otimes \mathbb{Q}$ tel que $I^3 \subset \delta S$ et $N(I)^3 = N(\delta)$. Si on se restreint aux formes projectives, I est inversible et $N(I^3) = N(I)^3$, d’où $I^3 = (\delta)$. L’application $\mathrm{Cl}(\mathrm{Sym}^3 \mathbb{Z}^2; D) \rightarrow \mathrm{Cl}_3(D)$ donnée par $(S, I, \delta) \rightarrow I$ est un morphisme surjectif dont le noyau est de cardinal $\#(S^*/S^{*3})$. En particulier si S est un anneau de Dedekind, ce morphisme est un isomorphisme si $D < -3$, et a un noyau d’ordre 3 sinon (cette construction remonte à Eisenstein [24], voir aussi Hoffman-Morales [28]).

- $\mathbb{Z}^2 \otimes \text{Sym}^2 \mathbb{Z}^2$, $\text{SL}_2(\mathbb{Z}) \times \text{SL}_2(\mathbb{Z})$: paires de formes quadratiques binaires et triplets $(S, I_1, I_2, I_3 = I_2)$ où (I_1, I_2, I_2) est un triplet équilibré. L'application naturelle $\mathbb{Z}^2 \otimes \text{Sym}^2 \mathbb{Z}^2 \rightarrow (\text{Sym}^2 \mathbb{Z}^2)^*$ donnée par $A \mapsto Q_3^A$ devient un isomorphisme par passage aux quotients, si on la restreint aux classes projectives $(I_1 I_2 I_3)$ étant principal, si $I_2 = I_3$ l'est, I_1 aussi).
- $(\text{Sym}^2 \mathbb{Z}^2)^*$, $\text{SL}_2(\mathbb{Z})$: formes quadratiques binaires, c'est le cas considéré par Gauss.
- $\mathbb{Z}^2 \otimes \Lambda^2 \mathbb{Z}^4$, $\text{SL}_2(\mathbb{Z}) \times \text{SL}_4(\mathbb{Z})$: paires de 2-formes alternées de rang 4. Elles paramètrent les paires $(S, (I, M))$, où M est un « S -idéal de rang 2 » et (I, M) est équilibré.
- $\Lambda^3 \mathbb{Z}^6$, $\text{SL}_6(\mathbb{Z})$: 3-formes alternées de rang 6. Elles paramètrent les paires (S, M) où M est un « S -idéal de rang 3 » équilibré.

5. PARAMÉTRISATIONS EN DEGRÉ SUPÉRIEUR

5.1. Anneaux cubiques

THÉORÈME 5.1 (Delone-Faddeev [22], Gan-Gross-Savin [26])

Il existe une bijection canonique entre les deux ensembles suivants :

- les classes d'isomorphismes d'anneaux cubiques,
- les formes cubiques binaires entières, soit $(\text{Sym}^3 \mathbb{Z}^2)^*$, modulo l'action de $\text{GL}_2(\mathbb{Z})$.

Cette bijection préserve le discriminant. La classe contenant la forme de coefficients (a, b, c, d) est associée au \mathbb{Z} -module libre $R = \langle 1, \omega, \theta \rangle_{\mathbb{Z}}$, muni de la multiplication

$$\begin{aligned}\omega\theta &= -ad, \\ \omega^2 &= -ac + b\omega - a\theta, \\ \theta^2 &= -bd + d\omega - c\theta.\end{aligned}$$

Preuve. — Vérification explicite, facilitée par le choix d'une base de R telle que $\omega\theta \in \mathbb{Z}$, toujours possible par translation de ω et θ . Une autre démonstration pour R intègre consiste à comparer deux applications classiques : la forme indice (qui, à un ordre de rang n , associe une forme de $(\text{Sym}^{n(n-1)/2} \mathbb{Z}^{n-1})^*$, modulo $\text{GL}_{n-1}(\mathbb{Z})$) et l'ordre de Dedekind (qui, à une forme irréductible de $(\text{Sym}^n \mathbb{Z}^2)^*$, associe un ordre de degré n). Elles sont compatibles si et seulement si $n = 3$, et inverses l'une de l'autre dans ce cas. \square

Par exemple

$$(18) \quad (0, 0, 0, 0) \longleftrightarrow \mathbb{Z}[\omega, \theta]/(\omega^2, \theta^2, \omega\theta),$$

$$(19) \quad (a, b, c, d) \longleftrightarrow \mathbb{Z}[a\omega, a\omega^2 + b\omega]/(a\omega^3 + b\omega^2 + c\omega + d) \quad \text{si } a \neq 0.$$

Zagier [58] a donné une jolie interprétation de l’application réciproque, en associant à R l’application

$$\begin{aligned}\Phi_{3,2} : R/\mathbb{Z} &\rightarrow \Lambda^3 R \cong \mathbb{Z} \\ \xi &\mapsto 1 \wedge \xi \wedge \xi^2,\end{aligned}$$

que l’on identifie à une forme cubique en choisissant une base $\langle \alpha, \beta \rangle$ du \mathbb{Z} -module R/\mathbb{Z} de rang 2 et en posant $\xi = x\alpha + y\beta$. La construction est bien définie modulo $\mathrm{GL}_2(\mathbb{Z})$.

5.2. Paramétrisations cubiques

L’espace $V_{\mathbb{Z}} = \mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$ peut se représenter comme l’ensemble des boîtes $2 \times 3 \times 3$ à sommets entiers, ou encore les paires (A, B) de matrices 3×3 . De façon analogue au §4.1, on le munit d’une action de $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{GL}_3(\mathbb{Z}) \times \mathrm{GL}_3(\mathbb{Z})$. On se restreint à $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ puisque $(-\mathrm{Id}_2, \mathrm{Id}_3, -\mathrm{Id}_3)$ et $(-\mathrm{Id}_2, -\mathrm{Id}_3, \mathrm{Id}_3)$ agissent trivialement (cette action est fidèle). Soit $f(x, y)$ la forme cubique binaire $\det(xA - yB)$, on note

$$\mathrm{Disc}((A, B)) = \mathrm{Disc}(\det(xA - yB)) = \mathrm{Disc}(f),$$

qui est l’unique $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ -invariant sur $V_{\mathbb{Z}}$. D’après le Théorème 5.1, on associe un ordre cubique R de discriminant $\mathrm{Disc}(f)$ à f .

THÉORÈME 5.2. — *Il y a une bijection canonique entre*

- l’ensemble des $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ -orbites de discriminant $D \neq 0$ sur $\mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$,
- l’ensemble des classes d’isomorphismes de paires $(R, (I, I'))$, où R est un anneau cubique de discriminant $D \neq 0$, et (I, I') est une classe d’équivalence de paires équilibrées de R -idéaux fractionnaires de $R \otimes \mathbb{Q}$.

THÉORÈME 5.3. — *Il y a une bijection canonique entre*

- l’ensemble des $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ -orbites de discriminant $D \neq 0$ sur $\mathbb{Z}^2 \otimes \mathrm{Sym}^2 \mathbb{Z}^3$,
- l’ensemble des classes d’isomorphismes de triplets (R, I, δ) , où R est un anneau cubique de discriminant $D \neq 0$, et I est un R -idéal et δ est un élément inversible de $R \otimes \mathbb{Q}$, tels que $I^2 \subset (\delta)$, $N(\delta) = N(I)^2$.

Tout comme au §4, on obtient des applications naturelles

$$\mathbb{Z}^2 \otimes \mathrm{Sym}^2 \mathbb{Z}^3 \longrightarrow \mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3 \longrightarrow \mathbb{Z}^2 \otimes \Lambda^2 \mathbb{Z}^6,$$

munies d’actions de $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$, $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z}) \times \mathrm{SL}_3(\mathbb{Z})$ et $\mathrm{GL}_2(\mathbb{Z}) \times \mathrm{SL}_6(\mathbb{Z})$, respectivement. Restreintes à des sous-espaces convenables et modulo ces actions, elles deviennent des morphismes de groupes. Par exemple, pour R fixé et I, I' projectifs comme R -modules (*i.e.* inversibles), la restriction de la bijection du Théorème 5.2 associe à $(A, B) \in \mathbb{Z}^2 \otimes \mathbb{Z}^3 \otimes \mathbb{Z}^3$ la classe d’idéaux de I . C’est un isomorphisme sur le groupe des classes des R -idéaux inversibles $\mathrm{Cl}(R)$.

5.3. Les cas quartiques et quintiques

La paramétrisation des anneaux quartiques requiert de nouveaux préliminaires.

DÉFINITION 5.4. — *Pour un anneau R de degré n , soit I_R l'idéal de $R^{\otimes n}$ engendré par les éléments de la forme*

$$(x \otimes 1 \otimes \cdots \otimes 1) + (1 \otimes x \otimes \cdots \otimes 1) + \cdots + (1 \otimes 1 \otimes \cdots \otimes x) - \text{Tr}(x) \times (1 \otimes \cdots \otimes 1).$$

La S_n -clôture \hat{R} de R est l'anneau de degré $n!$ donné par $\hat{R} = M/M_{\text{tor}}$, où $M := R^{\otimes n}/I_R$ et M_{tor} est le sous-groupe de torsion de M .

Si R est intègre de corps des fractions K , \hat{R} est la \mathbb{Z} -algèbre engendrée par les conjugués des éléments de R et son corps des fractions $\text{Frac}(\hat{R})$ est une clôture galoisienne de K/\mathbb{Q} . On fixe un plongement $x \mapsto (x \otimes 1 \otimes \cdots \otimes 1)$ de R dans \hat{R} . Les n conjugués de x sont les $(x \otimes 1 \otimes \cdots \otimes 1), \dots, (1 \otimes 1 \otimes \cdots \otimes x)$. Il y a une action naturelle du groupe symétrique S_n sur \hat{R} et $\hat{R}^{S_n} = \mathbb{Z} \otimes \cdots \otimes \mathbb{Z} = \mathbb{Z}$.

DÉFINITION 5.5. — *Soit Q un anneau quartique de S_4 -clôture \hat{Q} . Pour $x \in Q$, on note x, x', x'', x''' ses S_4 -conjugués et on définit $\Phi_{4,3} : Q \rightarrow \hat{Q}$ par $\Phi_{4,3}(x) = xx' + x''x'''$. On note*

$$R^{\text{inv}}(Q) = \mathbb{Z}[\{\Phi_{4,3}(x) : x \in Q\}] \subset \hat{Q}.$$

On démontre que $R^{\text{inv}}(Q)$ est inclus dans un anneau cubique, le sous-anneau de \hat{R} fixe par un sous-groupe $D_4 \subset S_4$ diédral d'ordre 8.

DÉFINITION 5.6. — *Soit Q un anneau quartique. Une résolvante cubique de Q est un anneau cubique R tel que $\text{Disc}(R) = \text{Disc}(Q)$ et $R^{\text{inv}}(Q) \subset R$.*

Bhargava a aussi donné une description plus fonctorielle des résolvantes cubiques qui n'utilise pas la notion de S_n -clôture. L'idée est de voir une résolvante cubique R d'un anneau quartique Q comme un anneau cubique muni d'une application quadratique $\Phi_{4,3} : Q \rightarrow R$ satisfaisant des propriétés formelles convenables.

THÉORÈME 5.7. — *Tout anneau quartique admet au moins une résolvante cubique R , qui est unique si et seulement si Q est « de contenu 1 », c'est-à-dire si Q/\mathbb{Z} n'est pas de la forme $n(Q'/\mathbb{Z})$ pour un $n > 1$ et un anneau quartique Q' . Dans ce cas, $R = R^{\text{inv}}(Q)$.*

Si R est une résolvante cubique de Q , $\Phi_{4,3}$ induit une application quadratique de Q/\mathbb{Z} dans R/\mathbb{Z} , c'est-à-dire de \mathbb{Z}^3 dans \mathbb{Z}^2 aux changements de base près. C'est donc une paire de formes quadratiques ternaires modulo $\Gamma = \text{GL}_3(\mathbb{Z}) \times \text{GL}_2(\mathbb{Z})$. Explicitement $(g_3, g_2) \in \Gamma$ opère sur $(A, B) \in \mathbb{Z}^2 \otimes (\text{Sym}^3 \mathbb{Z}^2)^*$, où A, B sont vues comme matrices symétriques 3×3 à coefficients demi-entiers en dehors de la diagonale, par

$$(g_3, g_2) \cdot (A, B) = (r \cdot g_3 A g_3^t + s \cdot g_3 B g_3^t, t \cdot g_3 A g_3^t + u \cdot g_3 B g_3^t),$$

avec $g_2 = \begin{pmatrix} r & s \\ t & u \end{pmatrix}$. Soit f la forme cubique binaire $4 \det(xA - yB)$ — le facteur 4 assure l'intégralité. La forme f est invariante sous l'action du facteur $\text{GL}_3(\mathbb{Z})$ de Γ , et le facteur $\text{GL}_2(\mathbb{Z})$ agit via $\begin{pmatrix} r & \\ -t & -u \end{pmatrix} \cdot f$. L'action de Γ a un unique invariant $\text{Disc}((A, B)) := \text{Disc}(f)$.

THÉORÈME 5.8. — *Il existe une bijection entre les ensembles suivants :*

- les paires de formes quadratiques ternaires (A, B) de discriminant $\text{Disc}((A, B)) = D$, soit $\mathbb{Z}^2 \otimes (\text{Sym}^2 \mathbb{Z}^3)^*$, modulo l'action de $\text{GL}_3(\mathbb{Z}) \times \text{GL}_2(\mathbb{Z})$,
- les classes d'isomorphismes de triplets (Q, R, Φ) , où Q, R sont des anneaux respectivement quartique et cubique, $\Phi : Q/\mathbb{Z} \rightarrow R/\mathbb{Z}$ est une application quadratique, R est une résolvante cubique de Q et $\text{Disc } Q = \text{Disc } R = D$. La classe d'isomorphisme de R est donnée par la $\text{GL}_2(\mathbb{Z})$ -orbite de la forme cubique binaire $f(x, y) = 4 \det(xA - yB)$.

On notera $Q(A, B)$ l'anneau quartique associé à la paire (A, B) . Ce théorème est un analogue du cas cubique, qui associe à un anneau cubique R de discriminant D , une résolvante quadratique $S(D)$ et une application cubique $\Phi_{3,2} : R/\mathbb{Z} \rightarrow S/\mathbb{Z}$. Ces deux données étant entièrement déterminées par R , elles n'apparaissent pas explicitement. Dans la paramétrisation ci-dessus, Φ est en fait déterminée par Q et R si leur discriminant commun est non-nul. De surcroît, comme nous l'avons vu au Théorème 5.7, Q de contenu 1 détermine R .

Ce théorème se démontre « explicitement » comme le Théorème 4.6, la grande difficulté venant du fait qu'il n'y a pas de choix naturel pour la résolvante cubique R . Pour un choix particulier d'un représentant (A, B) dans la classe modulo $\text{GL}_3(\mathbb{Z}) \times \text{GL}_2(\mathbb{Z})$, Bhargava montre que la structure multiplicative de Q et R est fixée, en écrivant explicitement des lois de multiplication génériques analogues à (17), et en résolvant les conditions de compatibilité, associativité, commutativité, au terme d'un processus où le mot « miracle » apparaît plusieurs fois [2]. On vérifie que R est bien une résolvante cubique de Q , et que (Q, R) a bien pour image (A, B) .

Bhargava [3, 4] annonce d'autres paramétrisations, dont le résultat suivant :

THÉORÈME 5.9. — *Il existe une bijection entre les ensembles suivants :*

- les quadruplets de 2-formes alternées de rang 5, soit $\mathbb{Z}^4 \otimes \Lambda^2 \mathbb{Z}^5$, modulo l'action de $\text{GL}_4(\mathbb{Z}) \times \text{SL}_5(\mathbb{Z})$,
- les paires (P, \mathcal{S}) , où P est une classe d'isomorphisme d'anneaux quintiques, et \mathcal{S} est une résolvante sextique de P .

Cette bijection préserve le discriminant.

6. COMPTAGES PAR DISCRIMINANT ET DENSITÉS

6.1. La conjecture de Malle

Soit k un corps de nombres dont on fixe une clôture algébrique \bar{k} , $G \subset S_n$ un groupe de permutations sur n lettres et K/k une extension finie de degré $[K : k] = n$. Par abus de notation, on écrit « $\text{Gal}(K/k) = G$ » si le groupe de Galois de la clôture galoisienne

de K/k , vu comme groupe de permutation sur les n k -plongements de K dans \bar{k} , est isomorphe à G . Soit

$$\mathcal{F}_{n,k}(X, G) = \{K/k: K \subset \bar{k}, \text{Gal}(K/k) = G, N_{k/\mathbb{Q}}(d_{K/k}) \leq X\} / \text{Gal}(\bar{k}/k),$$

l'ensemble des classes d'isomorphismes d'extensions de k de groupe de Galois G , au sens précédent, dont la norme du discriminant relatif $d_{K/k}$ est bornée par X . On note

$$N_{n,k}(X, G) := \sum_{K \in \mathcal{F}_{n,k}(X, G)} \frac{1}{\#\text{Aut}_k K}.$$

La pondération par le nombre d'automorphismes, qui ne dépend que de k et G , nous permettra de formuler plus naturellement certaines densités.

DÉFINITION 6.1. — Pour $\sigma \in S_n$, on définit l'indice de σ par

$$\text{ind}(\sigma) := n - \#\{\text{cycles de } \sigma\}$$

(il ne dépend que de la classe de conjugaison de σ). Si $G \neq \{\text{Id}\}$ est un sous-groupe de S_n , on note

$$\text{ind}(G) := \min_{\sigma \in G - \{\text{Id}\}} \text{ind}(\sigma), \quad a(G) = 1/\text{ind}(G) \in]0, 1].$$

On note $b(G, k) \geq 1$ le nombre de k -classes de conjugaison de G , c'est-à-dire de classes modulo l'action naturelle de $\text{Gal}(\bar{k}/k)$, dont l'indice est $\text{ind}(G)$.

En particulier, $a(G) = 1$ si et seulement si G contient une transposition ; dans ce cas, $b(G, k) = 1$ pour tout corps de nombres k (Malle [38]). Malle [38, 39] conjecture une estimation relativement précise pour $N_{n,k}(X, G)$:

CONJECTURE 6.2. — Soit G un groupe de permutation transitif et k un corps de nombres. Il existe une constante $c(G, k) > 0$ telle que

$$N_{n,k}(X, G) \sim c(G, k) X^{a(G)} (\log X)^{b(G,k)-1}.$$

On en déduit que le nombre $N_{n,k}(X) = \sum_{G \subset S_n} N_{n,k}(X, G)$ d'extensions de k de degré n dont on ne fixe plus le groupe de Galois vérifierait

$$N_{n,k}(X) \stackrel{?}{\sim} X \cdot \sum_{\substack{G \subset S_n \\ a(G)=1}} c(G, k).$$

Bien sûr, cette conjecture implique une solution positive au problème de Galois inverse. Malle conjecture qu'en fixant la structure de k_v -algèbre de $K \otimes k_v$ pour un nombre fini de places v de k , on garde les mêmes exposants (mais la constante change) pour peu qu'au moins une telle extension K existe. Les $c(G, k)$ ne sont pas précisés par la conjecture.

Celle-ci est démontrée si G est abélien⁽⁵⁾ (Mäki [37], Wright [55]) ou de petit cardinal : $n = 3$ et $G = S_3$ (Davenport et Heilbronn [21], Datskovsky et Wright [18]), $n = 4$ et $G = D_4$ (Cohen, Diaz y Diaz, Olivier [14] ; malgré le titre, le cas $k \neq \mathbb{Q}$ est traité). Voir [11] pour un survol plus détaillé.

Citons pour finir trois résultats récents. Pour un groupe G nilpotent en représentation régulière, Klüners et Malle [35] obtiennent la forme faible

$$\lim_{X \rightarrow +\infty} \frac{\log N_{|G|,k}(X, G)}{\log X} = a(G).$$

Klüners [34] a récemment annoncé que les groupes quaternioniens généralisés

$$Q_{4m} := \langle x, y \mid x^{2m} = 1, y^2 = x^m, y^{-1}xy = x^{-1} \rangle, \quad m = 2^l, l \geq 1,$$

vérifient la conjecture sur $k = \mathbb{Q}$, fournissant les premiers exemples non abéliens d'ordre arbitrairement élevé ; ici $G = Q_{4m} \subset S_{4m}$ est en représentation régulière, $a(G) = 2m$ et $b(G, \mathbb{Q}) = 1$. Le résultat le plus général à ce jour, dû à Ellenberg et Venkatesh [25] dit que pour tout $\varepsilon > 0$, tout $n \geq 1$, et tout corps de nombres k , on a

$$\limsup_{X \rightarrow +\infty} \frac{\log N_{n,k}(X)}{\log X} \ll_{\varepsilon} n^{\varepsilon} \quad \text{et} \quad \liminf_{X \rightarrow +\infty} \frac{\log N_{n,k}(X)}{\log X} \geq \frac{1}{2} + \frac{1}{n^2}$$

6.2. La conjecture de Bhargava

Pour v une place de k et $n \geq 1$, on définit

$$a_v(n) := \sum_{\substack{A \text{ étale}/k_v \\ [A:k_v]=n}} \frac{|d_{A/k_v}|_v}{\#\text{Aut}_{k_v} A}$$

où la somme porte sur les classes d'isomorphismes d'algèbres étales de degré n sur le complété k_v . Ici, d_{A/k_v} est le discriminant associé (1 pour v archimédienne), et $|\cdot|_v$ désigne la valeur absolue normalisée habituelle.

CONJECTURE 6.3 (Bhargava). — Pour $n \geq 2$, on a $N_{n,k}(X, S_n) \sim c(S_n, k)X$ quand $X \rightarrow +\infty$, avec

$$(20) \quad c(S_n, k) = \frac{1}{2} \text{Res}(\zeta_k, s = 1) \prod_v a_v(n)(1 - 1/Nv)$$

où ζ_k est la fonction zêta de Dedekind de k , et où on omet le facteur $(1 - 1/Nv)$ pour v archimédienne.

⁽⁵⁾Si G est abélien de cardinal n , $G \subset S_n$, p le plus petit diviseur premier de n , n_p le nombre d'éléments d'ordre p dans G , on montre que

$$1/a(G) = |G|(1 - 1/p) \quad \text{et} \quad b(G, k) = n_p/[k(e^{2i\pi/p}) : k].$$

(Il s'agit d'une reformulation d'une conjecture équivalente donnée par Bhargava [4].)

Comme les algèbres étales sur k_v sont les produits finis de corps sur k_v et que leurs k_v -automorphismes et discriminants sont les produits de ceux de leurs composantes, on obtient l'expression de la série génératrice

$$(21) \quad \sum_n a_v(n)T^n = \exp \left(\sum_{K \text{ corps}/k_v} \frac{|d_{K/k_v}|_v}{\#\text{Aut}_{k_v} K} T^{[K:k_v]} \right).$$

THÉORÈME 6.4. — On a

$$\sum_n a_v(n)T^n = \begin{cases} \exp(T) & \text{si } v \text{ est complexe,} \\ \exp \left(T + \frac{1}{2}T^2 \right) & \text{si } v \text{ est réelle,} \\ \prod_{k=1}^{+\infty} \frac{1}{1 - q^{k-1}T^k} & \text{si } v \text{ est finie, } q = (\mathbb{N}v)^{-1}. \end{cases}$$

Preuve. — Les deux premiers cas sont clairs à partir de (21), le troisième résulte de la formule de masse de Serre [47], sous la forme

$$\sum_{\substack{K \text{ corps}/k_v \\ [K:k_v]=n}} \frac{|d_{K/k_v}|_v}{\#\text{Aut}_{k_v} K} = \sum_{d|n} \frac{q^{n-d}}{d}.$$

La formule donnée dans [47] énonce que la somme sur les extensions totalement ramifiées est q^{n-1} . Pour chaque $d | n$, on l'applique à l'unique extension non ramifiée de degré d de k_v , pour compter ses extensions totalement ramifiées de degré $\frac{n}{d}$. \square

COROLLAIRE 6.5. — Si v est finie, $q = (\mathbb{N}v)^{-1}$, alors $a_v(n) = P_n(q)$ où $P_n \in \mathbb{N}[X]$ ne dépend pas de v . On a $P_n(q) = 1 + q + O(q^2)$ quand $q \rightarrow 0$ et le produit infini (20) converge.

Exemple 6.6. — Pour v réelle,

$$a_v(2) = 1, \quad a_v(3) = 2/3, \quad a_v(4) = 5/12, \quad a_v(5) = 13/60.$$

Pour v finie, $q = (\mathbb{N}v)^{-1}$, on obtient

$$\begin{aligned} (1-q)a_v(2) &= 1 - q^2, & (1-q)a_v(4) &= 1 + q^2 - q^3 - q^4, \\ (1-q)a_v(3) &= 1 - q^3, & (1-q)a_v(5) &= 1 + q^2 - q^4 - q^5. \end{aligned}$$

Exemple 6.7. — Si $n = 2$, la conjecture prédit

$$N_{\mathbb{Q},2}(X, S_2)/X \rightarrow \frac{1}{2\zeta(2)},$$

ce qui est un résultat classique. On le démontre par exemple à partir de l'identité $\mu^2(q) = \sum_{d^2|q} \mu(d)$ d'où on déduit la densité des entiers sans facteur carré. (Ne pas oublier que chaque corps est pondéré par $1/\#\text{Aut}_{\mathbb{Q}}(K) = 1/2$.)

La conjecture suit du principe heuristique suivant, analogue à celui de Cohen-Lenstra [12] et Cohen-Martinet [13] sur le comportement moyen des groupes de classes. On considère le nombre d’extensions K/k de degré n , avec $\text{Gal}(K/k) = S_n$ et $d_{K/k} = D$. En se demandant quelles collections de $A = K \otimes k_v$ peuvent intervenir et en supposant les comportements indépendants aux différentes places, on s’attend à ce qu’il soit égal à

$$\prod_v \sum_{\substack{A \text{ étale}/k_v \\ [A:k_v]=n \\ |d_{A/k_v}|_v = |D|_v}} \frac{1}{\#\text{Aut}_{k_v} A},$$

en moyenne sur D . La conjecture est vraie pour $n = 2$ (Wright [55], Cohen-Diaz y Diaz-Olivier [15]), $n = 3$ (Datskovsky-Wright [18]), et correspond aux valeurs annoncées par Bhargava sur \mathbb{Q} pour $n = 4, 5$. Une variation évidente fixe la structure de k_v -algèbre de $K \otimes k_v$ pour un nombre fini de places, par exemple la signature à l’infini : on remplace les $a_v(n)$ correspondants par la somme sur les algèbres de structure permise. Cette conjecture renforcée est vérifiée dans les mêmes cas que ci-dessus (voir [18, §4] pour $n = 2, 3$).

6.3. S_3 sur \mathbb{Q}

Ce cas, réinterprété et étendu dans le langage du §3.3 par Datskovsky et Wright [54, 17, 18], est originellement traité sur \mathbb{Q} par Davenport et Heilbronn [21] avec la méthode du domaine fondamental du §2. À partir du Théorème 5.1, un équivalent de $N_{\mathbb{Q},3}(X, S_3)$ s’obtient assez simplement : on compte les classes de formes cubiques associées aux ordres maximaux, une condition locale qui prend la forme suivante.

THÉORÈME 6.8 (Davenport-Heilbronn). — *La classe d’une forme cubique irréductible F est associée à un ordre R maximal en p (c’est-à-dire tel que $R \otimes \mathbb{Z}_p$ est maximal) si et seulement si $p \nmid F$ et F n’est pas $\text{GL}_2(\mathbb{Z})$ -équivalente à une forme (a, b, pc, p^2d) , $a, b, c, d \in \mathbb{Z}$.*

Ce n’est pas la formulation de Davenport et Heilbronn [21], qui ne connaissaient pas le résultat de Delone et Faddeev et faisaient des calculs locaux désagréables, mais elle lui est équivalente.

Preuve. — Si $p \mid F$, R est inclus dans l’ordre associé à F/p , donc non maximal en p . Si p ne divise pas le discriminant de F , égal à $\text{Disc } R$, il n’y a pas de problème. Sinon, en remplaçant au besoin F par une forme $\text{GL}_2(\mathbb{Z})$ -équivalente, on peut supposer que la racine double de F modulo p est en 0, et qu’il n’y a pas de racine à l’infini soit $F = (a, b, pc, pd)$, avec $p \nmid a$. Le résultat suit du critère de p -maximalité de Dedekind (cf. [10, §6.1.4]).

Explicitement, si $p \mid d$, l’ordre associé à $(pa, b, c, d/p)$ contient R (voir (19)). □

Posons $G = \mathrm{GL}_2$, $V = \mathrm{Sym}^3 \mathbb{A}^2$; il suffit de compter les points de $V_{\mathbb{Z}}/G_{\mathbb{Z}}$ de discriminant borné, satisfaisant les congruences ci-dessus. En effet, les points associés à des anneaux non intègres ou de corps des fractions cubiques cycliques sont absorbés dans un terme d'erreur. On trouve

$$\mathrm{Vol}(V_{\mathbb{R}}/G_{\mathbb{Z}} \cap \{F : |\mathrm{Disc} F| < X\}) = \left(\frac{1}{2} + \frac{1}{6}\right) \frac{\zeta(2)}{2} X = \frac{\zeta(2)}{3} X,$$

en séparant les deux composantes connexes de $(V - S)_{\mathbb{R}} = V^+ \cup V^-$ associées aux formes de discriminant respectivement positif ou négatif⁽⁶⁾. Le produit des densités p -adiques provenant du Théorème 6.8 est $1/(\zeta(2)\zeta(3))$, et l'on obtient

$$N_{\mathbb{Q},3}(X, S_3)/X \rightarrow \frac{1}{3\zeta(3)} = \frac{1}{2} \mathrm{Res}(\zeta, s=1) a_{\infty}(3) \prod_p (1 - 1/p) a_p(3).$$

On peut bien sûr séparer les densités des corps totalement réels et totalement complexes, la première étant trois fois plus faible que la seconde, en se restreignant à V^+ ou V^- au lieu d'additionner leurs contributions.

6.4. S_4 et S_5 sur \mathbb{Q}

Pour ce qui concerne S_4 et S_5 sur \mathbb{Q} , seul le cas des corps S_4 totalement réels est publié [2] à ce jour. Bhargava [4] annonce une solution complète qui corrobore sa conjecture pour $n = 4, 5$.

DÉFINITION 6.9. — *Une paire de formes quadratiques ternaires (A, B) est totalement réelle, resp. mixte, resp. totalement complexe si (A, B) a 4, resp. 2, resp. 0 zéros réels dans $\mathbb{P}^2(\mathbb{R})$. Elle est irréductible si A et B n'ont pas de zéro commun dans $\mathbb{P}^2(\mathbb{Q})$ et si la forme cubique binaire $\det(xA - yB)$ est irréductible sur \mathbb{Q} . Une paire (A, B) est maximale si l'anneau quartique $Q(A, B)$ est maximal.*

Toutes ces définitions ne dépendent que de la classe de (A, B) modulo $\mathrm{GL}_3(\mathbb{Z}) \times \mathrm{GL}_2(\mathbb{Z})$ et on a une description locale, analogue au Théorème 6.8, des paires (A, B) maximales.

THÉORÈME 6.10. — *Soit $Q(A, B)$ l'anneau quartique associé à une classe de paires de formes quadratiques (A, B) :*

- (A, B) est irréductible si et seulement si $Q(A, B)$ est un ordre d'une extension quartique de \mathbb{Q} de clôture galoisienne A_4 ou S_4 .
- (A, B) est totalement réelle si et seulement si Q est totalement réel ($Q \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{R}^4$).
- (A, B) est totalement réelle irréductible maximale si et seulement si $Q(A, B)$ est un ordre quartique maximal, dont le corps des fractions est totalement réel, de clôture galoisienne A_4 ou S_4 sur \mathbb{Q} .

⁽⁶⁾ V^+ (resp. V^-) correspond aux formes ayant trois racines réelles (resp. une seule racine réelle), dont le stabilisateur dans $\mathrm{GL}_2(\mathbb{R})$ est isomorphe à S_3 (resp. à S_2). D'où les coefficients $1/6$ et $1/2$.

– Le produit des densités locales associées aux (A, B) maximales est

$$\frac{1}{\zeta(2)^2 \zeta(3)} \prod_p (1 + p^{-2} - p^{-3} - p^{-4}).$$

On trouve cette fois-ci

$$\text{Vol}(V_{\mathbb{R}}/G_{\mathbb{Z}} \cap \{(A, B) : |\text{Disc}(A, B)| < X\}) = \left(\frac{1}{24} + \frac{1}{4} + \frac{1}{8} \right) \frac{\zeta(2)^2 \zeta(3)}{2} X,$$

où $V_{\mathbb{R}} = \mathbb{R}^2 \otimes \text{Sym}^2 \mathbb{R}^3$, $G_{\mathbb{Z}} = \text{GL}_2(\mathbb{Z}) \times \text{GL}_3(\mathbb{Z})$, et les coefficients $1/24$, $1/4$, $1/8$ proviennent respectivement des cas réels, mixtes et complexes (stabilisateurs S_4 , S_2^2 et S_2^3) qui sont les trois composantes connexes de $(V - S)_{\mathbb{R}}$.

THÉORÈME 6.11. — *Conformément à la conjecture de Bhargava, le nombre de classes d'isomorphismes de corps quartiques totalement réels de discriminant inférieur à X est équivalent à CX , où*

$$C = \frac{1}{2} \times \frac{1}{24} \times \prod_p a_4(p)(1 - 1/p).$$

Preuve. — Ce théorème suit formellement de ce qui précède, en montrant que sont négligeables

- le nombre d'extensions A_4 de discriminant inférieur à X ,
- les points associés aux (A, B) réductibles,
- le terme reste algébrique lié au passage à la limite dans le produit des facteurs locaux (analogue de (4)),
- le terme géométrique provenant du principe de Lipschitz (Théorème 2.2), lié aux volumes des projections d'un domaine fondamental.

Wong [53, Remark 9] démontre que $N_{\mathbb{Q},4}(X, A_4) = O(X^{7/8+\varepsilon})$ pour tout $\varepsilon > 0$, et les trois autres points sont démontrés par Bhargava. Le deuxième est de loin le plus délicat, faisant hélas appel à un domaine fondamental explicite et à de lourdes estimations directes (en dimension 12). Parmi les points entiers du domaine fondamental, les paires réductibles sont en fait largement majoritaires ; elles sont heureusement concentrées sur quelques sections hyperplanes qui peuvent être éliminées avant le décompte par principe de Lipschitz. \square

L'hypothèse « totalement réel » et le domaine fondamental explicite qu'elle permet sont utilisés pour traiter un cas dégénéré mais crucial. Je ne sais pas si Bhargava les supprime pour traiter les deux autres signatures possibles ou bien s'il doit refaire tous les calculs.

Si K est un corps de nombres et p un nombre premier, soit

$$r_p(K) = \dim_{\mathbb{F}_p} (\text{Cl}(K) \otimes_{\mathbb{Z}} (\mathbb{Z}/p\mathbb{Z}))$$

le p -rang du groupe des classes de K ; ainsi, $p^{r_p(K)}$ est le nombre d'éléments de p -torsion du groupe des classes, et $(p^{r_p(K)} - 1)/(p - 1)$ le nombre de sous-groupes d'indice p . La méthode de Davenport et Heilbronn du §6.3 permet la détermination de l'ordre moyen

de $3^{r_3(K)}$ quand K parcourt les corps quadratiques, et la vérification de l'heuristique de Cohen-Lenstra dans ce cadre. Bhargava obtient l'analogie suivant :

THÉORÈME 6.12. — *Conformément aux heuristiques de Cohen-Martinet, on a*

$$\sum_K (2^{r_2(K)} - 1) \sim \frac{1}{4} \sum_K 1, \quad \text{quand } X \rightarrow \infty,$$

où K parcourt les corps cubiques totalement réels de discriminant inférieur à X .

Preuve. — Soit K_4 un corps quartique de type A_4 ou S_4 , de clôture galoisienne L , $K_3 \subset L$ un corps cubique, et K_6 l'unique extension quadratique de K_3 dont la clôture galoisienne soit L . Alors K_6/K_3 est ramifiée en une place divisant p si et seulement si le type de ramification de p dans K_4 est (1^4) , (2^2) , ou $(1^2 2^2)$. Le corps K_4 est totalement réel si et seulement si L l'est, ce qui implique que K_3 et K_6 le sont aussi.

Réciproquement si K_3 est un corps cubique totalement réel et si K_6/K_3 est une extension quadratique non ramifiée, alors la clôture galoisienne L de K_6 est de type A_4 ou S_4 et contient un unique corps quartique K_4 à conjugaison près, et K_4 est totalement réel puisque K_6 et donc L le sont. Par théorie du corps de classes, $2^{r_2(K)} - 1$ est le nombre d'extensions quadratiques non ramifiées K_6 de K_3 . On vérifie par ailleurs que $\text{Disc } K_3 = \text{Disc } K_4$. Le membre de gauche compte donc les extensions K_4 totalement réelles de type A_4 ou S_4 et de ramification restreinte comme ci-dessus.

Bhargava démontre que, pour $Q(A, B)$ maximal, la structure de l'anneau quotient $Q(A, B)/(p)$ est donnée par les degrés résiduels aux points d'intersections des deux coniques définies par A et B sur \mathbb{F}_p . (Le résultat reste vrai pour $Q(A, B)$ non-maximal, s'il n'y a pas de point d'intersection quadruple.) En particulier le type de décomposition d'un premier p dans le corps des fractions K_4 de $Q(A, B)$ se voit par des congruences modulo p sur (A, B) . En renforçant la condition locale de p -maximalité de la démonstration précédente pour imposer l'absence de ramification de type (1^4) , (2^2) ou $(1^2 2^2)$, on obtient un équivalent du membre de gauche. Le théorème de Davenport-Heilbronn démontré au §6.3 donne un équivalent du membre de droite. \square

RÉFÉRENCES

- [1] K. BELABAS, Crible et 3-rang des corps quadratiques, *Ann. de l'Inst. Fourier* **46** (1996), pp. 909–949.
- [2] M. BHARGAVA, Higher composition laws, Ph.D. thesis, Princeton University, 2001.
- [3] M. BHARGAVA, Gauss composition and generalizations, in *Ants V, Sydney*, LNCS, no. 2369, Springer-Verlag, 2002, pp. 1–9.
- [4] M. BHARGAVA, The parametrization of algebraic structures, 2003, *Explicit Methods in Number Theory* (Oberwolfach). Exposé.
- [5] M. BHARGAVA, Higher composition laws. I. A new view on Gauss composition, and quadratic generalizations, *Ann. of Math. (2)* **159** (2004), no. 1, pp. 217–250.

- [6] M. BHARGAVA, Higher composition laws. II. On cubic analogues of Gauss composition, *Ann. of Math. (2)* **159** (2004), no. 2, pp. 865–886.
- [7] M. BHARGAVA, Higher composition laws. III, *Ann. of Math. (2)* **159** (2004), no. 3.
- [8] D. A. BUELL, *Binary quadratic forms*, Springer-Verlag, 1989.
- [9] F. CHAMIZO & H. IWANIEC, On the Gauss mean-value formula for class number, *Nagoya Math. J.* **151** (1998), pp. 199–208.
- [10] H. COHEN, *A course in computational algebraic number theory*, third ed., Springer-Verlag, 1996.
- [11] H. COHEN, Constructing and counting number fields, in *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, Higher Ed. Press, 2002, pp. 129–138.
- [12] H. COHEN & H. W. LENSTRA, JR., Heuristics on class groups of number fields, in *Number theory, Noordwijkerhout 1983* (Berlin), Lecture Notes in Math., vol. 1068, Springer, Berlin, 1984, pp. 33–62.
- [13] H. COHEN & J. MARTINET, Études heuristiques des groupes de classes des corps de nombres, *J. reine angew. Math.* **404** (1990), pp. 39–76.
- [14] H. COHEN, F. DIAZ Y DIAZ, & M. OLIVIER, Enumerating quartic dihedral extensions of \mathbb{Q} , *Compositio Math.* **133** (2002), no. 1, pp. 65–93.
- [15] H. COHEN, F. DIAZ Y DIAZ, & M. OLIVIER, On the density of discriminants of cyclic extensions of prime degree, *J. Reine Angew. Math.* **550** (2002), pp. 169–209.
- [16] D. COX, *Primes of the form $x^2 + ny^2$* , Wiley-Interscience, 1989.
- [17] B. DATSKOVSKY & D. J. WRIGHT, The adelic zeta function associated to the space of binary cubic forms. II. Local theory, *J. Reine Angew. Math.* **367** (1986), pp. 27–75.
- [18] B. DATSKOVSKY & D. J. WRIGHT, Density of discriminants of cubic extensions, *J. Reine Angew. Math.* **386** (1988), pp. 116–138.
- [19] B. A. DATSKOVSKY, On Dirichlet series whose coefficients are class numbers of binary quadratic forms, *Nagoya Math. J.* **142** (1996), pp. 95–132.
- [20] H. DAVENPORT, On a principle of Lipschitz, *J. Lond. Math. Soc.* **26** (1951), pp. 179–183, corrigendum *idid* **39** (1964), p. 580.
- [21] H. DAVENPORT & H. HEILBRONN, On the density of discriminants of cubic fields (ii), *Proc. Roy. Soc. Lond. A* **322** (1971), pp. 405–420.
- [22] B. N. DELONE & D. K. FADDEEV, *The theory of irrationalities of the third degree*, Translations of Mathematical Monographs, vol. 10, American Mathematical Society, 1964.
- [23] P. G. L. DIRICHLET, *Vorlesungen über Zahlentheorie*, Herausgegeben und mit Zusätzen versehen von R. Dedekind. Vierte, umgearbeitete und vermehrte Auflage, Chelsea Publishing Co., New York, 1968.
- [24] G. EISENSTEIN, Untersuchungen über die cubischen Formen mit zwei Variabeln, *J. reine angew. Math.* **27** (1844), pp. 89–104.

- [25] J. ELLENBERG & A. VENKATESH, The number of extensions of a number field with fixed degree and bounded discriminant, preprint.
- [26] W. T. GAN, B. GROSS, & G. SAVIN, Fourier coefficients of modular forms on G_2 , *Duke Math. J.* **115** (2002), no. 1, pp. 105–169.
- [27] D. GOLDFELD & J. HOFFSTEIN, Eisenstein series of $\frac{1}{2}$ -integral weight and the mean value of real Dirichlet L -series, *Invent. Math.* **80** (1985), no. 2, pp. 185–208.
- [28] J. W. HOFFMAN & J. MORALES, Arithmetic of binary cubic forms, *Enseign. Math. (2)* **46** (2000), no. 1-2, pp. 61–94.
- [29] M. N. HUXLEY, Integer points, exponential sums and the Riemann zeta function, in *Number theory for the millennium, II (Urbana, IL, 2000)*, A K Peters, 2002, pp. 275–290.
- [30] J.-I. IGUSA, *An introduction to the theory of local zeta functions*, AMS/IP Studies in Advanced Mathematics, vol. 14, American Mathematical Society, Providence, RI, 2000.
- [31] A. C. KABLE & A. YUKIE, Prehomogeneous vector spaces and field extensions. II, *Invent. Math.* **130** (1997), no. 2, pp. 315–344.
- [32] M. KASHIWARA, B -functions and holonomic systems. Rationality of roots of B -functions, *Invent. Math.* **38** (1976/77), no. 1, pp. 33–53.
- [33] T. KIMURA, The b -functions and holonomy diagrams of irreducible regular prehomogeneous vector spaces, *Nagoya Math. J.* **85** (1982), pp. 1–80.
- [34] J. KLÜNERS, On the asymptotics of number fields with given Galois group, 2003, *Explicit Methods in Number Theory* (Oberwolfach). Exposé.
- [35] J. KLÜNERS & G. MALLE, Counting nilpotent Galois extensions, *J. reine angew. Math.*, à paraître.
- [36] R. LIPSCHITZ, Über die asymptotischen Gesetze von gewissen Gattungen zahlen-theoretischer Funktionen, *Monatsber. der Berl. Acad* (1865), §174sqq.
- [37] S. MÄKI, On the density of abelian number fields, *Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes* (1985), no. 54.
- [38] G. MALLE, On the distribution of Galois groups, *J. Number Theory* **92** (2002), no. 2, pp. 315–329.
- [39] G. MALLE, On the distribution of Galois groups, II, 2002, preprint.
- [40] D. P. ROBERTS, Density of cubic field discriminants, *Math. Comp.* **70** (2001), no. 236, pp. 1699–1705.
- [41] H. RUBENTHALER, Formes réelles des espaces préhomogènes irréductibles de type parabolique, *Ann. Inst. Fourier (Grenoble)* **36** (1986), no. 1, pp. 1–38.
- [42] H. SAITO, On a classification of prehomogeneous vector spaces over local and global fields, *J. Algebra* **187** (1997), no. 2, pp. 510–536.
- [43] H. SAITO, Convergence of the zeta functions of prehomogeneous vector spaces, *Nagoya Math. J.* **170** (2003), pp. 1–31.
- [44] M. SATO & T. KIMURA, A classification of irreducible prehomogeneous vector spaces and their relative invariants, *Nagoya Math. J.* **65** (1977), pp. 1–155.

- [45] M. SATO & T. SHINTANI, On zeta functions associated with prehomogeneous vector spaces, *Ann. of Math.* **100** (1974), pp. 131–170.
- [46] M. SATO, Theory of prehomogeneous vector spaces (algebraic part)—the English translation of Sato’s lecture from Shintani’s note, *Nagoya Math. J.* **120** (1990), pp. 1–34, traduit du japonais par M. Muro.
- [47] J.-P. SERRE, Une “formule de masse” pour les extensions totalement ramifiées de degré donné d’un corps local, *C. R. Acad. Sci. Paris Sér. A-B* **286** (1978), no. 22, pp. A1031–A1036.
- [48] D. SHANKS, On Gauss and composition. I, II, in *Number theory and applications (Banff, AB, 1988)*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 265, Kluwer Acad. Publ., 1989, pp. 163–178, 179–204.
- [49] T. SHINTANI, On Dirichlet series whose coefficients are class numbers of integral binary cubic forms, *J. Math. Soc. Japan* **24** (1972), pp. 132–188.
- [50] T. SHINTANI, On zeta-functions associated with the vector space of quadratic forms, *J. Fac. Sci. Univ. Tokyo, Sec. Ia* **22** (1975), pp. 25–66.
- [51] C. L. SIEGEL, The average measure of quadratic forms with given determinant and signature, *Ann. of Math. (2)* **45** (1944), pp. 667–685.
- [52] È. B. VINBERG, The classification of nilpotent elements of graded Lie algebras, *Dokl. Akad. Nauk SSSR* **225** (1975), no. 4, pp. 745–748.
- [53] S. WONG, Automorphic forms on $GL(2)$ and the rank of class groups, *J. Reine Angew. Math.* **515** (1999), pp. 125–153.
- [54] D. J. WRIGHT, The adelic zeta function associated to the space of binary cubic forms. I. Global theory, *Math. Ann.* **270** (1985), no. 4, pp. 503–534.
- [55] D. J. WRIGHT, Distribution of discriminants of abelian extensions, *Proc. London Math. Soc. (3)* **58** (1989), no. 1, pp. 17–50.
- [56] D. J. WRIGHT & A. YUKIE, Prehomogeneous vector spaces and field extensions, *Invent. Math.* **110** (1992), no. 2, pp. 283–314.
- [57] A. YUKIE, *Shintani zeta functions*, London Mathematical Society Lecture Note Series, vol. 183, Cambridge University Press, Cambridge, 1993.
- [58] D. ZAGIER, Cubic forms and cubic rings, 2001, *Explicit Methods in Number Theory* (Oberwolfach). Exposé.

Karim BELABAS

Université Paris-Sud

Département de Mathématiques (Bât. 425)

F-91405 Orsay (France)

Karim.Belabas@math.u-psud.fr

CAPACITÉ ANALYTIQUE ET LE PROBLÈME DE PAINLEVÉ

par HERVÉ PAJOT

1. INTRODUCTION

Nous dirons qu'un compact E du plan complexe est *effaçable* pour les fonctions holomorphes bornées si pour tout ouvert $U \subset \mathbb{C}$ contenant E , toute fonction holomorphe bornée $f : U \setminus E \rightarrow \mathbb{C}$ admet une extension analytique dans U tout entier. Le *problème de Painlevé* consiste à donner une caractérisation métrique et/ou géométrique de ces ensembles effaçables. À ma connaissance, cette question est explicitement posée pour la première fois dans un article de Lars Ahlfors [1] en 1947. Dans ce papier, il introduit la capacité analytique $\gamma(E)$ définie par

$$\gamma(E) = \sup\{|f'(\infty)|; f : \mathbb{C} \setminus E \rightarrow \mathbb{C} \text{ est analytique et bornée avec } \|f\|_\infty \leq 1\}$$

(où $f'(\infty) = \lim_{z \rightarrow \infty} z(f(z) - f(\infty))$) et il démontre que le compact E est effaçable si et seulement si $\gamma(E) = 0$. Cependant, comme l'écrit Ahlfors lui-même, cette réponse n'est pas satisfaisante, dans la mesure où l'étude de la capacité analytique n'est pas aisée. Avant la caractérisation des ensembles effaçables donnée par X. Tolsa [47] en 2003, les résultats connus étaient les suivants. Dans les faits 1, 2, 3 et 4, E est un compact du plan complexe.

Fait 1 : Si $H^1(E) = 0$ (où H^1 désigne la mesure de Hausdorff de dimension 1), alors E est effaçable.

Fait 2 : Si la dimension de Hausdorff de E (que nous noterons $\text{Hdim}(E)$) est strictement plus grande que 1, alors E n'est pas effaçable.

Les faits 1 et 2 traduisent l'idée intuitive que la propriété d'être effaçable ou non devrait dépendre de la taille de l'ensemble. Ainsi, d'après un théorème de Riemann, un singleton est effaçable. D'un autre côté, grâce au théorème de représentation de Riemann, nous pouvons facilement nous convaincre qu'un continuum (c'est-à-dire un ensemble compact et connexe) du plan complexe qui n'est pas réduit à un point ne l'est pas. Cependant, des exemples dus à Vitushkin, Ivanov et Garnett montrent

que les seules considérations de taille ne suffisent pas à caractériser les ensembles effaçables. Les propriétés de rectifiabilité, c'est-à-dire d'approximation par des courbes lipschitziennes, doivent aussi jouer un rôle. Remarquons que les faits 1 et 2 impliquent que le cas «intéressant» est celui des compacts du plan de dimension de Hausdorff 1 qui satisfont à $H^1(E) > 0$.

Fait 3 : Supposons que E soit contenu dans une courbe lipschitzienne Γ . Alors, E est effaçable si et seulement si $H^1(E) = 0$.

Fait 4 : Si $H^1(E) < +\infty$, alors E est effaçable si et seulement si $H^1(E \cap \Gamma) = 0$ pour toute courbe lipschitzienne Γ .

Notons que le fait 1 implique le sens indirect du fait 3 et que celui-ci implique le sens direct du fait 4. Les démonstrations des faits 3 et 4 sont beaucoup plus difficiles que celles des faits 1 et 2. Alors que ces énoncés semblent plutôt liés à l'analyse complexe, leurs preuves requièrent des techniques fines d'analyse réelle (théorie des intégrales singulières de Calderón-Zygmund, en particulier dans les espaces non doublants) et de théorie de la mesure géométrique (problème du voyageur de commerce géométrique de Peter Jones, théorie de la rectifiabilité uniforme de David et Semmes). La condition de finitude de la mesure de Hausdorff dans l'énoncé du fait 4 est nécessaire. La solution proposée par Tolsa se passe de cette condition et donne donc une caractérisation complète des ensembles effaçables. Les progrès sur le problème de Painlevé ont été relativement lents. Il faut souligner que l'intérêt pour la capacité analytique a été relancé par les travaux de A.G. Vitushkin en théorie de l'approximation rationnelle [51] dans les années 1960. Le fait 1 est généralement attribué à Paul Painlevé lui-même [38]. Le fait 2 apparaît (sous une forme un peu différente) dans la thèse de Lennart Carleson en 1950. Avant qu'une preuve complète en soit donnée, l'énoncé du fait 3 s'appelait «Conjecture de Denjoy». En effet, A. Denjoy [14] l'avait démontré dans le cas où la courbe Γ est une droite et il pensait que sa démonstration s'étendait au cas des courbes lipschitziennes générales. Il n'en était rien, comme l'ont remarqué L. Ahlfors et A. Beurling [2]. En 1978, A.P. Calderón [3] démontrait la continuité L^2 de l'opérateur de Cauchy sur les graphes lipschitziens (à condition que la constante de Lipschitz soit suffisamment petite). Un corollaire de ce résultat (connu des experts, mais pas de Calderón) était le fait 3. Le fait 4 avait été conjecturé par A.G. Vitushkin (sans d'ailleurs l'hypothèse $H^1(E) < +\infty$ et sous une forme légèrement différente, voir section 2.1.2). À la suite d'une série de travaux de M. Christ, P. Jones, P. Mattila, M. Melnikov, J. Verdera entre autres, Guy David [9] donnera une preuve complète du fait 4 en 1998. Un outil d'apparence élémentaire joue un rôle important dans cette histoire. Il s'agit de la courbure de Menger. Étant donnés 3 points z_1, z_2, z_3 distincts et non alignés du plan complexe \mathbb{C} ,

on définit leur courbure de Menger par

$$c(z_1, z_2, z_3) = \frac{1}{R(z_1, z_2, z_3)}$$

où $R(z_1, z_2, z_3)$ est le rayon du cercle circonscrit aux trois points. Si ces points ne sont pas distincts ou sont alignés, on pose $c(z_1, z_2, z_3) = 0$. Si μ est une mesure de Radon positive dans \mathbb{C} , on définit sa courbure de Menger par

$$c^2(\mu) = \int \int \int c(x, y, z)^2 d\mu(x) d\mu(y) d\mu(z).$$

Nous pouvons maintenant énoncer le résultat de Tolsa.

THÉORÈME 1.1 (X. Tolsa). — *Soit E un compact du plan complexe. Alors, E n'est pas effaçable si et seulement si E supporte une mesure de Radon positive (non nulle) μ telle que*

- (T1) *Il existe $C_0 > 0$ tel que $\mu(D(z, r)) \leq C_0 r$ pour tout $z \in \mathbb{C}$, tout $r > 0$ ($D(z, r)$ est le disque ouvert de centre z et de rayon r).*
 (T2) $c^2(\mu) < +\infty$.

Nous verrons dans la suite que (T1) est une condition sur la taille de l'ensemble, alors que (T2) est une condition de nature plus géométrique. Dans ce texte, une mesure satisfaisant (T1) sera dite à croissance linéaire (du volume).

Dans le paragraphe 2, nous introduirons les outils et les techniques d'analyse réelle et de théorie de la mesure géométrique dont nous aurons besoin pour expliquer la stratégie de démonstration de Tolsa. En particulier, nous verrons comment l'analyse harmonique s'est débarrassée de la condition de doublement du volume (travaux de Nazarov-Treil-Volberg et de Tolsa). Nous montrerons aussi comment déduire les faits 1, 2, 3 et 4 du théorème 1.1. Les démonstrations directes sont souvent plus simples, mais ceci permettra au lecteur de se convaincre que le résultat de Tolsa contient ce qui était connu avant et de se familiariser avec les diverses notions introduites au cours du texte. Dans le paragraphe 3, nous expliquerons comment la caractérisation de Tolsa permet de résoudre le problème de l'invariance par homéomorphismes bilipschitziens de la classe des ensembles effaçables. Nous discuterons dans le paragraphe 4 de problèmes ouverts autour de la capacité analytique et de la longueur de Favard, qui nous permettront de voir que la courbure de Menger n'est pas un objet aussi simple qu'il n'y paraît. Enfin, le paragraphe 5 sera consacré au cas (largement ouvert) de la dimension supérieure.

Je remercie L. Chevalier, G. David, J. Garnett, Y. Meyer, X. Tolsa, J. Verdera et A. Volberg pour leurs remarques sur des versions préliminaires de ce texte. L'auteur est partiellement supporté par le réseau européen TMR «Harmonic Analysis and Related Problems (HARP)».

Je dédie avec beaucoup d'amour ces quelques lignes à mon fils, Hervé-Maurice, né et décédé le même jour. Un sombre jour de mai 2004.

2. DE L'ANALYSE COMPLEXE VERS L'ANALYSE RÉELLE

Nous allons dans un premier temps voir comment les propriétés de rectifiabilité peuvent être contrôlées par des quantités géométriques (comme les nombres β de Peter Jones et la courbure de Menger). Cela nous permettra de relier ces propriétés de rectifiabilité à la continuité L^2 de l'opérateur de Cauchy (voir par exemple le théorème 2.13). Nous pourrions alors donner quelques idées de démonstration du fait 4 (voir le début de la section 2.2.2) et du théorème 1.1.

2.1. Nombres géométriques et rectifiabilité

2.1.1. Mesures et dimension de Hausdorff. — Soit $E \subset \mathbb{R}^n$ et soit $d \in [0, n]$ un réel. Nous définissons la d -mesure de Hausdorff de E par

$$H^d(E) = \lim_{\delta \rightarrow 0} \left(\inf \left\{ \sum_i (\text{diam} U_i)^d; E \subset \cup_i U_i, \text{diam} U_i < \delta \right\} \right).$$

Notons qu'ici les recouvrements considérés sont (au plus) dénombrables. Alors, H^d est une mesure de Borel régulière. Elle n'est pas de Radon (sauf si $d = n$) car elle n'est pas localement finie. Si $d = 0$, nous retrouvons la mesure de comptage. Pour une courbe de Jordan rectifiable Γ de \mathbb{R}^n , $H^1(\Gamma)$ est sa longueur. Enfin, H^n est égale (à une constante multiplicative près) à la mesure de Lebesgue \mathcal{L}^n de \mathbb{R}^n . Par un argument élémentaire, nous pouvons démontrer que $\inf\{s; H^s(E) = 0\} = \sup\{t; H^t(E) = +\infty\}$. Cette valeur commune est appelée la dimension de Hausdorff de E . Nous la noterons $\text{Hdim}(E)$. Par exemple, la dimension de Hausdorff de l'ensemble triadique de Cantor est $\log(2)/\log(3)$, celle d'une courbe de longueur finie est 1. Enfin, $\text{Hdim}(\mathbb{R}^n) = n$. Un point-clé pour nous est le lemme de Frostman (voir [27] pour une démonstration, ainsi que pour plus de détails sur la théorie de la mesure géométrique) que voici :

THÉORÈME 2.1. — *Soit E un borélien de \mathbb{R}^n et soit $d \geq 0$. Alors, $H^d(E) > 0$ si et seulement si E supporte une mesure de Radon positive μ (à support compact) telle que*

- (i) $\mu(\mathbb{R}^n) = 1$ (μ est une mesure de probabilité).
- (ii) *Il existe une constante $C_0 > 0$ telle que $\mu(B(x, R)) \leq C_0 R^d$ pour tout $x \in \mathbb{R}^n$, tout $R > 0$. Ici, et dans la suite, $B(x, R)$ désigne la boule euclidienne de centre x et de rayon R .*

Nous pouvons maintenant voir que le théorème de Tolsa implique le fait 1. En effet, soit E un compact du plan complexe tel que $H^1(E) = 0$. Alors, d'après le lemme de Frostman, E ne peut supporter une mesure μ à croissance linéaire (condition (T1) du théorème 1.1). Donc, E est effaçable. Signalons que le fait 1 peut se démontrer directement (et facilement) grâce à la formule de Cauchy (voir le théorème 64 dans [40]).

2.1.2. Rectifiabilité. — Soit $d \in \mathbb{N}^*$. Un ensemble $E \subset \mathbb{R}^n$ est dit d -rectifiable (au sens de la théorie géométrique de la mesure) s'il existe une famille dénombrable d'applications lipschitziennes $f_j : \mathbb{R}^d \rightarrow \mathbb{R}^n$ telles que $H^d(E \setminus \cup_j f_j(\mathbb{R}^d)) = 0$. Une courbe rectifiable Γ de \mathbb{R}^n est 1-rectifiable. D'un autre côté, un ensemble E est dit purement non d -rectifiable si et seulement si $H^d(E \cap F) = 0$ pour tout ensemble d -rectifiable F de \mathbb{R}^n . Dans le cas particulier où $d = 1$, cette condition est équivalente à $H^1(E \cap \Gamma) = 0$ pour toute courbe rectifiable Γ de \mathbb{R}^n . Donc, le fait 4 dit que si $H^1(E) < +\infty$, alors E est effaçable si et seulement si E est purement non 1-rectifiable. Un exemple d'un tel ensemble est l'ensemble de Cantor 4-coins que nous allons maintenant construire. Soit $E_0 = [0, 1]^2$ le carré unité dans \mathbb{C} . Découpons E_0 en 16 carrés égaux de longueur de côté $\frac{1}{4}$. L'ensemble E_1 est l'union des 4 carrés situés dans les coins de E_0 . Puis, on découpe chacun des 4 carrés en 16 carrés identiques et l'ensemble E_2 est l'union des 16 carrés qui sont situés dans les coins des 4 carrés de E_1 . En itérant cette construction, nous pouvons exhiber une suite de sous-ensembles $(E_j)_{j \in \mathbb{N}}$ de \mathbb{C} , chacun des E_j étant formé de 4^j carrés de longueur de côté 4^{-j} qui sont situés dans les coins des carrés de E_{j-1} . Soit $\mathcal{C} = \bigcap_{j \in \mathbb{N}} E_j$. Alors, $\text{Hdim}(\mathcal{C}) = 1$ et $H^1(\mathcal{C}) = \sqrt{2}$. Le fait que \mathcal{C} est purement non rectifiable découle de la remarque suivante. Si Γ est une courbe lipschitzienne du plan complexe, sa projection dans toute direction sauf peut-être une (penser au cas où Γ est une droite) est de mesure de Lebesgue (dans \mathbb{R}) non nulle. Il en est de même de tout sous-ensemble de Γ de mesure strictement positive. Donc, s'il existait une courbe lipschitzienne Γ telle que $H^1(\mathcal{C} \cap \Gamma) > 0$ alors \mathcal{C} ne pourrait pas avoir des projections de mesure de Lebesgue nulle dans deux directions distinctes. Ce qui est pourtant le cas ! Donc, \mathcal{C} est purement non rectifiable. L'ensemble de Cantor 4-coins est un exemple d'ensemble de 1-mesure de Hausdorff non nulle mais qui est effaçable (voir [16]).

Tout ensemble $E \subset \mathbb{R}^n$ avec $H^d(E) < +\infty$ se décompose en une bonne et une mauvaise partie :

$$E = E_{\text{rect}} \cup E_{\text{nonrect}}$$

où E_{rect} (respectivement E_{nonrect}) est d -rectifiable (respectivement purement non d -rectifiable). Il est important de noter qu'il existe diverses caractérisations des ensembles rectifiables/purement non rectifiables (en termes de densité, d'existence de tangentes, de taille des projections, voir [27]) dans le cas des ensembles de mesure de Hausdorff finie. Cependant, les propriétés de rectifiabilité pour des ensembles qui ne sont pas de mesure de Hausdorff σ -finie sont très mal connues. Donnons maintenant un critère de non rectifiabilité pour les sous-ensembles du plan complexe. Commençons par une définition. Soit $E \subset \mathbb{C}$. Pour tout $\theta \in [0, \pi]$, notons $\Pi_\theta(E)$ la projection orthogonale de E sur la direction θ et $|\Pi_\theta(E)|$ la mesure de Lebesgue (dans \mathbb{R}) de cette projection. Nous définissons la longueur de Favard de E par

$$\text{Fav}(E) = \frac{2}{\pi} \int_0^\pi |\Pi_\theta(E)| d\theta.$$

Le théorème des projections de Besicovitch s'énonce ainsi :

THÉORÈME 2.2. — Soit $E \subset \mathbb{C}$ tel que $H^1(E) < +\infty$. Alors, E est purement non rectifiable si et seulement si $\text{Fav}(E) = 0$.

Donc, d’après le fait 4, si $H^1(E) < +\infty$, alors E est effaçable si et seulement si $\text{Fav}(E) = 0$. Vituskhin avait conjecturé ce résultat sans l’hypothèse $H^1(E) < +\infty$. Nous discuterons de ce cas dans le paragraphe 4.

2.1.3. *Le problème géométrique du voyageur de commerce.* — Nous allons, dans ce paragraphe, donner des conditions quantitatives de rectifiabilité. Le point de départ est un article de P. Jones [18] qui a eu l’idée d’introduire les nombres β qui mesurent en tout point et à toutes les échelles la qualité de l’approximation d’un ensemble $E \subset \mathbb{R}^n$ par des droites. Pour tout $x \in \mathbb{R}^n$ et tout $t > 0$, nous définissons

$$\beta_\infty^E(x, t) = \inf_L \sup_{y \in E \cap B(x, t)} \frac{d(y, L)}{t}$$

où la borne inférieure est prise sur toutes les droites L de \mathbb{R}^n .

Posons $\beta(E) = \int_{\mathbb{R}^n} \int_0^{+\infty} \beta_\infty^E(x, t)^2 d\mathcal{L}^n(x) \frac{dt}{t^n}$, où \mathcal{L}^n est la mesure de Lebesgue sur \mathbb{R}^n . Si E est, par exemple, le graphe d’une fonction lipschitzienne, alors en presque tout $x \in E$, il existe une tangente au graphe et donc $\beta_\infty^E(x, t)$ tend vers 0 avec t . En fait, nous allons voir que les propriétés de rectifiabilité sont liées à des estimations L^2 des nombres β .

THÉORÈME 2.3 ([18] [37]). — Soit E un sous-ensemble compact de \mathbb{R}^n . L’ensemble E est contenu dans une courbe de longueur finie Γ si et seulement si $\beta(E) < +\infty$. De plus, si c’est le cas,

$$C^{-1}(\beta(E) + \text{diam}E) \leq \inf_{\Gamma \supset E} l(\Gamma) \leq C(\beta(E) + \text{diam}E)$$

où $C \geq 1$ est une constante ne dépendant que de n .

Ce résultat apparaît comme une version géométrique du problème classique du voyageur de commerce, dans la mesure où l’ensemble E des villes que le voyageur doit visiter peut être infini. Le théorème donne alors une condition nécessaire et suffisante pour que le voyageur puisse remplir sa tâche en un «temps fini». Nous allons maintenant donner une version du théorème 2.3 qui a un analogue pour des ensembles de dimension strictement plus grande que 1. Commençons par une définition. Nous dirons qu’un ensemble (non réduit à un point) $E \subset \mathbb{R}^n$ est Ahlfors-régulier (de dimension 1) s’il existe une constante $C_0 > 0$ telle que

$$C_0^{-1}R \leq H^1(E \cap B(x, R)) \leq C_0R$$

pour tout $x \in E$, tout $R \in]0, \text{diam}E[$. Notons qu’un tel ensemble est alors de dimension de Hausdorff égale à 1. Cette condition d’Ahlfors-régularité impose peu de restriction sur la géométrie de l’ensemble. Par exemple, sont Ahlfors-réguliers des ensembles rectifiables comme les graphes lipschitziens et des ensembles purement non rectifiables comme l’ensemble de Cantor 4 coins. Une courbe Ahlfors-régulière Γ est

une courbe localement rectifiable telle qu'il existe une constante $C_0 > 0$ telle que $H^1(\Gamma \cap B(x, r)) \leq C_0 r$ pour tout $x \in \Gamma$, tout $r > 0$. Notons que, par connexité, nous avons aussi $H^1(\Gamma \cap B(x, r)) \geq C_0^{-1} r$ si $x \in \Gamma$ et r est assez petit. Suivant David et Semmes [11] [12], nous dirons qu'un ensemble Ahlfors-régulier E est uniformément rectifiable s'il existe une courbe Ahlfors-régulière Γ telle que $E \subset \Gamma$.

THÉORÈME 2.4. — *Supposons que $E \subset \mathbb{C}$ soit compact et Ahlfors-régulier (de dimension 1).*

- (i) *L'ensemble E est uniformément rectifiable si et seulement si il existe $C > 0$ telle que, pour tout $x \in E$, tout $R \in]0, \text{diam} E[$,*

$$\int_{y \in E \cap B(x, R)} \int_0^R \beta_\infty^E(y, t)^2 dH^1(y) \frac{dt}{t} \leq CR.$$

- (ii) *L'ensemble E est rectifiable si et seulement si $\int_0^{\text{diam} E} \beta_\infty^E(x, t)^2 \frac{dt}{t} < \infty$ pour H^1 -presque tout $x \in E$.*

Le (i) s'obtient en adaptant les arguments de la démonstration du théorème 2.3. Il existe bien d'autres caractérisations des ensembles uniformément rectifiables (voir [11] et [12]). Le (ii) découle en partie de (i) (voir [39]).

2.1.4. La courbure de Menger. — Nous avons rencontré dans l'introduction un autre «nombre géométrique», à savoir la courbure de Menger. Son côté magique repose dans la formule élémentaire suivante.

$$(1) \quad \left(\frac{4S(z_1, z_2, z_3)}{|z_1 - z_2||z_2 - z_3||z_3 - z_1|} \right)^2 = c(z_1, z_2, z_3)^2 = \sum_{\sigma} \frac{1}{(z_{\sigma(1)} - z_{\sigma(2)})(z_{\sigma(1)} - z_{\sigma(3)})}$$

où $S(z_1, z_2, z_3)$ est l'aire du triangle de sommets z_1, z_2 et z_3 , et la somme se fait sur les permutations σ sur $\{1, 2, 3\}$. Le membre de gauche mesure la «platitude» du triplet z_1, z_2, z_3 , alors que nous verrons un peu plus tard que le membre de droite est relié à l'intégrale de Cauchy. Ainsi, la courbure de Menger va nous permettre de faire le lien entre propriétés géométriques (rectifiabilité) et propriétés analytiques (continuité de l'opérateur de Cauchy). Commençons par le premier aspect.

THÉORÈME 2.5 ([21]). — *Soit E un sous-ensemble compact de \mathbb{C} tel que $H^1(E) < +\infty$. Si $c^2(E) < +\infty$, alors E est 1-rectifiable.*

Ici, $c^2(E)$ désigne la courbure de Menger de la restriction de H^1 à E . Dans le cas Ahlfors-régulier, la courbure de Menger permet de caractériser la rectifiabilité uniforme.

THÉORÈME 2.6. — *Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier de dimension 1. Alors, E est uniformément rectifiable si et seulement si il existe une constante $C_0 > 0$ telle que pour tout disque D de \mathbb{C} ,*

$$(2) \quad \int \int \int_{(E \cap D)^3} c(x, y, z)^2 dH^1(x) dH^1(y) dH^1(z) \leq C \text{diam} D.$$

La condition (2) s'appelle la condition de courbure locale. Ce résultat est énoncé et démontré dans [30]. Nous pouvons aussi le voir comme une conséquence du théorème 2.4 (i) et des deux résultats suivants qui font le lien entre courbure de Menger et nombres β (voir [40]).

Attention, les «exposants» dans les nombres β précisent à quel ensemble les β réfèrent.

THÉORÈME 2.7. — *Soit Γ une courbe rectifiable de \mathbb{C} et soit μ une mesure de Radon positive telle que $\text{supp}\mu$ soit un sous-ensemble compact de Γ . Supposons que μ est à croissance linéaire du volume (c'est-à-dire vérifie (T1)) avec une constante $1/2$:*

$$\mu(D) \leq \text{diam}D \text{ pour tout disque } D \subset \mathbb{C}.$$

Alors, $c^2(\mu) \leq C \int_{\mathbb{C}} \int_0^{+\infty} \beta_{\infty}^{\Gamma}(x, t)^2 d\mu(x) \frac{dt}{t}$. Si de plus, Γ est Ahlfors-régulière, alors $c^2(\mu) \leq C\mu(\mathbb{C})$ où $C > 0$ ne dépend que de la constante de régularité de Γ .

THÉORÈME 2.8. — *Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier de dimension 1. Alors, pour tout disque D de \mathbb{C} ,*

$$\int_{E \cap D} \int_0^{\text{diam}D} \beta_{\infty}^E(x, t)^2 dH^1(x) \frac{dt}{t} \leq C \int \int \int_{(E \cap C \cdot D)^3} c(x, y, z)^2 dH^1(x) dH^1(y) dH^1(z)$$

où $C > 0$ ne dépend que de la constante de régularité de E .

Dans la seconde intégrale, $C \cdot D$ désigne le disque de même centre que D mais de diamètre $C \cdot \text{diam}D$. Il est important de noter que les nombres β sont plus faciles à manipuler en pratique que la courbure de Menger, d'où l'intérêt d'avoir des estimations qui permettent de passer de l'un à l'autre. Nous allons maintenant voir comment le théorème de Tolsa implique les faits 3 et 4. Tout d'abord, soit $E \subset \Gamma$, où Γ est une courbe rectifiable, tel que $H^1(E) > 0$. Alors, d'après le lemme de Frostman, il existe une mesure de probabilité μ supportée sur E à croissance linéaire. Donc, d'après les théorèmes 2.3 et 2.7, $c^2(\mu) < +\infty$. D'où, E n'est pas effaçable. Passons maintenant au fait 4 et supposons que $E \subset \mathbb{C}$ ne soit pas effaçable. Alors, il existe une mesure μ supportée sur E qui est à croissance linéaire et de courbure finie. Donc, d'après le théorème 2.5 (et avec un peu de travail), le support de μ est rectifiable. D'où, il existe une courbe lipschitzienne Γ telle que $H^1(E \cap \Gamma) > 0$.

Il est amusant de noter que la courbure de Menger d'une mesure supportée sur un ensemble de dimension plus grande que 1 peut être finie, sans que la géométrie de cet ensemble n'entre en compte. En effet, nous avons le

THÉORÈME 2.9 ([28]). — *Soit $h : [0, +\infty] \rightarrow [0, \infty]$ une fonction strictement croissante telle que $\int_0^{+\infty} r^{-3} h^2(r) dr < \infty$ et soit μ une mesure de Borel finie dans \mathbb{C} telle que, pour tout disque D de rayon $r > 0$, $\mu(D) \leq Ch(r)$. Alors,*

$$c^2(\mu) \leq C\mu(\mathbb{C}) \int_0^{+\infty} r^{-3}h^2(r)dr.$$

En particulier, si E est de dimension de Hausdorff strictement plus grande que 1, alors il existe $d > 1$ tel que $H^d(E) > 0$. Donc, d'après le lemme de Frostman, il existe une mesure de probabilité μ supportée par E telle que $\mu(D) \leq CR^d$ pour tout disque D de \mathbb{C} de rayon $R > 0$. Or, d'après le théorème 2.9 (prendre $h(t) = t^d$), $c^2(\mu) < +\infty$. Donc, E n'est pas effaçable. Le théorème de Tolsa implique donc le fait 2. Pour une preuve directe, voir [40], théorème 64.

2.2. Analyse harmonique sans doublement du volume

2.2.1. La théorie classique. — Un espace de type homogène au sens de Coifman-Weiss est un triplet (X, δ, μ) où X est un ensemble (pour nous, un sous-ensemble de \mathbb{C}), δ est une distance (ou une quasi-distance) sur X et μ est une mesure (de Radon positive), supportée sur X , qui satisfait la condition de doublement du volume, c'est-à-dire qu'il existe une constante $C_{dv} \geq 1$ telle que

$$(DV) \quad \mu(B(x, 2R)) \leq C_{dv}\mu(B(x, R))$$

pour toute boule $B(x, R)$ relativement à δ de centre $x \in X$ et de rayon $R > 0$. Beaucoup de faits classiques d'analyse dans les espaces euclidiens restent vrais dans ces espaces homogènes. Le point-clé est que la condition (DV) permet d'utiliser des théorèmes de recouvrement de type Vitali (voir par exemple le premier chapitre de [44]). Le fait que ces espaces soient un cadre assez général pour faire de l'analyse harmonique apparaît après les travaux de Coifman et Weiss [6] au début des années 1970. Il est clair que \mathbb{C} muni de sa structure euclidienne et de sa mesure de Lebesgue est un espace de type homogène. Il en est de même si X est un sous-ensemble de \mathbb{C} qui est Ahlfors-régulier (de dimension quelconque d). Dans ce cas, on munit X de la distance euclidienne induite et de la restriction de la mesure de Hausdorff H^d sur X (et c'est ce que nous ferons systématiquement dans la suite). En particulier, peuvent être équipés d'une structure d'espace homogène toute droite (vue comme sous-ensemble de \mathbb{C}), tout graphe lipschitzien de \mathbb{C} (c'est-à-dire le graphe, à une rotation près d'une fonction lipschitzienne $A : \mathbb{R} \rightarrow \mathbb{R}$), ou l'ensemble de Cantor 4-coins. Décrivons maintenant quelques éléments de la théorie des intégrales singulières dans les espaces homogènes (voir par exemple [4] pour plus de détails), en nous restreignant au cas des espaces homogènes contenus dans \mathbb{C} .

Un noyau standard antisymétrique pour la dimension 1 est une fonction $K(x, y) : \mathbb{C} \times \mathbb{C} \setminus \{x = y\} \rightarrow \mathbb{R}$ telle qu'il existe des constantes $s \in]0, 1]$ et $C > 0$ vérifiant, pour tout $x \neq y$,

- (i) $K(x, y) = -K(y, x)$;
- (ii) $|K(x, y)| \leq C \frac{1}{|x - y|}$;

$$(iii) \quad |K(x, y) - K(x', y)| \leq C \frac{|x - x'|^s}{|x - y|^{1+s}} \text{ si } |x - x'| < \frac{1}{2}|x - y|.$$

Ainsi, le comportement de $K(x, y)$ est comparable à celui de $\frac{1}{x - y}$. Considérons maintenant une mesure de Radon positive μ sur \mathbb{C} . Dans ce cas, si μ vérifie (DV), l'espace homogène considéré est le support de la mesure muni de la distance euclidienne induite et de la mesure μ . Nous définissons l'opérateur tronqué T_μ^ε associé au noyau standard K par

$$T_\mu^\varepsilon(f)(x) = \int_{|x-y|>\varepsilon} K(x, y)f(y)d\mu(y)$$

pour $f \in L^1(\mu)$ et $x \in \mathbb{C}$.

Nous dirons que l'opérateur T_μ associé au noyau K est continu sur $L^p(\mu)$ s'il existe $C > 0$ telle que pour tout $\varepsilon > 0$, toute fonction $f \in L^p(\mu)$,

$$\int |T_\mu^\varepsilon(f)(x)|^p d\mu(x) \leq C \int |f(x)|^p d\mu(x).$$

En d'autres termes, T_μ^ε est borné sur $L^p(\mu)$ indépendamment de ε . Cette définition de la continuité L^p est équivalente à la plupart des définitions disponibles dans la littérature et permet d'éviter d'avoir à définir proprement $T_\mu(f)$. Un premier fait important est que si μ satisfait la condition de doublement du volume (et donc son support est un espace homogène) et est à croissance linéaire, alors le théorème de Calderón-Zygmund nous dit que, si un opérateur T_μ est continu sur $L^2(\mu)$, alors il est continu sur $L^p(\mu)$ pour $1 < p < \infty$ et il est de type faible $(1, 1)$, c'est-à-dire qu'il existe une constante $C > 0$ indépendante de ε telle que, pour tout $M > 0$,

$$\mu(\{x \in \mathbb{C}; |T_\mu^\varepsilon(f)(x)| > M\}) \leq \frac{C}{M} \int |f| d\mu$$

pour toute fonction f dans $L^1(\mu)$. Ainsi, compte tenu de ce résultat, la continuité L^2 qui peut être obtenue par des méthodes d'analyse de Fourier joue un rôle crucial dans cette théorie. Considérons par exemple le cas du noyau de Cauchy $K(x, y) = \frac{1}{x - y}$.

Dans ce cas, nous noterons \mathcal{C}_μ l'opérateur associé. Si la mesure μ est la restriction de H^1 à \mathbb{R} , l'opérateur s'appelle l'opérateur de Hilbert et sa continuité L^2 s'obtient facilement en passant en transformée de Fourier et en appliquant le théorème de Plancherel. Dans le cas où μ est la restriction de H^1 à un graphe lipschitzien, le problème est plus ardu. Cette continuité a été conjecturée par Calderón et Zygmund dans les années 1950, puis démontrée par Calderón [3] dans le cas où la constante de Lipschitz du graphe est assez petite. Le cas général a été résolu par Coifman, Mc Intosh et Meyer [7]. Il est amusant de noter que de nombreuses preuves très différentes ont été proposées depuis. Un problème important dans les années 1980 est alors d'établir des critères de continuité L^2 pour les opérateurs d'intégrales singulières, en s'inspirant du cas de l'opérateur de Cauchy. Le premier de ces critères a été obtenu par G. David et J.-L. Journé et porte le nom de théorème $T(1)$.

THÉORÈME 2.10. — Soit μ une mesure doublante et à croissance linéaire, et soit K un noyau antisymétrique dans \mathbb{C} . Alors, T_μ est continu sur $L^2(\mu)$ si et seulement si, pour tout disque D de \mathbb{C} , tout $\varepsilon > 0$,

$$(3) \quad \int_D |T_\mu^\varepsilon(\chi_D)(x)|^2 d\mu(x) \leq C\mu(D)$$

où $C > 0$ est une constante indépendante de ε et de D .

La condition (3) est équivalente au fait que $T_\mu(1)$ (quand on donne un sens convenable à l'image de la fonction 1 par l'opérateur T_μ) est dans l'espace $BMO(\mu)$ de John et Nirenberg, d'où le nom du théorème. Il existe un critère plus flexible (appelé théorème $T(b)$) pour lequel nous testons l'opérateur sur une fonction raisonnable b que nous pouvons choisir. Au lieu de le donner, nous allons en énoncer une version locale due à M. Christ [5], en ne considérant que le cas Ahlfors-régulier. Commençons par équiper un ensemble Ahlfors-régulier (de dimension 1) d'une famille de partitions qui joue le rôle des intervalles dyadiques de \mathbb{R} .

THÉORÈME 2.11. — Soit $E \subset \mathbb{R}^2$ un ensemble Ahlfors-régulier de dimension 1. Il existe une famille de partitions Δ_j , $j \in \mathbb{Z}$ tel que $2^j \leq \text{diam}E$, de E en « intervalles dyadiques » $Q \subset E$ avec :

- (i) Si $j \geq k$, $Q \in \Delta_j$, $Q' \in \Delta_k$, alors soit $Q \cap Q' = \emptyset$, soit $Q' \subset Q$.
- (ii) Si $Q \in \Delta_j$, alors

$$\begin{aligned} C^{-1}2^j &\leq \text{diam}Q \leq C2^j \\ C^{-1}2^j &\leq H^1(Q) \leq C2^j, \end{aligned}$$

(où $C > 0$ ne dépend que de E). De plus, les « intervalles dyadiques » ont une petite frontière, au sens où

- (iii) Pour tout $Q \in \cup_j \Delta_j$, tout $\tau \in]0, 1[$,

$$H^1(\{z \in Q; d(z, E \setminus Q) \leq \tau \text{diam}Q\}) \leq C\tau^{1/C}H^1(Q).$$

Notons que le nombre de fils d'un intervalle dyadique $Q \in \Delta_j$ (c'est-à-dire d'intervalles dyadiques de la génération suivante Δ_{j-1} contenus dans Q) est borné indépendamment de j . Le lecteur pourra vérifier aisément que, si $E = \mathbb{R}$, les intervalles dyadiques usuels, c'est-à-dire ceux de la forme $[k2^j, (k+1)2^j]$, satisfont les propriétés précédentes. Voir [8] pour une démonstration de 2.11, ainsi que [5] pour une construction dans le cas d'une mesure qui n'est que doublante. Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier (de dimension 1) et soit $\Delta = \cup_j \Delta_j$ une famille d'intervalles dyadiques comme donnés par le théorème précédent. Un système de fonctions $\{b_Q, Q \in \Delta\}$ est dit *pseudo-accrétif* s'il existe des constantes $C > 0$, $\varepsilon > 0$ telles que, pour tout $Q \in \Delta$, $\|b_Q\|_\infty \leq C$ et $\left| \int_Q b_Q dH^1 \right| \geq \varepsilon H^1(Q)$. Nous pouvons maintenant énoncer le théorème $T(b)$ local.

THÉORÈME 2.12. — *Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier de dimension 1 et soit T_μ un opérateur d'intégrale singulière associé à un noyau standard anti-symétrique et défini par rapport à la mesure $\mu = H_E^1$ (restriction de H^1 sur E). Supposons qu'il existe un système pseudo-accrétif (b_Q) tel que $\|T_\mu^\varepsilon b_Q\| \leq C$ pour tout $Q \in \Delta$ et tout $\varepsilon > 0$. Alors, T_μ est borné sur $L^2(\mu)$.*

Le côté local vient du fait qu'ici, nous devons construire une fonction b_Q sur chaque intervalle dyadique Q alors que, dans le théorème $T(b)$ classique, la fonction test b est choisie une fois pour toutes et définie sur E tout entier. Celle-ci doit vérifier une condition de moyenne sur les intervalles dyadiques qui est identique à celle des b_Q , mais la condition L^∞ est remplacée par une condition du type BMO (du style de celles du théorème $T(1)$ en remplaçant la fonction caractéristique par son produit par b), qui est plus faible. Signalons que l'on passe de la version classique à la version locale du théorème $T(b)$ en posant $b_Q = b \chi_Q$ et que, réciproquement, en «recollant» les b_Q , on peut construire une fonction b qui satisfait les hypothèses du théorème $T(b)$. Revenons à l'opérateur de Cauchy. Nous avons vu qu'il définit un opérateur borné (au sens L^2) sur les graphes lipschitziens. Mais, dans quelle mesure les propriétés de rectifiabilité du support de la mesure jouent-elles un rôle ? À la suite de travaux de G. David et S. Semmes [11], [12], P. Mattila, M. Melnikov et J. Verdera [30] ont démontré le résultat étonnant suivant.

THÉORÈME 2.13. — *Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier de dimension 1. Notons μ la restriction de H^1 sur E . Alors, l'opérateur de Cauchy \mathcal{C}_μ est bornée sur $L^2(\mu)$ si et seulement si E est uniformément rectifiable.*

Donnons une idée de la démonstration, car cela nous permettra de voir comment les ingrédients précédents peuvent être associés. Formellement, en oubliant les ε (voir [40] pour les détails techniques), la continuité L^2 de l'opérateur est équivalente, d'après le théorème $T(1)$, à

$$\int_D |\mathcal{C}_E(\chi_D)|^2 d\mu \leq C \text{diam} D$$

pour tout disque D de \mathbb{C} . Ici, nous avons utilisé le fait qu'un ensemble Ahlfors-régulier muni de sa mesure de Hausdorff est un espace de type homogène afin de pouvoir utiliser le théorème $T(1)$. Or, d'après la formule magique (1) sur la courbure,

$$\begin{aligned} \int_D |\mathcal{C}_\mu(\chi_D)(x)|^2 d\mu(x) &= \int_D \mathcal{C}_\mu(\chi_D)(x) \overline{\mathcal{C}_\mu(\chi_D)(x)} d\mu(x) \\ &= \int_D \int_D \int_D \frac{1}{(y-x)\overline{(z-x)}} d\mu(x) d\mu(y) d\mu(z) \\ &= 6 \int_D \int_D \int_D c(x, y, z)^2 d\mu(x) d\mu(y) d\mu(z). \end{aligned}$$

Il en résulte que la continuité L^2 de \mathcal{C}_μ est équivalente à la condition de courbure locale, qui est elle-même équivalente à la rectifiabilité uniforme (voir le théorème 2.6) !

Signalons que l'idée d'utiliser la courbure de Menger pour étudier l'opérateur de Cauchy apparaît pour la première fois dans [33].

2.2.2. Les théorèmes $T(1)$ et $T(b)$ sans doublement. — Essayons d'expliquer l'intérêt pour nous d'avoir des analogues dans le cas non homogène des théorèmes du style $T(1)$ en expliquant le plan de preuve d'un des sens du fait 4. Pour cela, considérons $E \subset \mathbb{C}$ tel que $H^1(E) < +\infty$ et $\gamma(E) > 0$. Commençons par le cas où E est Ahlfors-régulier (ce cas a été traité dans [30]).

Étape 1 : Puisque $\gamma(E) > 0$, il existe une fonction non constante, holomorphe, bornée $f : \mathbb{C} \setminus E \rightarrow \mathbb{C}$ telle que $\|f\|_\infty \leq 1$ et (quitte à multiplier f par une constante de module 1) $f'(\infty) = \gamma(E)$. Un argument standard [27] montre qu'il existe une fonction borélienne $h : E \rightarrow \mathbb{C}$ bornée telle que $f(z) = \int_E \frac{h(\xi)}{z - \xi} dH^1(\xi)$ pour tout $z \notin E$.

Étape 2 : L'idée (due à M. Christ dans [5]) est maintenant de construire à partir de h , en utilisant un argument de temps d'arrêt, un ensemble Ahlfors-régulier F et une fonction $b : F \rightarrow \mathbb{C}$ tels que $H^1(E \cap F) > 0$ et tels que les hypothèses du théorème $T(b)$ soient vérifiées pour l'opérateur de Cauchy sur F muni de H^1_F (restriction de H^1 sur F). Comme cet espace est de type homogène, on en déduit que l'opérateur de Cauchy est borné sur $L^2(F)$.

Étape 3 : D'après le théorème 2.13, F est uniformément rectifiable et donc il existe une courbe rectifiable Γ telle que $H^1(E \cap \Gamma) > 0$.

Le cas général a été résolu par G. David [9] (voir aussi [36]), après un premier pas important avec P. Mattila dans [10]. Expliquons rapidement comment modifier la preuve précédente. L'étape 1 est inchangée. Un des problèmes pour l'étape 2 était le manque d'un théorème de type $T(b)$ dans le cadre des espaces qui ne sont pas de type homogène. Guy David comblera cette lacune (voir plus bas) mais c'est en partie ce problème qui a motivé le développement de l'analyse harmonique dans ces espaces. Concernant l'étape 3, le même genre d'argument que pour démontrer le théorème 2.13 permet de passer d'estimations sur l'opérateur de Cauchy à des estimations sur la courbure de Menger puis nous pouvons appliquer le théorème 2.5 (pour obtenir une partie rectifiable dans E).

Une des premières versions du théorème $T(1)$ dans le cas non doublant concerne l'opérateur de Cauchy ; elle est due indépendamment à Nazarov-Treil-Volberg [34] et Tolsa [45]. Ce résultat s'énonce ainsi.

THÉORÈME 2.14. — *Soit μ une mesure de Borel positive qui est à croissance linéaire. Alors, l'opérateur de Cauchy \mathcal{C}_μ est borné sur $L^2(\mu)$ si et seulement si il existe une constante $C > 0$ telle que*

$$\int_D |\mathcal{C}_\mu^\varepsilon(\chi_D)|^2 d\mu \leq C\mu(D),$$

pour tout $\varepsilon > 0$ et tout disque D de \mathbb{C} .

Notons que la condition de croissance linéaire est nécessaire. Le lecteur attentif aura noté que l'énoncé du théorème 2.14 est le même que celui du théorème $T(1)$ dans le cas doublant. Ce qui est tout à fait incroyable ! La démonstration de Tolsa repose sur la courbure de Menger et ne permet pas de traiter le cas de noyaux plus généraux. La stratégie (qui s'inspire de celle utilisée dans le cas doublant) de Nazarov-Treil-Volberg leur permettra de considérer tous les noyaux standards (voir [35] par exemple). Expliquons le schéma (maintenant classique) de démonstration d'un théorème de type $T(1)$ ou $T(b)$ dans le cas d'un ensemble Ahlfors-régulier $E \subset \mathbb{C}$ de dimension 1. Rappelons que, dans ce cas, la mesure μ considérée est la restriction de H^1 à E . Tout d'abord, équipons E d'une famille d'intervalles dyadiques comme dans le théorème 2.11. Une idée naïve pour construire une telle famille est de tout simplement considérer la famille des intersections de tous les carrés dyadiques usuels de \mathbb{C} avec E . Ceci ne marche pas aussi simplement, mais cela donne une bonne idée de ce à quoi doit ressembler un «intervalle dyadique» de E . Considérons la base de Haar $(b_Q)_{Q \in \Delta}$ associée. Le but est alors d'estimer les coefficients de T_μ^ε dans cette base, c'est-à-dire les $\langle T_\mu^\varepsilon(b_Q), b_R \rangle$ pour Q, R dans Δ . Pour cela, nous devons distinguer suivant les positions relatives de R et Q . Ici, vont jouer un rôle important le fait que $\mu(Q)$ est comparable à $\text{diam} Q$ et la propriété de «petite frontière» des «intervalles» de Δ . Pour voir l'importance de la dernière propriété, considérons le cas où Q et R sont très proches. Alors, dans la mesure où le noyau $K(x, y)$ se comporte comme $\frac{1}{x-y}$, ses singularités sont proches de la diagonale $x = y$. La propriété de «petite frontière», qui dit que la mesure de l'ensemble des points $x \in Q$ et des points $y \in R$ qui sont très proches est petite, permet de compenser l'effet des singularités du noyau. Sous de bonnes hypothèses, la décroissance des coefficients de T_μ^ε dans la base de Haar au voisinage de la diagonale permet de démontrer la continuité L^2 de T_μ , en appliquant un argument classique du style lemme de Schur (voir [8]). Si notre ensemble E est seulement doublant, ces arguments s'adaptent (voir [5]). Dans le cas où l'espace E n'est plus homogène mais la restriction de H^1 à E est à croissance linéaire, il est possible de construire une famille d'intervalles dyadiques (voir [10]), puis d'adapter les idées précédentes pour obtenir un théorème du style $T(b)$ (voir [9]). Mais, tout ceci est long et très technique. L'idée de Nazarov-Treil-Volberg est de ne pas essayer de construire une bonne famille d'intervalles dyadiques, mais de considérer pour tout vecteur \vec{u} , la famille de carrés dyadiques obtenus en translatant de \vec{u} la famille usuelle de carrés dyadiques de \mathbb{C} . En intersectant ces carrés avec E , ils obtiennent une collection indexée par \vec{u} , et notée $\Delta_{\vec{u}}$, d'intervalles dyadiques. Il est clair que pour des choix aléatoires de \vec{u} , la famille $\Delta_{\vec{u}}$ ne satisfait pas les conclusions du théorème 2.11 et que, si nous estimons la matrice de T dans la base de Haar associée à $\Delta_{\vec{u}}$, nous n'obtenons pas la bonne décroissance près de la diagonale. Mais, ce que montrent Nazarov-Treil-Volberg, c'est qu'en moyenne, tout se passe comme dans le cas homogène, et que l'on peut ainsi conclure. Leurs démonstrations utilisent toute une panoplie d'astuces techniques étonnantes, comme par exemple les fonctions de Bellman. Ils obtiennent ainsi des théorèmes de type $T(1)$ ou $T(b)$ et des versions locales de $T(b)$

au sens de M. Christ pour des mesures non doublantes. Nous renvoyons à [52] et à sa liste de références pour plus de détails sur les travaux de Nazarov-Treil-Volberg. Pour un survol sur le développement de l'analyse harmonique non homogène, voir [50].

2.3. La stratégie de Tolsa

La démonstration de Tolsa se fait en deux temps et s'inspire de celle donnée dans [24] dans le cas des ensembles de Cantor de type 4-coins. Elle repose sur une comparaison avec une capacité plus facile à étudier. Celle-ci, notée γ_+ , est définie pour tout compact $E \subset \mathbb{C}$ par $\gamma_+(E) = \sup \mu(\mathbb{C})$ où la borne supérieure est prise sur toutes les mesures de Radon positives et supportées sur E telles que $(1/z) * \mu$ soit borné uniformément par 1. En fait, $\gamma_+(E) = \sup |f'(\infty)|$ où la borne supérieure est maintenant prise sur les fonctions f qui sont holomorphes et bornées par 1 en dehors de E et qui sont de la forme $(1/z) * \mu$, où μ est une mesure positive sur E .

Étape 1 [46]. *Il existe une constante $C \geq 1$ telle que pour tout compact $E \subset \mathbb{C}$,*

$$C^{-1} \sup_{\mu} \frac{\mu(\mathbb{C})^{\frac{3}{2}}}{(\mu(\mathbb{C}) + c^2(\mu))^{\frac{1}{2}}} \leq \gamma_+(E) \leq C \sup_{\mu} \frac{\mu(\mathbb{C})^{\frac{3}{2}}}{(\mu(\mathbb{C}) + c^2(\mu))^{\frac{1}{2}}},$$

où la borne supérieure est prise pour toutes les mesures de Radon positives μ supportées sur E à croissance linéaire (de constante 1) et de courbure finie.

L'inégalité de droite avait été démontrée par M. Melnikov [32] pour la capacité analytique, et c'est ce papier qui avait laissé espérer une solution possible au problème de Painlevé. La démonstration de l'étape 1 repose en particulier sur les liens entre courbure de Menger et continuité de l'opérateur de Cauchy. Notons aussi que ces estimations permettent de démontrer facilement que la capacité γ_+ est semi-additive, c'est-à-dire qu'il existe $C > 0$ tel que $\gamma_+(E \cup F) \leq C(\gamma_+(E) + \gamma_+(F))$ pour tous compacts E et F de \mathbb{C} .

Étape 2 [47]. *Il existe une constante $C \geq 1$ telle que, pour tout compact E de \mathbb{C} ,*

$$\gamma_+(E) \leq \gamma(E) \leq C\gamma_+(E).$$

L'inégalité de gauche est évidente par définition des capacités. Nous avons vu, au tout début de la section 2.2.2, que (sous l'hypothèse $H^1(E) < +\infty$) toute fonction holomorphe bornée dans $\mathbb{C} \setminus E$ est de la forme $\mu * (1/z)$ où μ est une mesure COMPLEXE supportée sur E . Le résultat de Tolsa est étonnant, dans la mesure où il nous dit que l'on peut se restreindre dans la définition de la capacité analytique aux seules fonctions holomorphes qui sont obtenues comme convolutions de $1/z$ avec des mesures positives ! Il est clair que les résultats des étapes 1 et 2 impliquent le théorème 1.1. Notons aussi que l'étape 2 et la semi-additivité de la capacité γ_+ (qui découle de l'étape 1, voir plus

haut) impliquent la semi-additivité de la capacité analytique, à savoir qu'il existe une constante $C > 0$ telle que

$$\gamma(E \cup F) \leq C(\gamma(E) + \gamma(F))$$

pour tous compacts E et F de \mathbb{C} . Ceci avait été conjecturé par A.G. Vituskhin dans les années 1960, en particulier à cause des applications potentielles en théorie de l'approximation rationnelle (voir [51]). Nous allons expliquer les idées de la démonstration de Tolsa de l'étape 2 dans le cas où le compact E est la N -ième génération du Cantor 4-coins de Garnett, noté E_N . Rappelons que $E_N = \cup_{j=1}^{4^N} Q_N^j$ où les Q_N^j sont des carrés dont la longueur de côté est 4^{-N} et sont situés dans les coins des carrés de E_{N-1} (voir la section 2.1.2). Nous allons en particulier constater qu'un des ingrédients principaux est le théorème $T(b)$ local de M. Christ. Dans un premier temps, nous devons rappeler que, d'après les travaux d'Eidermann et Tolsa (voir par exemple [45]), il existe une constante $C \geq 1$, telle que pour tout $N \in \mathbb{N}^*$,

$$\frac{C^{-1}}{\sqrt{N}} \leq \gamma_+(E_N) \leq \frac{C}{\sqrt{N}}.$$

Cette estimation découle de l'étape 1 et de calculs de courbure de Menger. En fait, ceci implique qu'il existe une constante $C_1 > 0$ telle que (*) $\gamma_+(E_{N/2}) \leq C_1 \gamma_+(E_N)$ pour tout N pair et (*) $\gamma_+(E_{(N+1)/2}) \leq C_1 \gamma_+(E_N)$ pour tout N impair. Supposons maintenant que la proposition suivante est vraie.

PROPOSITION 2.15. — *Soit N un entier pair (respectivement impair) et supposons qu'il existe une constante $C_0 > 1$ telle $\gamma(E_{\frac{N}{2}}) \leq C_0 \gamma(E_N)$ (respectivement $\gamma(E_{\frac{N+1}{2}}) \leq C_0 \gamma(E_N)$). Alors, il existe une constante absolue $A > 0$ telle que $\gamma(E_N) \leq C_0 A \gamma_+(E_{\frac{N}{2}})$ (respectivement $\gamma(E_N) \leq C_0 A \gamma_+(E_{\frac{N+1}{2}})$).*

Soit $N \in \mathbb{N}$ un entier pair (le cas impair est identique). Nous allons utiliser la proposition précédente pour démontrer que $\gamma(E_N) \leq C_2 \gamma_+(E_N)$ pour une certaine constante C_2 . Pour cela, raisonnons par récurrence et supposons que $\gamma(E_m) \leq AC_1^2 \gamma_+(E_m)$ pour tout $m < N$ (où C_1 est la constante de (*)).

Cas 1. $\gamma(E_N) \leq \frac{1}{C_1} \gamma(E_{\frac{N}{2}})$.

Alors,

$$\begin{aligned} \gamma(E_N) &\leq \frac{1}{C_1} \gamma(E_{\frac{N}{2}}) \\ &\leq AC_1 \gamma_+(E_{\frac{N}{2}}) \text{ d'après l'hypothèse de récurrence} \\ &\leq AC_1^2 \gamma_+(E_N) \text{ d'après (*).} \end{aligned}$$

Cas 2. $\gamma(E_{\frac{N}{2}}) \leq C_1 \gamma(E_N)$.

D'après la proposition 2.15, il vient

$$\begin{aligned} \gamma(E_N) &\leq AC_1 \gamma_+(E_{\frac{N}{2}}) \\ &\leq AC_1^2 \gamma_+(E_N) \text{ d'après (*).} \end{aligned}$$

Il nous reste donc à esquisser une démonstration de la proposition 2.15, en supposant par exemple que N est pair. Considérons la mesure $\lambda = \frac{1}{4}\gamma(E_N)\mu$ où μ est la mesure de longueur sur le bord de $E_{N/2}$. Alors, la mesure λ est portée par le bord de $E_{N/2}$ et satisfait :

- (i) λ est doublante (avec une constante ne dépendant pas de N) ;
- (ii) $\lambda(\mathbb{C}) = \gamma(E_N)$;
- (iii) l'opérateur de Cauchy \mathcal{C}_λ est borné sur $L^2(\lambda)$ et sa norme $\|\mathcal{C}_\lambda\|_{2,2}$ est bornée par $C C_0$ (où $C > 0$ est une constante absolue).

Les points (i) et (ii) sont clairs. Pour le point (iii), nous voulons appliquer le théorème $T(b)$ local de Christ, dans sa version «homogène» d'après le point (i). Ainsi, pour tout Q_n^j , $0 \leq n \leq \frac{N}{2}$, $1 \leq j \leq 4^n$, nous devons construire une fonction b_n^j à support sur Q_n^j (qui sont les «intervalles dyadiques» considérés) et telle que

- (a) $\|b_n^j\|_\infty \leq C C_0$;
- (b) $\|\mathcal{C}_\lambda^\varepsilon(b_n^j)\|_\infty \leq C$ (uniformément en ε);
- (c) $\left| \int_{Q_n^j} b_n^j d\lambda \right| \geq \delta \lambda(Q_n^j)$

où C et δ sont des constantes absolues. Rappelons que la démonstration du théorème de Christ donne aussi que $\|\mathcal{C}_\lambda\|_{2,2} \leq C C_0$ pour une constante absolue $C > 0$, c'est-à-dire l'estimation annoncée dans le point (iii). L'idée pour construire cette famille pseudo-accréitive est la suivante. Tout d'abord, il existe une fonction f , appelée fonction d'Ahlfors, pour laquelle la capacité analytique $\gamma(E_N)$ est atteinte (c'est-à-dire $f'(\infty) = \gamma(E_N)$). Alors, pour tout $z \notin E_N$,

$$f(z) = \frac{-1}{2i\pi} \int_{\partial E_N} \frac{f(s)t(s)}{s-z} ds$$

où $t(s)$ est le vecteur tangent unitaire. Posons $\nu = \frac{-1}{2i\pi} f(s)t(s) ds|_{\partial E_N}$. Alors, ν est supportée sur le bord de E_N et sa transformée de Cauchy est majorée par 1 en dehors de ∂E_N . Commençons par construire b_0^1 . Si nous avons $\nu = b d\lambda$, alors $b = b_0^1$ conviendrait. Mais, ν n'est pas absolument continue par rapport à λ (d'ailleurs, ces deux mesures ont des supports différents). L'idée est d'adapter et de considérer

$$b_0^1 = \sum_{j=1}^{4^{\frac{N}{2}}} \frac{\nu(Q_{\frac{N}{2}}^j)}{\lambda(Q_{\frac{N}{2}}^j)} \chi_{Q_{\frac{N}{2}}^j}.$$

L'estimation cruciale pour obtenir (a) est

$$|\nu(Q_{N/2}^j)| \leq C C_0 \lambda(Q_{N/2}^j)$$

pour tout $1 \leq j \leq 4^{N/2}$. Celle-ci découle d'arguments standards et de notre hypothèse (à savoir $\gamma(E_{\frac{N}{2}}) \leq C_0 \gamma(E_N)$). L'idée pour obtenir (b) est de comparer $\mathcal{C}_\lambda^\varepsilon(b_0^1)$ avec la

transformée de Cauchy de ν qui est bornée. Le point (c) est facile, il découle de l'égalité

$$\int_{Q_0^1} b_0^1 d\lambda = \sum_{j=1}^{4^{N/2}} \nu(Q_{N/2}^j) = \nu(\mathbb{C}) = f'(\infty) = \gamma(E_N) = \lambda(Q_0^1).$$

Pour les autres carrés, la construction précédente s'adapte assez facilement. Voir [40] (pages 108-112) pour une preuve complète. Nous avons maintenant à notre disposition une mesure λ qui satisfait les points (i), (ii) et (iii). Alors, d'après le théorème de Calderón-Zygmund dans le cas homogène (voir section 2.2.1), \mathcal{C}_λ est de type faible $(1, 1)$ avec une constante bornée par $C C_0$ et donc, par un argument de dualité utilisant Hahn-Banach (voir [40] théorème 71), il existe une fonction h à support sur $\partial E_{\frac{N}{2}}$ telle que $0 \leq h \leq 1$, $\lambda(\partial E_{\frac{N}{2}}) \leq 2 \int_{\mathbb{C}} h d\lambda$ et $|\mathcal{C}(hd\lambda)(z)| \leq C C_0$ pour tout $z \notin \partial E_{\frac{N}{2}}$. Ainsi, par définition de γ_+ , il existe une constante $A > 0$ telle que $\gamma(E_N) = \lambda(\mathbb{C}) \leq A C_0 \gamma_+(E_{\frac{N}{2}})$ et donc la proposition 2.15 est démontrée.

Pour la démonstration de l'étape 2 dans le cas d'un compact quelconque du plan complexe, les difficultés pour adapter l'argument précédent ne manquent pas ! En particulier, les espaces considérés ne sont plus homogènes. Donc, nous ne pouvons plus utiliser le théorème de Christ. De plus, la version du théorème $T(b)$ établie par G. David dans [9] pour résoudre la conjecture de Vitushkin s'est avérée insuffisante. La démonstration de X. Tolsa repose sur les versions données dans [35]. En fait, Nazarov-Treil-Volberg ont démontré ces résultats avec l'idée de les utiliser justement pour démontrer l'équivalence entre capacité analytique γ et capacité γ_+ . Il semble utile d'insister sur le fait que l'analyse harmonique non homogène a été développée non pas pour elle-même, mais afin de résoudre les problèmes liés à la capacité analytique.

3. CAPACITÉ ANALYTIQUE ET HOMÉOMORPHISMES BILIPSCHITZIENS

Dans [40] (pages 112-14), nous proposons deux tests afin de déterminer si la solution de Tolsa est une solution satisfaisante ou non au problème de Painlevé. En fait, nous énonçons deux problèmes de nature géométrique autour de la capacité analytique et nous demandons si la réponse de Tolsa permettait de les résoudre. Nous allons dans ce paragraphe expliquer comment Xavier Tolsa a répondu positivement à la question 1 de [40] (page 113). Le second problème est toujours ouvert et sera discuté dans la prochaine section. Commençons par rappeler qu'un homéomorphisme $\phi : \mathbb{C} \rightarrow \mathbb{C}$ est bilipschitzien s'il existe une constante $K \geq 1$ telle que

$$K^{-1}|x - y| \leq |\phi(x) - \phi(y)| \leq K|x - y|,$$

pour tout $x \in \mathbb{C}$, tout $y \in \mathbb{C}$. Le problème qui nous intéresse est de déterminer si les ensembles effaçables sont invariants par le groupe des homéomorphismes bilipschitziens. Dans [48], il est démontré le

THÉORÈME 3.1. — *Soit $\phi : \mathbb{C} \rightarrow \mathbb{C}$ un homéomorphisme bilipshitzien. Alors, il existe une constante $C_0 \geq 1$ ne dépendant que de ϕ telle que*

$$C_0^{-1}\gamma(E) \leq \gamma(\phi(E)) \leq C_0\gamma(E)$$

pour tout compact E dans \mathbb{C} .

Il s'en déduit immédiatement le

COROLLAIRE 3.2. — *Soit $\phi : \mathbb{C} \rightarrow \mathbb{C}$ un homéomorphisme bilipschitzien et soit $E \subset \mathbb{C}$ un ensemble compact et effaçable. Alors, $\phi(E)$ est effaçable.*

Notons que le cas des ensembles de Cantor de type 4-coins est traité dans [17]. La démonstration du théorème 3.1 repose sur le résultat suivant et des estimations contenues dans [47]. Nous pouvons tout de suite voir que le corollaire 3.2 découle du théorème de Tolsa et du théorème 3.3. Nous noterons μ_ϕ la mesure-image d'une mesure μ par un homéomorphisme ϕ .

THÉORÈME 3.3. — *Soit $\phi : \mathbb{C} \rightarrow \mathbb{C}$ un homéomorphisme bilipschitzien. Alors, il existe une constante $C_0 > 0$ ne dépendant que de ϕ telle que*

$$c^2(\mu_\phi) \leq C_0(c^2(\mu) + \mu(E)),$$

pour tout compact E de \mathbb{C} et toute mesure de Radon positive μ supportée sur E , à croissance linéaire (avec une constante 1) et de courbure de Menger finie.

Notons que ce résultat a un analogue pour l'opérateur de Cauchy. En effet, si μ est une mesure de Radon positive (sans atome) sur \mathbb{C} telle que l'opérateur de Cauchy associé \mathcal{C}_μ soit borné sur $L^2(\mu)$, il en est de même pour la mesure image de μ par tout homéomorphisme bilipschitzien du plan complexe (voir [48], théorème 1.3). Observons qu'une majoration du type $c^2(\mu_\phi) \leq C_0 c^2(\mu)$ dans le théorème 3.3 est impossible. En effet, prenons pour E un segment de \mathbb{C} . Alors, toute mesure supportée par E est de courbure de Menger nulle. Ce qui n'est pas obligatoirement le cas de la mesure image. Celle-ci est alors supportée par une courbe corde-arc et il découle du théorème 11 de [41] que $c^2(\mu_\phi) \leq C\mu(\mathbb{C})$ (voir aussi le théorème 2.7 ci-dessus). Donc, la majoration du théorème 3.3 est en quelque sorte optimale. Le point-clé de la démonstration du théorème 3.3 est une décomposition de la couronne de type géométrique inspirée des travaux de G. David et S. Semmes [11] [12]. L'idée est de regrouper les carrés dyadiques usuels de \mathbb{C} en régions de temps d'arrêt, dont le nombre est contrôlé. Sur chacune de ces régions, le support de μ est bien approximé par une courbe Ahlfors-régulière. Nous allons dans un premier temps donner les principes généraux de construction de la décomposition de la couronne de Tolsa, avant de donner un énoncé plus précis. Partons d'une mesure μ comme dans le théorème 3.3 et supposons que μ est supportée dans le carré unité $[0, 1] \times [0, 1]$. Notons Δ la famille des carrés dyadiques usuels de \mathbb{C} et rappelons que Δ est la réunion des familles Δ_j , $j \in \mathbb{Z}$, de carrés dyadiques de génération j : chaque carré $Q \in \Delta_j$ est de la forme $[k.2^j, (k+1).2^j] \times [l.2^j, (l+1).2^j]$. Tout carré $Q \in \Delta_j$ a un unique père, c'est-à-dire un carré $Q' \in \Delta_{j+1}$ tel que $Q \subset Q'$, et

a exactement 4 fils, c'est-à-dire des carrés $\tilde{Q} \in \Delta_{j-1}$ tels que $\tilde{Q} \subset Q$. Pour tout $Q \in \Delta$, posons $D_\mu(Q) = \mu(Q)/\text{diam}Q$. Une région de temps d'arrêt S est une famille de carrés dans Δ telle qu'il existe un carré dyadique (dit maximal) $Q(S) \in S$ tel que

- (i) Pour tout $Q \in S$, $Q \subset Q(S)$.
- (ii) Si $Q \in S$ et si $Q' \in \Delta$ vérifie $Q \subset Q' \subset Q(S)$, alors $Q' \in S$.

En suivant les idées de Tolsa, pour construire une région de temps d'arrêt en dessous d'un carré «raisonnable» Q_0 , nous considérons ses fils, puis ses petits-fils et nous stoppons quand nous rencontrons un carré R_0 pour lequel $D_\mu(Q_0) \gg D_\mu(R_0)$, ou $D_\mu(Q_0) \ll D_\mu(R_0)$ (c'est-à-dire quand la densité de μ dans R_0 est devenue trop «grande» ou trop «petite» par rapport à celle de Q_0) ou la courbure de Menger de μ dans R_0 est devenue trop «grande». Les hypothèses sur la mesure, à savoir qu'elle est à croissance linéaire et de courbure de Menger finie, nous permettent d'espérer que nous ne stoppons pas trop souvent en effectuant cette procédure. Ceci est primordial car, dans le cas contraire, nous aurions presque autant de cubes dyadiques que de régions de temps d'arrêt ! Pour nous, un cube «raisonnable» est un cube (a, b) -doublant, c'est-à-dire tel que $\mu(aQ) \leq b\mu(Q)$ avec a et b bien choisis. Cette notion est importante car, dans un tel cube, la mesure se comporte presque comme une mesure doublante. Pour des raisons techniques, les carrés maximaux de Tolsa ne sont pas dyadiques, mais des carrés 4-dyadiques, c'est-à-dire de la forme $[k.2^{-j}, (k+4).2^{-j}] \times [l.2^{-j}, (l+4).2^{-j}]$. En utilisant toutes ces idées, X. Tolsa construit une famille, notée $\text{TOP}(\mu)$, de carrés 4-dyadiques qui sont $(16, 5000)$ -doublants tels que

$$(P1) \quad \sum_{Q \in \text{TOP}(\mu)} D_\mu(Q)^2 \mu(Q) \leq C(\mu(\mathbb{C}) + c^2(\mu)).$$

Ces carrés de $\text{TOP}(\mu)$ sont les carrés maximaux décrits précédemment. Ainsi, (P1) est la formulation mathématique du fait que nous ne stoppons pas souvent. De plus, $\text{supp}(\mu)$ est bien approximé dans $Q \in \text{TOP}(\mu)$ par une courbe Ahlfors-régulière, en un sens que nous allons maintenant expliciter. Soit $Q \in \text{TOP}(\mu)$; notons $\text{Stop}(Q)$ l'ensemble des carrés $R \in \text{TOP}(\mu)$ tels que

- (i) $R \cap 3Q \neq \emptyset$.
- (ii) $\text{diam}(R) \leq (1/8)\text{diam}Q$.
- (iii) Il n'existe pas de $R' \in \text{TOP}(\mu)$ qui satisfait (i) et (ii) et qui contient R .

Les carrés de $\text{Stop}(Q)$ sont les carrés en dessous de Q pour lesquels nous avons stoppé dans la procédure décrite précédemment. De plus, notons $B_Q = (3Q \cap \text{Supp}\mu) \setminus (Z \cup \bigcup_{R \in \text{Stop}(Q)} R)$ (où Z est un ensemble de points qui ne sont pas contenus dans beaucoup

de carrés doublants ; il satisfait en fait à $\mu(Z) = 0$). Les points de B_Q sont ceux pour lesquels nous n'avons jamais stoppé. Le point-clé est que, par construction, nous contrôlons la densité ainsi que la courbure de Menger de μ dans la région de temps d'arrêt en dessous de Q (à toutes les échelles pour un point de B_Q et à l'échelle de R pour un point de $R \in \text{Stop}(Q)$). Nous pouvons alors utiliser les résultats du paragraphe 3.1. Ainsi, il existe une courbe Ahlfors-régulière Γ_Q de constante uniforme

(c'est-à-dire ne dépendant que des données de μ) telle que

(P2) $B_Q \subset \Gamma(Q)$.

(P3) Pour tout $R \in \text{Stop}(Q)$, il existe un carré \tilde{R} contenant R tel que $\delta(R, \tilde{R}) \leq CD_\mu(R)$ et $\tilde{R} \cap \Gamma_Q \neq \emptyset$.

Ici, $\delta(R, S) = \int_{S_R \setminus \tilde{R}} \frac{1}{|y - c_R|} d\mu(y)$ où c_R est le centre de R et S_R est le plus petit carré concentrique avec R et contenant S . Dans certains cas, (P3) dit que $d(R, \Gamma(Q)) \leq C \cdot \text{diam} R$, c'est-à-dire que les points de R sont proches de Γ_Q , à l'échelle de R . Enfin, les carrés de $\text{TOP}(\mu)$ vérifient une condition technique sur la densité d'un carré R par rapport à celle de son cube maximal Q , à savoir

(P4) Soit R un carré tel que $\text{diam} R \leq \text{diam} Q$. Supposons que $R \cap G_Q \neq \emptyset$, ou bien qu'il existe un carré $\tilde{R} \in \text{Stop}(Q)$ tel que $R \cap \tilde{R} \neq \emptyset$ et $\text{diam} \tilde{R} \leq \text{diam} R$. Alors $\mu(R) \leq CD_\mu(Q)l(R)$.

La décomposition de la couronne associée à μ est la donnée d'une famille de carrés $\text{TOP}(\mu)$ qui satisfait (P1), (P2), (P3) et (P4). Il est important de noter que ces conditions sont stables par homéomorphismes bilipschitziens. Notons aussi que cette décomposition de la couronne a permis à Tolsa [49] de démontrer que la continuité L^2 de l'opérateur de Cauchy implique la continuité L^2 de tous les opérateurs de Calderón-Zygmund (associés à des noyaux antisymétriques que nous devons supposer en plus suffisamment réguliers), tous ces opérateurs étant définis par respect à une mesure qui n'est pas supposée doublante.

4. CAPACITÉ ANALYTIQUE ET LONGUEUR DE FAVARD

D'après le fait 4 et le théorème 2.2, tout ensemble $E \subset \mathbb{C}$ est effaçable si et seulement si $\text{Fav}(E) = 0$ SOUS LA CONDITION QUE $H^1(E) < +\infty$. Si nous supprimons cette dernière hypothèse, le résultat devient faux comme l'avait démontré P. Mattila dans [25]. En fait, il prouvait que la condition d'être effaçable est invariante par représentations conformes, ce qui n'est pas le cas pour la condition d'être de longueur de Favard nulle (pour les ensembles de 1-mesure de Hausdorff infinie). Ainsi, son argument ne permettait pas de déterminer quelle implication était fautive. Plus tard, P. Jones et T. Murai [19] construisent un ensemble du plan complexe qui est de longueur de Favard nulle, mais qui n'est pas effaçable. Un exemple plus simple a été donné par H. Joyce et P. Mörters récemment [20]. Ceci nous amène donc au

Problème 1. Soit $E \subset \mathbb{C}$ compact. Supposons que $\text{Fav}(E) > 0$. L'ensemble est-il ou non effaçable ?

D'après le théorème de Tolsa, il suffit de construire une mesure supportée sur E qui soit à croissance linéaire et de courbure de Menger finie. Un exemple typique d'ensemble qui a de «grosses projections» est un continuum non réduit à un point.

Dans ce cas-là, une telle mesure est construite dans [40], mais il n'est pas clair que cet argument puisse être adapté au cas général. En fait, il faudrait dans un premier temps comprendre le lien entre la longueur de Favard et la courbure de Menger. Nous allons dans la suite de ce paragraphe discuter de quelques problèmes dans cette direction.

Nous avons vu que le théorème de Besicovitch (théorème 2.2) permet de caractériser les ensembles purement non rectifiables en termes de longueur de Favard (toujours sous l'hypothèse de finitude de la mesure de Hausdorff). Il serait intéressant d'avoir une version plus quantitative de ce résultat.

Problème 2. Soit $\delta > 0$. Existe-t-il une constante $C = C(\delta) > 0$ telle que, pour tout compact $E \subset \mathbb{C}$ vérifiant $H^1(E) < +\infty$ et $\text{Fav}(E) \geq \delta$, il existe une courbe rectifiable Γ satisfaisant $H^1(E \cap \Gamma) \geq C$?

Dans le même ordre d'idée, nous aimerions comprendre le lien entre longueur de Favard et rectifiabilité uniforme. N'oublions pas que celle-ci est équivalente à la condition de courbure locale (théorème 2.6).

Problème 3. Soit $E \subset \mathbb{C}$ un ensemble Ahlfors-régulier de dimension 1. Supposons qu'il existe une constante $\eta > 0$ telle que, pour tout $x \in E$, tout $R \in]0, \text{diam}E[$,

$$\text{Fav}(E \cap B(x, R)) \geq \eta R.$$

Alors, E est-il uniformément rectifiable ?

Cette question est abordée (sous une forme légèrement différente) dans [13] où il est démontré que la réponse est positive sous une hypothèse de platitude en termes des nombres β de Peter Jones. Cependant, les arguments de G. David et S. Semmes ne permettent pas de conclure dans le cas général. Notons que, d'après le théorème 2.2, un ensemble vérifiant les hypothèses du problème 3 est rectifiable. Considérons maintenant l'ensemble de Cantor 4-coins que nous noterons E . Nous avons vu au paragraphe 2.1.2 que $E = \bigcap_{N \in \mathbb{N}} E_N$ où E_N est formé de 4^N carrés de longueur de côtés 4^{-N} , qui sont situés dans les coins des carrés de E_{N-1} . Nous avons aussi vu que $\text{Fav}(E) = 0$, mais il serait intéressant de voir à quelle vitesse $\text{Fav}(E_N)$ tend vers 0.

Problème 4. Existe-t-il une constante $C \geq 1$ telle que $\text{Fav}(E_N) \leq C/N$ pour tout $N \in \mathbb{N}^*$?

Il découle du théorème 1.4 de [26] qu'il existe une constante $c \leq 1$ telle que $\text{Fav}(E_N) \geq c/N$ pour tout $N \in \mathbb{N}^*$. Une réponse positive au problème 4 impliquerait que $\text{Fav}(E_N)$ et $1/c^2(E_N)$ (où $c^2(E_N)$ est la courbure de la restriction de H^1 au bord de E_N) doivent être comparables, puisque $c^2(E_N)$ est comparable à N (voir [40]). Ceci est à première vue assez surprenant ! Dans [42], il est donné un majorant dans lequel le N est remplacé

par un logarithme itéré. Y. Peres et B. Solomyak proposent à la fin de leur article une approche plus générale, en relation avec la théorie quantitative de la rectifiabilité. Ainsi, si E est un sous-ensemble compact de \mathbb{C} , posons pour tout $\varepsilon > 0$,

$$l(E, \varepsilon) = \sup H_{\infty}^1(\Gamma(\varepsilon) \cap E)$$

où la borne supérieure est prise sur toutes les courbes rectifiables de \mathbb{C} de longueur 1 et où $\Gamma(\varepsilon)$ désigne le ε -voisinage de Γ . Le contenu de Hausdorff H_{∞}^1 est défini par

$$H_{\infty}^1(F) = \inf \left\{ \sum_i (\text{diam} U_i); F \subset \cup_i U_i \right\}.$$

En particulier, si F est compact, $H_{\infty}^1(F) \leq \text{diam} F$. Peres et Solomyak demandent si $\text{Fav}(E(\varepsilon)) = O(l(E, \varepsilon))$ quand $\varepsilon \rightarrow 0$ (où $E(\varepsilon)$ désigne toujours le ε -voisinage de E). Nous laissons le soin au lecteur de vérifier qu'une réponse positive à la question précédente implique une réponse positive au problème 4.

Soit $\lambda = (\lambda_j)$ une suite de réels telle que

- (i) Pour tout $j \in \mathbb{N}$, $\frac{1}{4} \leq \lambda_j \leq \frac{1}{3}$.
- (ii) La suite λ tend vers $1/4$ en décroissant.

Nous allons associer à cette suite λ un ensemble de Cantor de type 4-coins, noté $E(\lambda)$, de la façon suivante. Posons $\sigma_j = \prod_{k=1}^j \lambda_k$; soit E_0 le carré unité $[0, 1] \times [0, 1]$ de \mathbb{C} . L'ensemble E_1 est l'union des quatre carrés situés dans les coins de E_0 et de longueur de côté σ_1 . L'ensemble E_2 est l'union des 16 carrés de longueur de côté σ_2 situés dans les coins des carrés de E_1 . De manière générale, E_j est l'union de 4^j carrés de longueur de côté σ_j , chacun de ces carrés étant dans le coin d'un carré de E_{j-1} . Alors, $E(\lambda) = \bigcap_j E_j$. L'ensemble de Cantor 4-coins de Garnett correspond à la suite constante égale à $1/4$. Notons que $H^1(E(\lambda)) < +\infty$ si et seulement si $\sup_j 4^j \sigma_j < +\infty$. Nous savons grâce à [24] que $E(\lambda)$ est effaçable si et seulement si $\sum_j 4^{-2j} \sigma_j^{-2} = \infty$ (car dans ce cas, la courbure de Menger associée à la mesure naturelle sur $E(\lambda)$ est infinie, voir le théorème 44 de [40]). Cependant, les projections de $E(\lambda)$ sont mal comprises.

Problème 5. Pour quelles suites λ , a-t-on $\text{Fav}(E(\lambda)) = 0$?

Il est important de noter que la résolution d'un des problèmes précédents passe par une meilleure compréhension des rapports entre les nombres géométriques que sont les nombres β , la courbure de Menger et la longueur de Favard, et donc que résoudre un de ces problèmes devrait permettre d'avoir des éléments de réponse concernant les autres. Nous renvoyons à [29] pour d'autres problèmes concernant les projections et la rectifiabilité (et des commentaires sur les problèmes décrits plus haut).

5. ET EN DIMENSIONS SUPÉRIEURES ?

Un ensemble $E \subset \mathbb{R}^n$ est effaçable pour les fonctions harmoniques lipschitziennes si pour tout ouvert U contenant E , toute fonction localement lipschitzienne $f : U \rightarrow \mathbb{C}$ qui est harmonique dans $U \setminus E$ est harmonique dans U tout entier. Notre problème est alors de caractériser géométriquement ces ensembles effaçables. Ceci est un analogue naturel dans \mathbb{R}^n du problème de Painlevé, dans la mesure où nous pouvons aussi lui associer une capacité et des intégrales singulières comme nous allons le voir maintenant. Notons aussi que, si $n = 2$, l'effaçabilité pour les fonctions holomorphes bornées implique l'effaçabilité pour les fonctions harmoniques lipschitziennes, puisque si f est harmonique lipschitzienne, alors $\partial f / \partial z$ est holomorphe bornée. En fait, si nous nous restreignons aux ensembles de 1-mesure de Hausdorff finie, ces deux notions d'effaçabilité coïncident d'après le fait 4 et le résultat principal de [10]. Introduisons maintenant une famille de capacités. Soit α avec $0 < \alpha < n$. Pour tout compact $E \subset \mathbb{R}^n$, sa capacité associée aux potentiels de Riesz signés $R_{i,\alpha} = x_i/|x|^{1+\alpha}$ ($1 \leq i \leq n$), que nous noterons $\gamma_\alpha(E)$, est définie par

$$\gamma_\alpha(E) = \sup | \langle T, 1 \rangle |$$

où la borne supérieure est prise sur toutes les distributions réelles T supportées sur E telles que pour tout $1 \leq i \leq n$, $T * (x_i/|x|^{1+\alpha})$ soit une fonction bornée sur \mathbb{R}^n par 1. Ici, $\langle T, 1 \rangle$ est l'action de la distribution T sur la fonction constante 1. Le lien avec la capacité analytique γ est clair en notant que, pour tout ensemble compact $E \subset \mathbb{C}$, $\gamma(E) = \sup | \langle T, 1 \rangle |$, où la borne supérieure est prise sur toutes les distributions complexes T supportées sur E telles que $T * (1/z)$ soit une fonction bornée sur \mathbb{R}^2 par 1. Le cas $\alpha = n - 1$ est particulier et, dans ce cas-là, la capacité γ_{n-1} s'appelle la capacité harmonique lipschitzienne. Alors, $E \subset \mathbb{R}^n$ est effaçable (pour les fonctions harmoniques lipschitziennes) si et seulement si $\gamma_{n-1}(E) = 0$. Voir [31] où il est aussi démontré des analogues des faits 1, 2 et 3. Ainsi, si $H^{n-1}(E) = 0$, alors E est effaçable alors que, si $\text{Hdim}(E) > n - 1$, il n'est pas effaçable. De plus, si E est contenu dans une hypersurface régulière (par exemple, une image lipschitzienne ou bilipschitzienne de \mathbb{R}^{n-1}), $H^{n-1}(E)$ et $\gamma_{n-1}(E)$ sont comparables. Qu'en est-il du fait 4 ? Nous avons vu que l'ingrédient miraculeux dans \mathbb{R}^2 est la courbure de Menger qui permet de faire le lien entre les propriétés de rectifiabilité et la continuité de l'opérateur de Cauchy. Dans le cas de l'effaçabilité pour les fonctions harmoniques lipschitziennes dans \mathbb{R}^n , les intégrales singulières qui interviennent naturellement sont les transformées de Riesz R_i ($1 \leq i \leq n$) qui sont associées aux noyaux $x_i/|x|^n$. Nous ne savons pas si une courbure adaptée aux transformées de Riesz existe si $n > 2$ (et il est peut-être même impossible d'en trouver une, voir [15]). Dans ces conditions, c'est un véritable tour de force qu'a réalisé A. Volberg dans [52] où il démontre que la capacité harmonique lipschitzienne est semi-additive (en suivant plus ou moins le plan de la preuve de X. Tolsa dans le cas de la capacité analytique). Le manque de courbure est suppléé par les techniques d'analyse harmonique sans doublement du volume qu'ont développées Nazarov-Treil-Volberg.

Notons aussi que des estimations précises de la capacité harmonique lipschitzienne d'analogues n -dimensionnels des ensembles de Cantor 4-coins sont données dans [23]. Dans ce cas, le fait que ces ensembles sont décrits de façon explicite permet de pallier l'absence de courbure.

Signalons enfin qu'une étude systématique des capacités γ_α est initiée dans [22] et [43]. En particulier, L. Prat démontre que $\gamma_\alpha(E) = 0$ si $0 < \alpha < 1$ et E est un sous-ensemble compact de \mathbb{R}^n tel que $H^\alpha(E) < \infty$ ou si α n'est pas entier ($0 < \alpha < n$) et E est un sous-ensemble compact de \mathbb{R}^n qui est de plus Ahlfors-régulier de dimension α .

En paraphrasant un (très) récent premier ministre français, concluons en disant que la route menant à la résolution de tous ces problèmes dans \mathbb{R}^n paraît bien longue et plutôt sinueuse...

RÉFÉRENCES

- [1] L. AHLFORS – *Bounded analytic functions*, Duke Mathematical Journal **14** (1947), 1-11.
- [2] L. AHLFORS, A. BEURLING – *Conformal invariants and function theoretic null-sets*, Acta Mathematica **83** (1950), 101-129.
- [3] A.P. CALDERÓN – *Cauchy integrals on Lipschitz curves and related operators*, Proceedings of the National Academy of Sciences of the USA **74** (1977), 1324-1327.
- [4] M. CHRIST – *Lectures on singular integral operators*, Regional Conference Series in Mathematics **77**, American Mathematical Society (1990).
- [5] M. CHRIST – *A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral*, Colloquium Mathematicum **60-61** (1990), 601-628.
- [6] R. COIFMAN, G. WEISS – *Analyse harmonique non-commutative sur certains espaces homogènes*, Lecture Notes in Math. **242**, Springer-Verlag (1971).
- [7] R. COIFMAN, A. MCINTOSH, Y. MEYER – *L'opérateur de Cauchy définit un opérateur borné sur L^2 sur les courbes lipschitziennes*, Annals of Mathematics **116** (1982), 361-388.
- [8] G. DAVID – *Wavelets and singular integral operators on curves and surfaces*, Lecture Notes in Math. **1465** (1991), Springer-Verlag.
- [9] G. DAVID – *Unrectifiable 1-sets have vanishing analytic capacity*, Revista Matemática Iberoamericana **14** (1998), 369-479.
- [10] G. DAVID, P. MATTLA – *Removable sets for Lipschitz harmonic functions in the plane*, Revista Matemática Iberoamericana **16** (2000), 137-215.
- [11] G. DAVID, S. SEMMES – *Singular integrals and rectifiable sets in \mathbb{R}^n : Au-delà des graphes lipschitziens*, Astérisque **193** (1991).

- [12] G. DAVID, S. SEMMES – *Analysis of and on uniformly rectifiable sets*, Mathematical Surveys and Monographs **38**, American Mathematical Society (1993).
- [13] G. DAVID, S. SEMMES – *Quantitative rectifiability and Lipschitz mappings*, Transactions of the American Mathematical Society **337** (1993), 855-889.
- [14] A. DENJOY – *Sur les fonctions analytiques uniformes à singularités discontinues*, Comptes rendus de l'Académie des Sciences de Paris **149** (1909), 258-260.
- [15] H. FARAG – *The Riesz kernels do not give rise to higher dimensional analogues to the Menger-Melnikov curvature*, Publicacions Matemàtiques **43** (1999), 251-260.
- [16] J.B. GARNETT – *Positive length but zero analytic capacity*, Proceedings of the American Mathematical Society **21** (1970), 696-699.
- [17] J.B. GARNETT, J. VERDERA – *Analytic capacity, bilipschitz maps and Cantor sets*, Mathematical Research Letters **10** (2003), 515-522.
- [18] P. JONES – *Rectifiable sets and the traveling salesman problem*, Inventiones Math. **102** (1990), 1-15.
- [19] P. JONES, T. MURAI – *Positive analytic capacity, but zero Buffon needle probability*, Pacific Journal of Mathematics **133** (1988), 99-114.
- [20] H. JOYCE, P. MÖRTERS – *A set with finite curvature and projections of zero length*, Journal of Mathematical Analysis and Applications **247** (2000), 126-135.
- [21] J.-C. LÉGER – *Rectifiability and Menger curvature*, Annals of Mathematics **149** (1999), 831-869.
- [22] J. MATEU, L. PRAT, J. VERDERA – *The capacity associated to signed Riesz kernels and Wolff potentials*, à paraître dans Journal für die reine und angewandte Mathematik.
- [23] J. MATEU, X. TOLSA – *Riesz transforms and harmonic Lip_1 -capacity of Cantor sets*, à paraître dans Proceedings of the London Mathematical Society.
- [24] J. MATEU, X. TOLSA, J. VERDERA – *The planar Cantor sets of zero analytic capacity and the local $T(b)$ theorem*, Journal of the American Mathematical Society **16** (2003), 19-28.
- [25] P. MATTILA – *Smooth maps, null-sets for integral geometric measures and analytic capacity*, Annals of Mathematics **123** (1986), 303-309.
- [26] P. MATTILA – *Orthogonal projections, Riesz capacities, and Minkowski content*, Indiana University Journal **39** (1990), 185-198.
- [27] P. MATTILA – *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics **44**, Cambridge University Press (1995).
- [28] P. MATTILA – *On the analytic capacity and curvature of some Cantor sets with non- σ -finite length*, Publicacions Matemàtiques **40** (1996), 195-204.
- [29] P. MATTILA – *Hausdorff dimension, projections, and the Fourier transform*, Publicacions Matemàtiques **48** (2004), 3-48.
- [30] P. MATTILA, M. MELNIKOV, J. VERDERA – *The Cauchy integral, analytic capacity, and uniform rectifiability*, Annals of Mathematics **144** (1996), 127-136.

- [31] P. MATTILA, P. V. PARAMONOV – *On geometric properties of harmonic Lip_1 -capacity*, Pacific Journal of Mathematics **171** (1995), 469-491.
- [32] M. MELNIKOV – *Analytic capacity: discrete approach and curvature of measure*, Sbornik Mathematics **186** (1995), 827-846.
- [33] M. MELNIKOV, J. VERDERA – *A geometric proof of the L^2 boundedness of the Cauchy integral on Lipschitz curves*, International Mathematical Research Notices **7** (1995), 325-331.
- [34] F. NAZAROV, S. TREIL, A. VOLBERG – *Cauchy integral and Calderón-Zygmund operators on nonhomogeneous spaces*, International Mathematical Research Notices **15** (1997), 703-726.
- [35] F. NAZAROV, S. TREIL, A. VOLBERG – *T_b theorems on nonhomogeneous spaces*, Acta Mathematica **190** (2003), 151-239.
- [36] F. NAZAROV, S. TREIL, A. VOLBERG – *Nonhomogeneous T_b theorem which proves Vitushkin's conjecture*, Preprint n° 519, CRM Barcelona (2002).
- [37] K. OKIKIOLU – *Characterizations of subsets of rectifiable curves in \mathbb{R}^n* , Journal of the London Mathematical Society **46** (1992), 336-348.
- [38] P. PAINLEVÉ – *Leçons sur la théorie analytique des équations différentielles professées à Stockholm*, Hermann (1897).
- [39] H. PAJOT – *Conditions quantitatives de rectifiabilité*, Bulletin de la Société Mathématique de France **125** (1997), 15-53.
- [40] H. PAJOT – *Analytic capacity, rectifiability, Menger curvature and the Cauchy integral*, Lecture Notes in Math. **1799** (2002), Springer-Verlag, Berlin.
- [41] H. PAJOT – *Le problème géométrique du voyageur de commerce, et ses applications à l'analyse complexe et harmonique*, à paraître dans *Autour du centenaire de la mesure de Lebesgue*, Panoramas et Synthèses, Société Mathématique de France.
- [42] Y. PERES, B. SOLOMYAK – *How likely is Buffon's needle to fall near a planar Cantor set?*, Pacific Journal of Mathematics **204** (2002), 473-496.
- [43] L. PRAT – *Potential theory of signed Riesz kernels: Capacity and Hausdorff measures*, International Mathematics Research Notices **19** (2004), 937-981.
- [44] E. M. STEIN – *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press (1993).
- [45] X. TOLSA – *L^2 -boundedness of the Cauchy integral for continuous measures*, Duke Mathematical Journal **98** (1999), 269-304.
- [46] X. TOLSA – *On the analytic capacity γ_+* , Indiana University Journal **51** (2002), 317-343.
- [47] X. TOLSA – *Painlevé's problem and the semiadditivity of analytic capacity*, Acta Mathematica **190** (2003), 105-149.
- [48] X. TOLSA – *Bilipschitz maps, analytic capacity, and the Cauchy integral*, à paraître dans *Annals of Mathematics*.

- [49] X. TOLSA – *The L^2 boundedness of the Cauchy transform implies L^2 boundedness of all antisymmetric Calderón-Zygmund operators*, à paraître dans *Publicacions Matemàtiques*.
- [50] J. VERDERA – *The Fall of the doubling condition in Calderón-Zygmund theory*, in «Proceedings of the 6th International Conference on Harmonic Analysis and Partial Differential Equations (El Escorial 2000)», Numéro spécial de *Publicacions Matemàtiques* (2002), 275-292.
- [51] A. G. VITUSHKIN – *The analytic capacity of sets in approximation theory*, *Uspekhi Matematicheskikh Nauk* **22** (1967), 141-199 (en russe) ; traduction en anglais dans *Russian Mathematical Surveys* **22** (1967), 139-200.
- [52] A. VOLBERG – *Calderón-Zygmund capacities and operators on nonhomogeneous spaces*, *Regional Conference Series in Mathematics* **100**, American Mathematical Society (2003).

HERVÉ PAJOT

Institut Fourier

Université de Grenoble I

Mathématiques, UMR 5582

B.P. 74

F-38402 SAINT-MARTIN D'HÈRES cédex

E-mail : herve.pajot@ujf-grenoble.fr

ÉTATS QUASI-LIBRES LIBRES ET FACTEURS DE TYPE III
[d'après D. Shlyakhtenko]

par **Stefaan VAES**

Le but de cet exposé est de présenter une famille d'algèbres de von Neumann introduite par Shlyakhtenko et de donner un aperçu des résultats de classification des algèbres de cette famille. Ces algèbres de von Neumann sont construites dans le cadre des probabilités libres de Voiculescu.

Murray et von Neumann ont initié la classification des algèbres de von Neumann. Ils ont démontré que chaque algèbre de von Neumann s'écrit comme intégrale directe de facteurs (algèbres de von Neumann de centre trivial) et ils ont classifié les facteurs en différents types : I, II et III. Dans sa thèse [7], Connes a raffiné cette classification en introduisant les sous-types III_λ ($0 \leq \lambda \leq 1$). La construction de Shlyakhtenko donne tout un monde d'exemples de facteurs de type III_1 , c'est-à-dire le plus haut dans la « hiérarchie » des facteurs.

L'idée de Shlyakhtenko est de donner dans le cadre des probabilités libres de Voiculescu une version du foncteur CAR (relations d'anticommutation canoniques) et des états quasi-libres associés. Le foncteur CAR associe à chaque espace de Hilbert H la C^* -algèbre unifère universelle $CAR(H)$ engendrée par la famille $\{a(\xi) \mid \xi \in H\}$ telle que

1. $\xi \mapsto a(\xi)$ est linéaire,
2. les relations d'anticommutation canoniques sont vérifiées

$$\begin{aligned}a(\eta)a(\xi)^* + a(\xi)^*a(\eta) &= \langle \eta, \xi \rangle 1, \\ a(\xi)a(\eta) + a(\eta)a(\xi) &= 0.\end{aligned}$$

Les relations d'anticommutation canoniques peuvent être réalisées par les opérateurs de création sur *l'espace de Fock antisymétrique*. Tout opérateur S agissant sur H tel que $0 \leq S \leq 1$ donne lieu à un état ω_S de la C^* -algèbre $CAR(H)$, qu'on appelle *état quasi-libre de covariance S* . Les représentations GNS [8] associées aux états quasi-libres ont été beaucoup étudiées [19, 15, 1]. Dans une telle représentation, le bicommutant $CAR(H)''$ est une algèbre de von Neumann. Chaque état quasi-libre sur $CAR(H)$ donne donc une algèbre de von Neumann qui s'avère être un facteur. Les travaux d'Araki & Woods [2] et de Powers & Størmer [15] permettent de déterminer le type de ces *facteurs*

d'Araki-Woods⁽¹⁾. Plus précisément, si $0 < \lambda \leq 1$, il existe exactement une classe d'isomorphisme de facteurs d'Araki-Woods de type III_λ et il existe une famille non-dénombrable de facteurs d'Araki-Woods de type III_0 et mutuellement non-isomorphes.

Remarquons que les facteurs d'Araki-Woods sont des *facteurs moyennables*. Dans [5] Connes a réussi à donner une classification complète des facteurs moyennables. C'est un des résultats les plus profonds de la théorie des algèbres de von Neumann. Pour chacun des types $\text{II}_1, \text{II}_\infty, \text{III}_\lambda$ ($0 < \lambda \leq 1$) il existe un unique facteur moyennable. Les facteurs moyennables de type III_0 sont classifiés par un invariant en théorie ergodique qu'on appelle *flot des poids*. L'unicité du facteur moyennable de type III_1 est dû à Haagerup [10]. Dans le cadre des probabilités libres, nous allons obtenir des facteurs non-moyennables. En particulier Shlyakhtenko construit une famille non-dénombrable de facteurs de type III_1 , non-moyennables et mutuellement non-isomorphes.

L'analogie en probabilités libres du foncteur CAR associé à un espace de Hilbert réel $H_{\mathbb{R}}$, la C^* -algèbre universelle $\Gamma(H_{\mathbb{R}})$ engendrée par la famille $\{s(\xi) \mid \xi \in H_{\mathbb{R}}\}$ telle que

- $s(\xi)$ soit auto-adjoint pour tout $\xi \in H_{\mathbb{R}}$,
- $\xi \mapsto s(\xi)$ soit \mathbb{R} -linéaire,
- $\|s(\xi)\| \leq \|\xi\|$ pour tout $\xi \in H_{\mathbb{R}}$.

Pour chaque plongement isométrique $H_{\mathbb{R}} \hookrightarrow H$ de $H_{\mathbb{R}}$ dans un espace de Hilbert H , on obtient une représentation de $\Gamma(H_{\mathbb{R}})$ sur *l'espace de Fock plein*

$$\mathcal{F}(H) = \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H^{\otimes n},$$

en posant $s(\xi) = (\ell(\xi) + \ell(\xi)^*)/2$ où $\ell(\xi)$ est l'opérateur de création.

La construction des facteurs d'Araki-Woods libres suppose la donnée d'un groupe à un paramètre (U_t) de transformations orthogonales d'un espace de Hilbert réel $H_{\mathbb{R}}$ qui induit un plongement isométrique $H_{\mathbb{R}} \hookrightarrow H$ de $H_{\mathbb{R}}$ dans le complexifié H . On obtient une représentation de $\Gamma(H_{\mathbb{R}})$ dont l'image est notée $\Gamma(H_{\mathbb{R}}, U_t)$. Le facteur d'Araki-Woods libre $\Gamma(H_{\mathbb{R}}, U_t)''$ est l'algèbre de von Neumann engendrée par $\Gamma(H_{\mathbb{R}}, U_t)$. La restriction de l'état du vide au facteur $\Gamma(H_{\mathbb{R}}, U_t)''$ est *l'état quasi-libre libre* noté φ_U . Alors, $\Gamma(H_{\mathbb{R}}, U_t)''$ est un facteur de type III, sauf si $U_t = \text{id}$ pour tout $t \in \mathbb{R}$.

Nous commençons cet exposé par rappeler la classification des facteurs de Connes et les probabilités libres de Voiculescu et par une présentation de plusieurs points de vue sur nos données essentielles, les représentations orthogonales de \mathbb{R} . Au §2 nous définissons l'algèbre de von Neumann $\Gamma(H_{\mathbb{R}}, U_t)''$ avec l'état quasi-libre libre φ_U . Nous étudions en détail le cas $H_{\mathbb{R}} = \mathbb{R}^2$ muni de la représentation de \mathbb{R} par rotations. Au §3 nous présentons les principaux résultats de classification et de non-isomorphisme des facteurs d'Araki-Woods libres obtenus par Shlyakhtenko :

⁽¹⁾Il découle des travaux de Powers & Størmer que les facteurs associés aux états quasi-libres sont des produits tensoriels infinis de matrices 2 fois 2 (des facteurs ITPFI₂). Plus généralement, Araki & Woods étudient et déterminent le type des produits tensoriels infinis de facteurs de type I (les ITPFI) et il y a des ITPFI qui ne sont pas ITPFI₂.

- Classification complète des facteurs d’Araki-Woods libres associés à une représentation (U_t) presque-périodique.
- Construction d’une famille non-dénombrable de facteurs d’Araki-Woods libres non-presque-périodiques et mutuellement non-isomorphes. Ces facteurs sont distingués par leur invariant τ de Connes.
- Construction de deux facteurs d’Araki-Woods libres non-isomorphes et ayant le même invariant τ de Connes. Ceci est une application de la notion de *dimension entropique libre* introduite par Voiculescu [29, 30].
- Démonstration que la classe d’isomorphisme d’un facteur Araki-Woods libre peut dépendre de la multiplicité de la représentation (U_t) . Ce résultat est démontré à l’aide de la notion d’algèbre de von Neumann solide due à Ozawa [12].

La classification complète des facteurs d’Araki-Woods libres reste un problème ouvert. Les résultats de non-isomorphisme présentés au §3 montrent qu’un facteur d’Araki-Woods libre $\Gamma(H_{\mathbb{R}}, U_t)''$ dépend fortement de la classe de la mesure spectrale de la représentation (U_t) . Ce problème de classification est beaucoup plus difficile que la classification des facteurs d’Araki-Woods, pour la raison suivante. Powers et Størmer [15] démontrent essentiellement que deux facteurs d’Araki-Woods associés à des états quasi-libres sont isomorphes si leurs opérateurs de covariance diffèrent d’un opérateur d’Hilbert-Schmidt. En particulier, d’après un résultat de von Neumann, il suffit de considérer le cas d’un opérateur de covariance diagonalisable. Ceci n’est plus le cas pour les facteurs d’Araki-Woods libres. Une grande partie des facteurs d’Araki-Woods libres ne peut être obtenue par des représentations orthogonales presque-périodiques.

Dans le dernier §4 nous présentons un nouveau résultat sur les produits libres, ce qui permet au §2.3 de démontrer un résultat un peu plus général que dans l’article [23].

Il y a un certain nombre de résultats et d’applications dans la théorie des facteurs d’Araki-Woods libres dont on ne parlera pas en détail dans cet exposé. Notons que Shlyakhtenko a démontré dans [22] que les facteurs d’Araki-Woods libres T_λ de type III_λ ($0 < \lambda < 1$) sont des *facteurs premiers* : ils ne peuvent être écrits comme produit tensoriel de deux facteurs diffus (sans projecteurs minimaux). Dans [13] Pisier et Shlyakhtenko utilisent des facteurs d’Araki-Woods libres comme modèles pour démontrer une *inégalité de Grothendieck* pour les espaces d’opérateurs. Dans [28] les facteurs d’Araki-Woods libres sont utilisés pour construire des *actions extérieures* de groupes quantiques localement compacts.

Je remercie S. Baaĵ, E. Germain, D. Shlyakhtenko et G. Skandalis pour leur aide pendant la préparation de cet exposé.

1. RAPPELS

1.1. Algèbres de von Neumann. Type des facteurs

On rappelle que pour un ensemble X d'opérateurs bornés sur un espace de Hilbert H , $X \subset B(H)$, on appelle *commutant* de X et on note X' l'ensemble de tous les opérateurs $T \in B(H)$ qui commutent à X . Si $A \subset B(H)$ est une sous-algèbre involutive qui agit d'une façon non-dégénérée sur H , le *bicommutant* A'' coïncide avec l'adhérence de A dans $B(H)$ pour la topologie faible, c'est-à-dire la topologie donnée par les semi-normes $T \mapsto |\langle T\xi, \eta \rangle|$, où $\xi, \eta \in H$.

On appelle *algèbre de von Neumann* toute sous-algèbre involutive $M \subset B(H)$ qui est égale à son bicommutant : $M = M''$, ce qui équivaut à dire que M est faiblement fermé et $1 \in M$. Un *facteur* est une algèbre de von Neumann dont le centre est réduit aux scalaires. Si G est un groupe localement compact, l'algèbre de von Neumann du groupe G notée $L(G)$ est le bicommutant $\{\lambda_g \mid g \in G\}''$ où (λ_g) est la représentation régulière du groupe G sur l'espace de Hilbert $L^2(G)$.

Murray et von Neumann ont classifié les facteurs en types I, II et III. Les *facteurs de type I* sont ceux qui possèdent des projecteurs minimaux. Ils sont isomorphes à $M_n(\mathbb{C})$ (type I_n) ou $B(\ell^2)$ (type I_∞). Les *facteurs de type II_1* sont ceux qui admettent une trace finie et qui sont de dimension infinie (pour exclure le cas I_n). L'exemple type d'un facteur II_1 est donné par l'algèbre de von Neumann $L(G)$ d'un groupe discret G dont $\{e\}$ est la seule classe de conjugaison finie (on dit que G est CCI). Les groupes libres \mathbb{F}_n à n générateurs sont des exemples de groupes CCI (n peut être ∞). Les *facteurs de type II_∞* sont ceux qui sont de la forme $N \otimes B(\ell^2)$ avec N un facteur II_1 . Ce sont exactement les facteurs qui admettent une trace infinie semi-finie et qui ne sont pas de type I. Un facteur de type I ou II est dit *semi-fini*. Il admet toujours une trace semi-finie. Finalement les *facteurs de type III* sont ceux qui n'admettent pas de trace non-nulle.

Remarque 1.1. — Dans tout l'exposé les espaces de Hilbert sont supposés *séparables* et les algèbres de von Neumann à *préduale séparable*, i.e. admettant une représentation fidèle sur un espace de Hilbert séparable.

1.2. Classification des facteurs de type III d'après Connes

Un *état normal* d'une algèbre de von Neumann M est une forme linéaire faiblement continue $\omega : M \rightarrow \mathbb{C}$ positive ($\omega(x) \geq 0$ quand $x \geq 0$) qui satisfait $\omega(1) = 1$. Un état est dit *fidèle* si $x = 0$ dès que $\omega(x) = 0$ et $x \geq 0$. Toute algèbre de von Neumann à préduale séparable admet un état fidèle.

La théorie de Tomita-Takesaki associe à tout état fidèle ω un groupe à un paramètre (σ_t^ω) d'automorphismes de M appelé *groupe modulaire* et caractérisé par

$$- \omega \sigma_t^\omega = \omega \text{ pour tout } t \in \mathbb{R},$$

- ω satisfait la condition KMS par rapport à (σ_t^ω) : pour tout $x, y \in M$, il existe une fonction continue $f : \{z \in \mathbb{C} \mid 0 \leq \text{Im } z \leq 1\} \rightarrow \mathbb{C}$ qui est analytique à l'intérieur de la bande et qui satisfait

$$f(t) = \omega(x\sigma_t^\omega(y)) \quad \text{et} \quad f(t+i) = \omega(\sigma_t^\omega(y)x) \quad \text{pour tout } t \in \mathbb{R} .$$

Une bonne introduction à la *théorie modulaire de Tomita-Takesaki* se trouve dans [27].

Le *théorème de Radon-Nikodym* de Connes [7] permet de comparer les groupes modulaires de deux états fidèles ω, μ sur M . En effet, il existe une application faiblement continue $t \mapsto u_t$ de \mathbb{R} dans le groupe unitaire de M telle que

- $u_{t+s} = u_t \sigma_t^\omega(u_s)$,
- $\sigma_t^\mu(x) = u_t \sigma_t^\omega(x) u_t^*$.

Les groupes modulaires de deux états fidèles diffèrent donc par une perturbation intérieure. Il s'en suit qu'une algèbre de von Neumann a une dynamique intrinsèque donnée par les groupes modulaires des états fidèles et déterminée à perturbation intérieure près.

Définissons le groupe polonais $\text{Aut } M$ des automorphismes de M muni de la topologie induite par les distances $d(\alpha, \beta) = \|\omega\alpha - \omega\beta\|$ et $d(\alpha, \beta) = \|\omega\alpha^{-1} - \omega\beta^{-1}\|$, où ω parcourt les états de M . À chaque unitaire $u \in M$, on associe l'*automorphisme intérieur* $\text{Ad } u$ défini par $(\text{Ad } u)(x) = uxu^*$. Les automorphismes intérieurs forment un sous-groupe distingué $\text{Int } M$ de $\text{Aut } M$. Le groupe quotient est noté $\text{Out } M$. Le théorème de Radon-Nikodym permet de définir un homomorphisme $\delta : \mathbb{R} \rightarrow \text{Out } M$ qui envoie t à la classe de σ_t^ω et qui ne dépend pas du choix de ω .

Invariant T. — Soit ω un état fidèle sur un facteur M avec groupe modulaire (σ_t) . Dans [7] Connes a introduit le sous-groupe $T(M)$ de \mathbb{R} :

$$T(M) = \{t \in \mathbb{R} \mid \sigma_t^\omega \in \text{Int } M\} .$$

D'après le théorème de Radon-Nikodym, l'invariant $T(M)$ ne dépend pas du choix de l'état ω . C'est donc un invariant de l'algèbre de von Neumann M .

Flot des poids. — À chaque facteur M est associée une algèbre de von Neumann semi-finie : c'est le produit croisé $M \rtimes_{(\sigma_t)} \mathbb{R}$ de M par le groupe modulaire (σ_t) d'un état fidèle sur M . Sur le produit croisé $M \rtimes_{(\sigma_t)} \mathbb{R}$, il existe une trace semi-finie canonique. Le produit croisé $M \rtimes_{(\sigma_t)} \mathbb{R}$ admet une *action duale* (θ_s) de \mathbb{R}_+^* par automorphismes. La restriction de l'action (θ_s) au centre du produit croisé s'appelle le *flot des poids* de M . Grâce au théorème de Radon-Nikodym, le flot des poids ne dépend pas du choix de l'état fidèle.

Facteurs de type III $_\lambda$. — Notons que le flot des poids est une action ergodique de \mathbb{R}_+^* sur un espace mesuré. Or une telle action est ou bien transitive ou bien proprement ergodique, d'où la classification suivante :

- M est semi-fini si c'est l'action de \mathbb{R}_+^* sur \mathbb{R}_+^* ,
- M est de type III $_\lambda$ avec $0 < \lambda < 1$ si c'est l'action de \mathbb{R}_+^* sur $\mathbb{R}_+^*/\lambda^{\mathbb{Z}}$,

- M est de type III_1 si c'est l'action de \mathbb{R}_+^* sur un point,
- M est de type III_0 si c'est une action proprement ergodique.

Combes a donné un aperçu des résultats de classification de Connes dans [4].

Murray et von Neumann ont construit deux facteurs de type II_1 non-isomorphes : le facteur hyperfini \mathcal{R} et le facteur $L(\mathbb{F}_2)$ du groupe libre à 2 générateurs. Ils sont non-isomorphes car $\text{Int } L(\mathbb{F}_2)$ est fermé dans $\text{Aut } L(\mathbb{F}_2)$ tandis que $\text{Int } \mathcal{R}$ est un sous-groupe dense non-trivial de $\text{Aut } \mathcal{R}$.

Facteurs pleins et invariant τ . — Un facteur M est dit *plein* si $\text{Int } M$ est un sous-groupe fermé de $\text{Aut } M$. Pour un facteur plein, le groupe quotient $\text{Out } M$ est un groupe polonais. Pour un tel facteur plein Connes [6] introduit un nouvel invariant :

$\tau(M)$ = la topologie la plus faible sur \mathbb{R} qui rend continue l'application $\delta : \mathbb{R} \rightarrow \text{Out } M$.

Les facteurs d'Araki-Woods libres sont des facteurs pleins. Nous remarquons qu'un facteur plein ne peut être de type III_0 .

1.3. Probabilités libres d'après Voiculescu

Une introduction plus complète aux probabilités libres de Voiculescu se trouve dans le livre [31] ou dans [26].

Un *espace de probabilités non-commutatif* est une paire (A, φ) où A est une algèbre unifère et φ est une forme linéaire vérifiant $\varphi(1) = 1$. Dans cet exposé on s'intéresse surtout aux algèbres de von Neumann : A est une algèbre de von Neumann et φ un état normal. Les éléments de A s'appellent toujours variables aléatoires. La *distribution* d'un élément $x \in A$ est l'application qui à un polynôme $P \in \mathbb{C}[X]$ associe $\varphi(P(x))$. Si A est une algèbre de von Neumann et x un élément auto-adjoint, la distribution de x est une mesure de probabilités dont le support est contenu dans $[-\|x\|, \|x\|]$.

NOTATION 1.2. — Si (M, φ) et (N, μ) sont des algèbres de von Neumann munies d'états φ et μ , la notation $(M, \varphi) \cong (N, \mu)$ signifie qu'il existe un **-isomorphisme* $\alpha : M \rightarrow N$ tel que $\mu\alpha = \varphi$.

DÉFINITION 1.3. — Soit (A, φ) un espace de probabilités non-commutatif. Une famille $(A_i)_{i \in I}$ de sous-algèbres est dite *libre* si pour tout k et toute suite d'éléments $a_j \in A_{i_j}$ ($j = 1, \dots, k$) satisfaisant $\varphi(a_j) = 0$ et $i_j \neq i_{j+1}$, on a $\varphi(a_1 \cdots a_k) = 0$.

Une famille d'éléments $(x_i)_{i \in I}$ est dite *libre* (resp. **-libre*) si les algèbres (resp. **-algèbres*) A_i engendrées par x_i forment une famille libre de sous-algèbres de A .

Les produits libres fournissent des exemples de familles libres.

PROPOSITION 1.4. — Soit (M_i, φ_i) une famille d'algèbres de von Neumann munies d'un état fidèle. Alors, il existe, à isomorphisme près, une unique paire (M, φ) d'une algèbre von Neumann munie d'un état fidèle, telle que

- (M_i, φ_i) se plonge dans (M, φ) en préservant l'état,

– M est engendrée par la famille de sous-algèbres (M_i) qui est une famille libre dans (M, φ) .

On appelle (M, φ) le produit libre des (M_i, φ_i) et on note $(M, \varphi) = \ast_{i \in I} (M_i, \varphi_i)$.

1.4. Représentations orthogonales de \mathbb{R}

La donnée de la construction des états quasi-libres libres et des facteurs d'Araki-Woods libres associés est une représentation de \mathbb{R} par transformations orthogonales.

TERMINOLOGIE 1.5. — On appelle représentation orthogonale de \mathbb{R} tout groupe à un paramètre de transformations orthogonales d'un espace de Hilbert réel $H_{\mathbb{R}}$.

Soit (U_t) une représentation orthogonale de \mathbb{R} sur l'espace de Hilbert réel $H_{\mathbb{R}}$. Le complexifié $H = H_{\mathbb{R}} \otimes \mathbb{C}$ admet une involution anti-unitaire J (l'opérateur de conjugaison complexe) et les transformations orthogonales (U_t) s'étendent en un groupe à un paramètre d'unitaires sur H , qu'on notera toujours (U_t) .

Il existe alors un unique opérateur auto-adjoint strictement positif A sur H tel que $U_t = A^{it}$ pour tout $t \in \mathbb{R}$. On a $JAJ = A^{-1}$. L'opérateur A permet de définir un nouveau plongement isométrique

$$H_{\mathbb{R}} \hookrightarrow H : \xi \mapsto \left(\frac{2}{A^{-1} + 1} \right)^{1/2} \xi .$$

En effet, si $\xi \in H_{\mathbb{R}}$, on a $J\xi = \xi$ et donc

$$\begin{aligned} \left\| \left(\frac{2}{A^{-1} + 1} \right)^{1/2} \xi \right\|^2 &= \left\langle \frac{1}{A^{-1} + 1} \xi, \xi \right\rangle + \left\langle \frac{1}{A^{-1} + 1} J\xi, J\xi \right\rangle = \left\langle \frac{A}{A + 1} \xi, \xi \right\rangle + \left\langle J \frac{1}{A + 1} \xi, J\xi \right\rangle \\ &= \left\langle \frac{A + 1}{A + 1} \xi, \xi \right\rangle = \|\xi\|^2 . \end{aligned}$$

On notera $K_{\mathbb{R}}$ l'image de $H_{\mathbb{R}}$ par ce plongement. Alors, $K_{\mathbb{R}}$ est un espace de Hilbert réel, isométriquement plongé dans un espace de Hilbert complexe H vérifiant la propriété suivante :

$$(\star) \quad K_{\mathbb{R}} \cap iK_{\mathbb{R}} = \{0\} \text{ et } K_{\mathbb{R}} + iK_{\mathbb{R}} \text{ est dense dans } H .$$

Dans [18] on démontre que chaque plongement isométrique $K_{\mathbb{R}} \subset H$ satisfaisant la condition (\star) provient d'une représentation orthogonale de \mathbb{R} sur un espace de Hilbert réel $H_{\mathbb{R}}$ par la construction présentée ci-dessus.

Écrivons $T = JA^{-1/2}$. Alors T est un opérateur anti-linéaire fermé et inversible sur H qui satisfait $T = T^{-1}$. Un tel opérateur s'appelle une *involution sur H* . Réciproquement une telle involution T admet une décomposition polaire $T = JA^{-1/2}$ dans laquelle J est une involution anti-unitaire sur H et A est un opérateur auto-adjoint strictement positif satisfaisant $JAJ = A^{-1}$. Posons $H_{\mathbb{R}} = \{\xi \in H \mid J\xi = \xi\}$ et $U_t = A^{it}$. On obtient ainsi une représentation orthogonale de \mathbb{R} . On remarquera que l'espace $K_{\mathbb{R}}$ correspondant consiste en les vecteurs ξ dans le domaine de T qui satisfont $T\xi = \xi$.

On a alors obtenu plusieurs points de vue différents sur les représentations orthogonales de \mathbb{R} .

1. Un groupe à un paramètre de transformations orthogonales d'un espace de Hilbert réel.
2. Un plongement isométrique d'un espace de Hilbert réel dans un espace de Hilbert complexe vérifiant (\star) .
3. Une involution T sur un espace de Hilbert.

Finalement on peut considérer la décomposition spectrale de l'opérateur $\log A$. Comme $J(\log A)J = -\log A$, la classe de la mesure spectrale de $\log A$ est symétrique. Les représentations orthogonales de \mathbb{R} sont donc classifiées par une classe de mesures symétrique sur \mathbb{R} et une fonction de multiplicité symétrique.

2. FACTEURS D'ARAKI-WOODS LIBRES

Le foncteur CAR associe à tout espace de Hilbert H la C^* -algèbre $\text{CAR}(H)$ (voir introduction). Oubliant la structure complexe de H on peut écrire $\text{CAR}(H)$ comme une algèbre de Clifford. Soit $H_{\mathbb{R}}$ un espace de Hilbert réel. On note $\text{Cliff}(H_{\mathbb{R}})$ et on appelle *algèbre de Clifford* la C^* -algèbre universelle engendrée par la famille $\{s(\xi) \mid \xi \in H_{\mathbb{R}}\}$ telle que $s(\xi)$ est auto-adjoint pour tout $\xi \in H_{\mathbb{R}}$, $\xi \mapsto s(\xi)$ est \mathbb{R} -linéaire et

$$s(\xi)s(\eta) + s(\eta)s(\xi) = 2\langle \xi, \eta \rangle 1 .$$

Cette dernière condition étant équivalente à $s(\xi)^2 = \|\xi\|^2 1$ pour tout $\xi \in H_{\mathbb{R}}$, on voit comment le foncteur $H_{\mathbb{R}} \mapsto \Gamma(H_{\mathbb{R}})$ est une version libre du foncteur Cliff .

À chaque plongement isométrique $H_{\mathbb{R}} \hookrightarrow H$ de $H_{\mathbb{R}}$ dans un espace de Hilbert complexe H est associée une représentation de $\text{Cliff}(H_{\mathbb{R}})$ sur *l'espace de Fock anti-symétrique (ou fermionique)* :

$$\mathcal{F}_{\text{as}}(H) = \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H^{\wedge n}$$

posant $s(\xi) = a(\xi)^* + a(\xi)$ où $a(\xi)$ est l'opérateur de création à gauche. Remarquons que cette représentation de $\text{Cliff}(H_{\mathbb{R}})$ est en fait la représentation GNS d'un état quasi-libre. L'algèbre de von Neumann engendrée par les opérateurs $s(\xi), \xi \in H_{\mathbb{R}}$ est un *facteur d'Araki-Woods*.

Shlyakhtenko donne une version libre de la construction précédente et appelle le facteur engendré *facteur d'Araki-Woods libre*.

2.1. États quasi-libres libres

Donnons-nous une représentation orthogonale (U_t) de \mathbb{R} sur l'espace de Hilbert réel $H_{\mathbb{R}}$. Comme au §1.4, nous regardons le complexifié H de $H_{\mathbb{R}}$ avec l'involution anti-unitaire J et l'opérateur auto-adjoint strictement positif A tel que $U_t = A^{it}$. Introduisons *l'espace de Fock plein* de H :

$$\mathcal{F}(H) = \mathbb{C}\Omega \oplus \bigoplus_{n=1}^{\infty} H^{\otimes n} .$$

Le vecteur unité Ω s'appelle *vecteur du vide*. Pour chaque vecteur $\xi \in H$, nous disposons de l'opérateur de création à gauche

$$\ell(\xi) : \mathcal{F}(H) \rightarrow \mathcal{F}(H) : \begin{cases} \ell(\xi)\Omega = \xi, \\ \ell(\xi)(\xi_1 \otimes \cdots \otimes \xi_n) = \xi \otimes \xi_1 \otimes \cdots \otimes \xi_n. \end{cases}$$

L'adjoint $\ell(\xi)^*$ s'appelle *opérateur d'annihilation*.

Pour chaque vecteur $\xi \in H$, notons $s(\xi)$ la partie réelle de $\ell(\xi)$ donnée par

$$s(\xi) = \frac{\ell(\xi) + \ell(\xi)^*}{2}.$$

Un résultat crucial de Voiculescu [31] dit que la distribution de l'opérateur $s(\xi)$ par rapport à l'état vectoriel du vide donné par $\varphi(x) = \langle x\Omega, \Omega \rangle$ est la loi semi-circulaire de Wigner supportée par l'intervalle $[-\|\xi\|, \|\xi\|]$.

Rappelons que l'opérateur A permet de définir un plongement de $H_{\mathbb{R}}$ dans H dont l'image est notée $K_{\mathbb{R}}$. On peut alors formuler la définition centrale de cet exposé [25].

DÉFINITION 2.1. — Soit (U_t) une représentation orthogonale de \mathbb{R} sur l'espace de Hilbert réel $H_{\mathbb{R}}$. Le facteur d'Araki-Woods libre⁽²⁾ noté $\Gamma(H_{\mathbb{R}}, U_t)''$ est défini par

$$\Gamma(H_{\mathbb{R}}, U_t)'' = \{s(\xi) \mid \xi \in K_{\mathbb{R}}\}''.$$

L'état vectoriel $\varphi_U(x) = \langle x\Omega, \Omega \rangle$ est appelé état quasi-libre libre.

Rappelons que $T = JA^{-1/2}$ est l'involution sur H associée à (U_t) . Pour $\xi, \eta \in K_{\mathbb{R}}$, on vérifie que

$$2s(\xi) + 2is(\eta) = \ell(\zeta) + \ell(T\zeta)^*$$

où $\zeta = \xi + i\eta$. On conclut que $\Gamma(H_{\mathbb{R}}, U_t)''$ est également l'algèbre de von Neumann engendrée par les opérateurs $\ell(\zeta) + \ell(T\zeta)^*$ où ζ appartient au domaine de T .

Le résultat suivant est facile à démontrer.

PROPOSITION 2.2. — L'état quasi-libre libre φ_U sur $\Gamma(H_{\mathbb{R}}, U_t)''$ est fidèle. Le groupe modulaire (σ_t) de l'état φ_U est donné par

$$\sigma_t(s(\xi)) = s(U_t\xi) \quad \text{pour tout } t \in \mathbb{R}, \xi \in K_{\mathbb{R}}.$$

La construction des facteurs d'Araki-Woods libres est fonctorielle dans un sens précis. En effet on considère la catégorie dont les objets sont les paires $(H_{\mathbb{R}}, U_t)$ et les morphismes sont les contractions entre espaces de Hilbert qui entrelacent les représentations. À chaque morphisme $(H_{\mathbb{R}}^{(1)}, U_t^{(1)}) \rightarrow (H_{\mathbb{R}}^{(2)}, U_t^{(2)})$ correspond une application complètement positive $\Gamma(H_{\mathbb{R}}^{(1)}, U_t^{(1)})'' \rightarrow \Gamma(H_{\mathbb{R}}^{(2)}, U_t^{(2)})''$ normale et unifère, préservant les états quasi-libres libres. On notera Γ'' ce foncteur.

La catégorie des paires $(H_{\mathbb{R}}, U_t)$ admet une structure additive : la somme directe. Shlyakhtenko démontre que le foncteur Γ'' entrelace les opérations somme directe et produit libre.

⁽²⁾Nous verrons que $\Gamma(H_{\mathbb{R}}, U_t)''$ est effectivement un facteur dès que $\dim H_{\mathbb{R}} \geq 2$.

PROPOSITION 2.3. — Soit $(H_{\mathbb{R}}^{(i)}, U_t^{(i)})_{i \in I}$ une représentation orthogonale de \mathbb{R} . Posons $(H_{\mathbb{R}}, U_t) = \bigoplus_i (H_{\mathbb{R}}^{(i)}, U_t^{(i)})$. Alors,

$$(\Gamma(H_{\mathbb{R}}, U_t)'' , \varphi_U) \cong \ast_{i \in I} (\Gamma(H_{\mathbb{R}}^{(i)}, U_t^{(i)})'' , \varphi_{U^{(i)}}) .$$

2.2. Variables circulaires généralisées

Pour comprendre la structure des algèbres de von Neumann $(\Gamma(H_{\mathbb{R}}, U_t)'' , \varphi_U)$ il est naturel de considérer d’abord les représentations orthogonales irréductibles de \mathbb{R} . Le cas $H_{\mathbb{R}} = \mathbb{R}$ et $U_t = \text{id}$ est facile : l’algèbre est engendré par un seul opérateur dont la distribution par rapport à φ_U est la loi semi-circulaire, d’après le résultat de Voiculescu. On trouve donc

$$(\Gamma(\mathbb{R}, \text{id})'' , \varphi_U) \cong (L^\infty[-1, 1], \mu)$$

où μ est la mesure semi-circulaire sur $[-1, 1]$. Si on combine ce résultat avec la proposition 2.3, on obtient

$$(1) \quad (\Gamma(H_{\mathbb{R}}, \text{id})'' , \varphi_U) \cong (L(\mathbb{F}_n), \text{tr})$$

où $L(\mathbb{F}_n)$ est l’algèbre de von Neumann du groupe libre à $n = \dim H_{\mathbb{R}}$ générateurs.

Prenons maintenant $H_{\mathbb{R}} = \mathbb{R}^2$ et $0 < \lambda < 1$. Posons

$$(2) \quad U_t = \begin{pmatrix} \cos(t \log \lambda) & -\sin(t \log \lambda) \\ \sin(t \log \lambda) & \cos(t \log \lambda) \end{pmatrix} .$$

En prenant la base orthonormale $\xi_1 = \frac{1}{\sqrt{2}}(1, -i)$, $\xi_2 = \frac{1}{\sqrt{2}}(1, i)$ du complexifié $H = \mathbb{C}^2$, on voit que l’algèbre de von Neumann $\Gamma(H_{\mathbb{R}}, U_t)''$ est engendrée par l’opérateur $\ell(\xi_2) + \sqrt{\lambda}\ell(\xi_1)^*$ sur l’espace de Fock plein $\mathcal{F}(\mathbb{C}^2)$.

NOTATION 2.4. — On notera $(T_\lambda, \varphi_\lambda) := \Gamma(H_{\mathbb{R}}, U_t)''$ où $H_{\mathbb{R}} = \mathbb{R}^2$ et U_t est donné par l’égalité (2).

Pour comprendre l’algèbre T_λ , il faut étudier la \ast -distribution de l’élément $\ell(\xi_2) + \sqrt{\lambda}\ell(\xi_1)^*$ par rapport à l’état vectoriel du vide. Un tel élément s’appelle *élément circulaire généralisé*. Dans le cas $\lambda = 1$, on retrouve l’élément circulaire y de Voiculescu [31]. Voiculescu a démontré que la décomposition polaire $y = ub$ d’un élément circulaire donne un *unitaire de Haar* u et un opérateur b *quart-circulaire*. Ceci veut dire que la distribution de u est la distribution uniforme sur le cercle et que la distribution de b suit la loi quart-circulaire supportée par l’intervalle $[0, 1]$. Shlyakthenko a démontré dans [25] un résultat analogue pour les éléments circulaires généralisés. Ce résultat permet de donner une description alternative de $(T_\lambda, \varphi_\lambda)$.

THÉORÈME 2.5. — Soit $0 < \lambda < 1$ et soit $y = \ell(\xi_1) + \sqrt{\lambda}\ell(\xi_2)^*$ l’élément circulaire généralisé associé dans $(T_\lambda, \varphi_\lambda)$. Notons $y = vb$ la décomposition polaire de y . Alors v est une isométrie non-unitaire qui satisfait $\varphi_\lambda(v^k(v^*)^l) = \delta_{kl}\lambda^k$. La distribution de l’opérateur b est sans atomes. Les éléments u et b sont \ast -libres.

Un corollaire immédiat de ce résultat est que

$$(3) \quad (T_\lambda, \varphi_\lambda) \cong (B(\ell^2(\mathbb{N})), \omega_\lambda) * (L^\infty[-1, 1], \mu)$$

où $\omega_\lambda(e_{ij}) = \delta_{ij}\lambda^j(1 - \lambda)$ et μ est la loi semi-circulaire sur $[-1, 1]$. Bien évidemment, au lieu de μ on pourrait prendre n'importe quelle autre mesure de probabilités sans atomes.

L'isomorphisme (3) est crucial. Il permet de réaliser $(T_\lambda, \varphi_\lambda)$ en représentant d'une manière libre $(B(\ell^2(\mathbb{N})), \omega_\lambda)$ et $(L^\infty[-1, 1], \mu)$ dans un espace de probabilité non-commutatif. Shlyakhtenko trouve dans [25] de telles représentations qui permettent de comprendre la réduction de l'algèbre T_λ par un projecteur minimal de $B(\ell^2(\mathbb{N}))$. On les appelle *modèles matriciels*. C'est un outil puissant qui permet de démontrer des résultats *d'absorption libre*.

THÉORÈME 2.6. — *On a*

$$(T_\lambda, \varphi_\lambda) \cong (T_\lambda, \varphi_\lambda) * (L^\infty[-1, 1], \mu) \cong (T_\lambda, \varphi_\lambda) * (L(\mathbb{F}_\infty), \text{tr}),$$

où μ est la mesure semi-circulaire et tr est la trace sur le facteur $L(\mathbb{F}_\infty)$ du groupe libre à une infinité de générateurs.

2.3. Type des facteurs d'Araki-Woods libres

À l'aide du théorème 2.6, on peut finalement démontrer que $\Gamma(H_{\mathbb{R}}, U_t)''$ est toujours un facteur quand la dimension de $H_{\mathbb{R}}$ est au moins 2. On peut en même temps déterminer le type de ce facteur et son invariant τ .

THÉORÈME 2.7. — *Soit (U_t) une représentation orthogonale de \mathbb{R} sur l'espace de Hilbert réel $H_{\mathbb{R}}$ de dimension au moins 2. Notons $M = \Gamma(H_{\mathbb{R}}, U_t)''$.*

1. *M est un facteur plein.*
2. *M est de type II_1 ssi $U_t = \text{id}$ pour tout $t \in \mathbb{R}$.*
3. *M est de type III_λ ($0 < \lambda < 1$) ssi (U_t) est périodique de période $\frac{2\pi}{|\log \lambda|}$.*
4. *M est de type III_1 dans les autres cas.*
5. *L'invariant $\tau(M)$ est la topologie la plus faible sur \mathbb{R} qui rend continue l'application $t \mapsto U_t$ de \mathbb{R} dans le groupe orthogonal de $H_{\mathbb{R}}$ muni de la topologie faible.*
6. *Le facteur M admet des états presque-périodiques ssi (U_t) est presque-périodique.*

Nous donnons ici plus de détails pour la démonstration de ce théorème. Shlyakhtenko détermine l'invariant $\tau(M)$ dans [23], mais en supposant que la représentation orthogonale (U_t) contient ou bien une représentation périodique ou bien une représentation triviale de dimension 2. Nous suivons la même méthode que Shlyakhtenko, mais utilisons le nouveau lemme 4.1 qui est plus fort que le lemme des 14ε de Barnett [3] utilisé par Shlyakhtenko. Shlyakhtenko démontre (1)–(4) dans [25] pour les représentations presque-périodiques et dans [24] pour le cas général, mais par d'autres méthodes que nous.

Preuve du théorème 2.7. — Il suffit de démontrer (1) et (5). En effet, un facteur plein est semi-fini (c'est-à-dire, de type I ou II) ssi $\tau(M)$ est la topologie grossière. Dans ce cas-là, on conclut de (5) que $U_t = \text{id}$ pour tout $t \in \mathbb{R}$ et d'après l'isomorphisme (1), M est un facteur II_1 . Ceci démontre (2). Un facteur plein n'est jamais de type III_0 . Comme $\tau(M)$ est la topologie la plus faible qui rend continue l'application $\delta : \mathbb{R} \rightarrow \text{Out}(M)$, on conclut de (5) que $\delta(t) = 1$ ssi $U_t = \text{id}$. Ceci démontre (3) et (4). Finalement, démontrons (6). Si M admet un état presque-périodique, le groupe \mathbb{R} muni de la topologie $\tau(M)$ peut être complété en un groupe compact. Il existe donc un groupe compact G , un plongement $\mathbb{R} \subset G$ et une extension de $t \mapsto U_t$ en un homomorphisme continu $G \rightarrow O(H_{\mathbb{R}})$. Ceci veut dire que (U_t) est presque-périodique [8]. Réciproquement, si (U_t) est presque-périodique, l'état quasi-libre libre est un état presque-périodique.

Il nous reste à démontrer (1) et (5). Ceci est évident quand $U_t = \text{id}$ pour tout $t \in \mathbb{R}$. Le deuxième cas qu'on considère est celui où (U_t) contient la représentation donnée par l'égalité (2) avec $0 < \lambda < 1$. Notons son complément par (U'_t) agissant sur $H'_{\mathbb{R}}$. D'après le théorème 2.6 on a

$$\begin{aligned} (M, \varphi) &\cong (T_{\lambda}, \varphi_{\lambda}) * (\Gamma(H'_{\mathbb{R}}, U'_t)'', \varphi_{U'}) \\ &\cong ((T_{\lambda}, \varphi_{\lambda}) * (L^{\infty}[-1, 1], \mu)) * (L^{\infty}([-1, 1], \mu) * (\Gamma(H'_{\mathbb{R}}, U'_t)'', \varphi_{U'})) . \end{aligned}$$

Comme $(L^{\infty}[-1, 1], \mu)$ contient un unitaire de Haar, on peut appliquer la proposition 4.2. On conclut que M est un facteur plein et que l'invariant $\tau(M)$ est la topologie la plus faible sur \mathbb{R} qui rend continues les deux applications $t \mapsto \sigma_t^{\varphi_{\lambda}}$ et $t \mapsto \sigma_t^{\varphi_{U'}}$. Par la proposition 2.2 ceci est exactement la topologie la plus faible sur \mathbb{R} qui rend continue l'application $t \mapsto U_t$.

Finalement, nous considérons le cas où la représentation (U_t) ne contient pas de représentation périodique et n'est pas triviale. Il est alors clair qu'on peut décomposer U_t en trois composantes non-triviales $U_t = U_t^{(1)} \oplus U_t^{(2)} \oplus U_t^{(3)}$. Les énoncés (1) et (5) découlent des propositions 2.3 et 4.2 ainsi que du lemme 4.3. \square

Pour chaque représentation orthogonale non-périodique (U_t) de \mathbb{R} , le facteur d'Araki-Woods libre $\Gamma(H_{\mathbb{R}}, U_t)''$ est donc un facteur de type III_1 dont l'invariant τ est la topologie la plus faible qui rend continue l'application $t \mapsto U_t$.

Remarque 2.8. — Dans [6] Connes part d'une mesure finie μ sur \mathbb{R}_+^* telle que $\int \lambda d\mu(\lambda) < \infty$. On y associe la représentation unitaire (U_t) de \mathbb{R} sur $L^2(\mathbb{R}_+^*, \mu)$, défini par $(U_t \xi)(\lambda) = \lambda^{it} \xi(\lambda)$. On suppose que (U_t) est non-périodique.

Connes définit $P = M_2(L^{\infty}(\mathbb{R}_+^*, \mu))$ muni de l'état φ proportionnel à la forme positive

$$\omega \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} = \int f_{11}(\lambda) d\mu(\lambda) + \int \lambda f_{22}(\lambda) d\mu(\lambda) .$$

Prenons un groupe discret infini G et définissons le produit tensoriel infini

$$P_{\infty} = \bigotimes_{g \in G} (P, \varphi) .$$

Alors G agit sur P_∞ par automorphismes de décalage des facteurs tensoriels et on considère le produit croisé $M = P_\infty \rtimes G$. De cette manière M est un facteur de type III_1 . Connes démontre que pour $G = \mathbb{F}_n$ ($n = 2, \dots, +\infty$), M est un facteur plein et l'invariant $\tau(M)$ est la topologie la plus faible qui rend continue l'application $t \mapsto U_t$. Les facteurs de type III_1 de Connes et ceux de Shlyakhtenko peuvent-ils être isomorphes ?

3. CLASSIFICATION DES FACTEURS D'ARAKI-WOODS LIBRES

3.1. Le cas presque-périodique

Supposons d'abord que (U_t) est une représentation orthogonale *presque-périodique*. Ceci veut dire que l'opérateur A , qui était défini sur le complexifié H de $H_\mathbb{R}$ par $U_t = A^{it}$, a un spectre purement ponctuel. Soit $G \subset \mathbb{R}_+^*$ le sous-groupe engendré par le spectre ponctuel de A . Shlyakhtenko [25] démontre que ce sous-groupe *classifie les facteurs d'Araki-Woods libres presque-périodiques*.

THÉORÈME 3.1. — *Soit (U_t) une représentation orthogonale presque-périodique et non-triviale. Soit G le sous-groupe de \mathbb{R}_+^* engendré par le spectre ponctuel de A . Alors, $(\Gamma(H_\mathbb{R}, U_t)'' , \varphi_U)$ ne dépend que de G à des isomorphismes qui préservent l'état quasi-libre libre près.*

Réciproquement, le groupe G coïncide avec l'invariant S_{discret} du facteur $\Gamma(H_\mathbb{R}, U_t)''$ [6], qui classifie donc les facteurs d'Araki-Woods libres presque-périodiques et non-triviales.

En particulier, il découle de ce théorème et du théorème 2.7 que $(T_\lambda, \varphi_\lambda)$ est le seul facteur d'Araki-Woods libre de type III_λ ($0 < \lambda < 1$).

Remarquons que le cas où $U_t = \text{id}$ pour tout $t \in \mathbb{R}$ reste ouvert. En effet, d'après l'isomorphisme (1), on sait qu'on obtient le facteur du groupe libre à n générateurs : décider si ces facteurs dépendent de n est un des problèmes ouverts en algèbres d'opérateurs.

3.2. Le facteur de type II_∞ associé

Au §1 nous avons vu qu'on associe, à chaque algèbre de von Neumann M , une algèbre de von Neumann semi-finie $M \rtimes_{(\sigma_t)} \mathbb{R}$ où (σ_t) est le groupe modulaire d'un état fidèle sur M . On sait que M est un facteur de type III_1 ssi le produit croisé $M \rtimes_{(\sigma_t)} \mathbb{R}$ est un facteur de type II_∞ . On l'appelle le facteur II_∞ associé au facteur M de type III_1 .

Si M est un facteur de type III_λ ($0 < \lambda < 1$), on peut prendre un état fidèle sur M tel que le groupe modulaire correspondant (σ_t) admette $\frac{2\pi}{|\log \lambda|}$ comme période. Le groupe modulaire donne donc une action du cercle \mathbb{T} sur M . Le produit croisé $M \rtimes \mathbb{T}$ est un facteur de type II_∞ et on a un isomorphisme canonique

$$M \rtimes_{(\sigma_t)} \mathbb{R} \cong (M \rtimes \mathbb{T}) \otimes L^\infty(\mathbb{T}) .$$

Le facteur $M \rtimes \mathbb{T}$ de type II_∞ s'appelle également le facteur II_∞ associé à M .

Les facteurs II_∞ associés aux facteurs de type III_λ ($0 < \lambda \leq 1$) retiennent une certaine partie de la structure de M . Dans ce paragraphe on s'en sert pour démontrer certains résultats de *non-isomorphisme* entre les facteurs d'Araki-Woods libre.

Dans [25] Shlyakhtenko calcule le facteur II_∞ associé au facteur d'Araki-Woods libre $(T_\lambda, \varphi_\lambda)$ de type III_λ . Il identifie, grâce aux modèles matriciels, le facteur T_λ au facteur suivant étudié par Rădulescu [17].

$$D_\lambda := (M_2(\mathbb{C}), \omega_\lambda) * (L^\infty[-1, 1], \mu) ,$$

où $\omega_\lambda(e_{ij}) = \delta_{ij} \frac{\lambda^j}{1+\lambda}$ pour $i, j = 0, 1$, et μ est la mesure semi-circulaire. Dans [17] Rădulescu démontre que le facteur II_∞ associé à D_λ est isomorphe à $L(\mathbb{F}_\infty) \otimes B(\ell^2)$. On obtient donc le résultat suivant.

PROPOSITION 3.2. — *Le facteur II_∞ associé au facteur d'Araki-Woods libre $(T_\lambda, \varphi_\lambda)$ est isomorphe à $L(\mathbb{F}_\infty) \otimes B(\ell^2)$.*

Dans [23, 24] une description plus systématique des facteurs II_∞ associés aux facteurs d'Araki-Woods libres est donnée. L'idée est la suivante : dans l'isomorphisme (1) nous avons vu que le facteur d'un groupe libre est engendré par une famille libre d'opérateurs semi-circulaires. Une telle famille peut être obtenue par des opérateurs de création sur un espace de Fock plein.

Shlyakhtenko généralise ceci et considère dans [23] une famille libre d'opérateurs semi-circulaires à *coefficients dans une algèbre de von Neumann* A . On retrouve le cas précédent quand $A = \mathbb{C}$. On peut construire une telle famille à coefficients dans A en remplaçant, dans la construction de l'espace de Fock plein, les espaces de Hilbert par des A -modules hilbertiens. Ceci permet d'engendrer le facteur II_∞ associé à un facteur d'Araki-Woods libre par une famille libre à coefficients dans $A = L^\infty(\mathbb{R})$.

De cette manière Shlyakhtenko démontre dans [21, 24] le résultat suivant.

THÉORÈME 3.3. — *Soit $(H_\mathbb{R}, U_t)$ un multiple fini ou infini de la représentation régulière de \mathbb{R} donnée par $(L^2(\mathbb{R}, \mathbb{R}), \lambda_t)$. Alors, le facteur II_∞ associé à $\Gamma(H_\mathbb{R}, U_t)''$ est isomorphe à $L(\mathbb{F}_\infty) \otimes B(\ell^2)$.*

Dans [24] Shlyakhtenko identifie l'action duale sur le facteur $L(\mathbb{F}_\infty) \otimes B(\ell^2)$ de type II_∞ associé à $\Gamma(L^2(\mathbb{R}, \mathbb{R}), \lambda_t)''$ avec l'action construite par Rădulescu dans [16].

3.3. Des résultats de non-isomorphisme

Comme on a vu au §1.4, on peut associer à chaque mesure symétrique μ sur \mathbb{R} , une représentation orthogonale (U_t) de \mathbb{R} sur l'espace de Hilbert réel $H_\mathbb{R}$ défini par

$$H_\mathbb{R} = \{ \xi \in L^2(\mathbb{R}, \mu) \mid \xi(-x) = \overline{\xi(x)} \} \quad \text{et} \quad (U_t \xi)(x) = e^{itx} \xi(x).$$

On notera $\tau(\mu)$ la topologie la plus faible sur \mathbb{R} qui rend continue l'application $t \rightarrow U_t$ de \mathbb{R} dans $O(H_\mathbb{R})$ muni de la topologie faible. D'après le théorème 2.7, $\tau(\mu)$ est exactement l'invariant τ du facteur d'Araki-Woods libre $\Gamma(H_\mathbb{R}, U_t)''$. Dans [21] Shlyakhtenko

démontre qu'il existe une famille non-dénombrable de mesures μ sans atomes telles que les topologies $\tau(\mu)$ soient distinctes. On obtient le résultat suivant.

PROPOSITION 3.4. — *Il existe une famille non-dénombrable de facteurs d'Araki-Woods libres mutuellement non-isomorphes et sans états presque-périodiques.*

Les algèbres de cette famille peuvent être distinguées par l'invariant τ . Néanmoins, on verra plus tard que l'invariant τ ne suffit pas pour distinguer tous les facteurs d'Araki-Woods libres.

Voiculescu a introduit [30] la notion d'entropie libre $\chi(x_1, \dots, x_n)$ pour des éléments auto-adjoints x_1, \dots, x_n dans une algèbre de von Neumann finie M munie d'une trace. Ceci est utilisé pour définir la dimension entropique libre $\delta(x_1, \dots, x_n)$. Une application spectaculaire de l'entropie libre a été donnée par Voiculescu dans [29] où il démontre que les facteurs des groupes libres n'admettent pas de sous-algèbre de Cartan.

Dans [21, 23] Shlyakhtenko utilise la dimension entropique libre pour démontrer que, dans certains cas, le facteur II_∞ associé à un facteur d'Araki-Woods libre ne peut être isomorphe à $L(\mathbb{F}_\infty) \otimes B(\ell^2)$.

THÉORÈME 3.5. — *Soit (U_t) une représentation orthogonale non-périodique de \mathbb{R} sur un espace de Hilbert réel $H_\mathbb{R}$. Supposons que la mesure spectrale de $\bigoplus_{n \geq 1} U_t^{\otimes n}$ est singulière par rapport à la mesure de Lebesgue. Alors, le facteur II_∞ associé à $\Gamma(H_\mathbb{R}, U_t)''$ n'est pas isomorphe à $L(\mathbb{F}_\infty) \otimes B(\ell^2)$.*

En particulier, $\Gamma(H_\mathbb{R}, U_t)''$ n'est pas isomorphe à $\Gamma(L^2(\mathbb{R}, \mathbb{R}), \lambda_t)''$, où (λ_t) est la représentation régulière de \mathbb{R} .

La condition du théorème précédent est satisfaite si la topologie la plus faible qui rend continue l'application $t \mapsto U_t$ est strictement plus faible que la topologie usuelle de \mathbb{R} .

Dans [21] Shlyakhtenko construit une mesure μ sur \mathbb{R} telle que toutes les mesures $\mu * \dots * \mu$ sont singulières par rapport à la mesure de Lebesgue, mais néanmoins $\tau(\mu)$ est la topologie usuelle de \mathbb{R} . Le théorème précédent admet donc le corollaire suivant.

COROLLAIRE 3.6. — *Il existe des facteurs d'Araki-Woods libres non-isomorphes ayant le même invariant τ .*

Comme l'invariant τ ne distingue pas tous les facteurs d'Araki-Woods libres, Shlyakhtenko propose dans [21] un *nouvel invariant* \mathcal{S} pour les facteurs pleins de type III. Introduisons quelques notations. Si μ est une mesure sur \mathbb{R} , notons \mathcal{C}_μ l'ensemble de toutes les mesures qui sont absolument continues par rapport à la mesure μ . Dans le cas où μ est la mesure spectrale d'un opérateur auto-adjoint et strictement positif A , on pose $\mathcal{C}_A := \mathcal{C}_\mu$. Ceci permet de définir

$$\mathcal{S}(M) := \bigcap_{\varphi \text{ état fidèle sur } M} \mathcal{C}_{\bigoplus_n \Delta_\varphi^{\otimes n}},$$

où Δ_φ est l'opérateur modulaire de l'état φ . On remarque que les mesures dans $\mathcal{S}(M)$ sont supportées par \mathbb{R}_+^* et que la mesure de Dirac δ_1 est toujours dans $\mathcal{S}(M)$.

Shlyakhtenko démontre dans [21] que cet invariant \mathcal{S} distingue certains facteurs d'Araki-Woods libres (non-isomorphes) qui ont le même invariant τ .

3.4. Le facteur d'Araki-Woods libre dépend-il de la multiplicité ?

À chaque mesure symétrique μ sur \mathbb{R} est associée une représentation orthogonale (voir §3.3). Notons $\Gamma(\mu, n)$, où $n \in \{1, \dots, +\infty\}$, le facteur d'Araki-Woods libre associé à la somme directe de n copies de cette représentation.

Il découle du théorème de classification 3.1 que, dans le cas où μ est une mesure atomique non-concentrée en $\{0\}$, le facteur $\Gamma(\mu, n)$ ne dépend pas de n . D'après le théorème 2.7, l'invariant τ d'un facteur $\Gamma(\mu, n)$ quelconque ne dépend pas de n . Néanmoins, Shlyakhtenko démontre dans [20] un résultat très surprenant.

THÉORÈME 3.7. — *Soient λ la mesure de Lebesgue sur \mathbb{R} et δ_0 la mesure de Dirac en 0. Alors, $\Gamma(\lambda + \delta_0, 1)$ et $\Gamma(\lambda + \delta_0, 2)$ ne sont pas isomorphes.*

Dans sa preuve Shlyakhtenko utilise la notion d'algèbre de von Neumann solide due à Ozawa [12] : une algèbre de von Neumann est dite solide si le commutant relatif de n'importe quelle sous-algèbre diffuse et unifère est injectif. Rappelons qu'une algèbre de von Neumann est dite diffuse si elle n'admet pas de projecteurs minimaux. Une algèbre de von Neumann solide est nécessairement finie.

Ozawa démontre dans [12] que l'algèbre de von Neumann $L(G)$ d'un groupe discret hyperbolique G (voir [9]) est solide. En particulier, les facteurs des groupes libres sont solides.

Notons N_n le facteur II_∞ associé à $\Gamma(\lambda + \delta_0, n)$. Shlyakhtenko démontre que $N_1 \cong L(\mathbb{F}_\infty) \otimes B(\ell^2)$. Ceci implique que pN_1p est une algèbre de von Neumann solide pour tout projecteur fini $p \in N_1$. Par contre, il construit également un projecteur fini $q \in N_2$ tel que qN_2q ne soit pas solide.

Remarquons qu'il découle des résultats de [21] que l'invariant \mathcal{S} ne distingue pas $\Gamma(\lambda + \delta_0, 1)$ et $\Gamma(\lambda + \delta_0, 2)$.

Soit (U_t) une représentation orthogonale qui contient une représentation périodique non-triviale. D'après le théorème 2.6 (et la proposition 2.3), on sait que le facteur d'Araki-Woods libre associé absorbe $(L(\mathbb{F}_\infty), \text{tr})$:

$$(\Gamma(H_{\mathbb{R}}, U_t)'' , \varphi_U) \cong (\Gamma(H_{\mathbb{R}}, U_t)'' , \varphi_U) * (L(\mathbb{F}_\infty), \text{tr}) .$$

Le deuxième résultat surprenant de [20] est qu'il existe des facteurs d'Araki-Woods libres qui n'absorbent pas $(L(\mathbb{F}_\infty), \text{tr})$.

THÉORÈME 3.8. — *Soit λ la mesure de Lebesgue sur \mathbb{R} . Alors,*

$$\Gamma(\lambda, 1) \not\cong (\Gamma(\lambda, 1), \varphi_{\lambda,1}) * (L(\mathbb{F}_\infty), \text{tr}) .$$

4. APPENDICE : SUR LE LEMME DES 14ε

Dans 4.1 nous démontrons une généralisation du lemme technique 4.1 de [28], qui était à son tour une généralisation du lemme des 14ε dû à Murray & von Neumann [11] (voir [3] pour une version adaptée aux produits libres de type III). On exploite le fait que le produit libre $G_1 * G_2$ de deux groupes non-triviaux G_1, G_2 est très non-moyennable⁽³⁾, sauf si $G_1 \cong G_2 \cong \mathbb{Z}/2\mathbb{Z}$. Ceci explique le lemme 4.1 : il nous faut un élément non-trivial dans N_1 et deux éléments non-triviaux dans N_2 .

Le lemme 4.1 permet de calculer l'invariant τ d'un certain nombre de produits libres, voir proposition 4.2.

LEMME 4.1. — Soit N_i une algèbre de von Neumann munie d'un état fidèle ω_i , ($i = 1, 2$). Posons $(N, \omega) = (N_1, \omega_1) * (N_2, \omega_2)$. Soient $a \in N_1$ et $b, c \in N_2$. Supposons que les éléments a, b et c appartiennent au domaine de $\sigma_{i/2}^\omega$, où (σ_t^ω) est le groupe modulaire de l'état ω . Soit α_i un automorphisme de N_i qui satisfait $\omega_i \alpha_i = \omega_i$, ($i = 1, 2$). Notons $\alpha = \alpha_1 * \alpha_2$. Alors, pour tout $x \in N$,

$$\|x - \omega(x)1\|_2 \leq \mathcal{E}(a, b, c) \max\{\|xa - \alpha(a)x\|_2, \|xb - \alpha(b)x\|_2, \|xc - \alpha(c)x\|_2\} + \mathcal{F}(a, b, c) \|x\|_2$$

$$\text{où } \mathcal{E}(a, b, c) = 6\|a\|^3 + 4\|b\|^3 + 4\|c\|^3 ,$$

$$\mathcal{F}(a, b, c) = 3\mathcal{C}(a) + 2\mathcal{C}(b) + 2\mathcal{C}(c) + 12|\omega(cb^*)| \|cb^*\| ,$$

$$\mathcal{C}(a) = 2\|a\|^3 \|\sigma_{i/2}^\omega(a) - a\| + 2\|a\|^2 \|a^*a - 1\| + 3(1 + \|a\|^2) \|aa^* - 1\| + 6|\omega(a)| \|a\| .$$

Démonstration. — Représentons N_i sur l'espace de Hilbert H_i de la représentation GNS de ω_i et soit ξ_i le vecteur cyclique associé. Posons $(H, \xi) = (H_1, \xi_1) * (H_2, \xi_2)$. On rappelle [31] que

$$H = \mathbb{C}\xi \oplus (\overset{\circ}{H}_1 \otimes H(2, l)) \oplus (\overset{\circ}{H}_2 \otimes H(1, l)) ,$$

$$\text{où } \overset{\circ}{H}_i = H_i \ominus \mathbb{C}\xi_i,$$

$$H(2, l) = \mathbb{C}\xi \oplus \overset{\circ}{H}_2 \oplus (\overset{\circ}{H}_2 \otimes \overset{\circ}{H}_1) \oplus (\overset{\circ}{H}_2 \otimes \overset{\circ}{H}_1 \otimes \overset{\circ}{H}_2) \oplus \dots ,$$

$$H(1, l) = \mathbb{C}\xi \oplus \overset{\circ}{H}_1 \oplus (\overset{\circ}{H}_1 \otimes \overset{\circ}{H}_2) \oplus (\overset{\circ}{H}_1 \otimes \overset{\circ}{H}_2 \otimes \overset{\circ}{H}_1) \oplus \dots .$$

Pour $\zeta \in H$ et $y \in N$, on définit l'action à droite de y sur ζ par $\zeta \cdot y := Jy^*J\zeta$ où J est la conjugaison modulaire de l'état ω .

Choisissons $x \in N$ et définissons $\eta = x\xi$. On écrit $\eta = \omega(x)\xi + \mu + \gamma$ avec $\mu \in \overset{\circ}{H}_1 \otimes H(2, l)$ et $\gamma \in \overset{\circ}{H}_2 \otimes H(1, l)$. Posons alors $\overset{\circ}{x} = x - \omega(x)1$, $\eta_0 = \mu + \gamma$, $\tilde{\eta} = \alpha(a^*) \cdot \eta \cdot a$,

⁽³⁾Plus précisément $G_1 * G_2$ n'est pas intérieurement moyennable. On construit également une décomposition paradoxale explicite de G .

$\tilde{\gamma} = \alpha(a^*) \cdot \gamma \cdot a$ et $\tilde{\zeta} = \eta_0 - \gamma - \tilde{\gamma}$. Bien évidemment

$$\begin{aligned} \|\mu\|^2 + \|\gamma\|^2 &= \|\tilde{\zeta} + \gamma + \tilde{\gamma}\|^2 \geq \|\tilde{\zeta}\|^2 + \|\gamma\|^2 + \|\tilde{\gamma}\|^2 - 2|\langle \tilde{\zeta}, \gamma \rangle| - 2|\langle \tilde{\zeta}, \tilde{\gamma} \rangle| - 2|\langle \gamma, \tilde{\gamma} \rangle| \\ &\geq 2\|\gamma\|^2 - \|\gamma\|^2 - \|\tilde{\gamma}\|^2 - 2|\langle \tilde{\zeta}, \gamma \rangle| - 2|\langle \tilde{\zeta}, \tilde{\gamma} \rangle| - 2|\langle \gamma, \tilde{\gamma} \rangle|. \end{aligned}$$

Exactement de la même manière que dans la démonstration du lemme 4.1 de [28], on sait estimer tous les termes négatifs. On conclut que

$$(4) \quad \|\gamma\|^2 \leq \|\mu\|^2 + 2\|a\|^3 \|xa - \alpha(a)x\|_2 \|\mathring{x}\|_2 + \mathcal{C}(a) \|x\|_2 \|\mathring{x}\|_2.$$

On obtient une estimation analogue à l'aide des éléments b et c . En effet on pose $\eta' = \alpha(b^*) \cdot \eta \cdot b$, $\eta'' = \alpha(c^*) \cdot \eta \cdot c$, $\mu' = \alpha(b^*) \cdot \mu \cdot b$ et $\mu'' = \alpha(c^*) \cdot \mu \cdot c$. On définit $\zeta' = \eta_0 - \mu - \mu' - \mu''$. On trouve que

$$\begin{aligned} \|\mu\|^2 + \|\gamma\|^2 &\geq 3\|\mu\|^2 - \|\mu\|^2 - \|\mu'\|^2 - \|\mu''\|^2 - 2|\langle \zeta', \mu \rangle| - 2|\langle \zeta', \mu' \rangle| \\ &\quad - 2|\langle \zeta', \mu'' \rangle| - 2|\langle \mu, \mu' \rangle| - 2|\langle \mu, \mu'' \rangle| - 2|\langle \mu', \mu'' \rangle|. \end{aligned}$$

On estime de nouveau tous les termes négatifs et on obtient

$$(5) \quad \begin{aligned} 2\|\mu\|^2 &\leq \|\gamma\|^2 + 2\|b\|^3 \|xb - \alpha(b)x\|_2 \|\mathring{x}\|_2 + 2\|c\|^3 \|xc - \alpha(c)x\|_2 \|\mathring{x}\|_2 \\ &\quad + (\mathcal{C}(b) + \mathcal{C}(c) + 6\|cb^*\| |\omega(cb^*)|) \|x\|_2 \|\mathring{x}\|_2. \end{aligned}$$

Comme $\|\mu\|^2 + \|\gamma\|^2 = \|\mathring{x}\|_2^2$, une combinaison des inégalités (4) et (5) donne l'inégalité du lemme. \square

A priori l'invariant τ d'un facteur plein est difficile à calculer car on munit \mathbb{R} de la topologie induite d'une topologie quotient. L'intérêt de la proposition suivante est de donner une formule pour l'invariant τ en termes du groupe modulaire d'un seul état fidèle, sans qu'il faille connaître $\text{Out } N$. De la même manière que Shlyakhtenko déduit du lemme des 14ε de Barnett son corollaire 8.4 dans [23], nous déduisons du lemme 4.1 le résultat suivant, utilisé dans le §2.2.

PROPOSITION 4.2. — *Soit N_i des algèbres de von Neumann munies d'un état fidèle ω_i , ($i = 1, 2$). Soit $(N, \omega) = (N_1, \omega_1) * (N_2, \omega_2)$. On suppose que N_1 contient une suite (a_n) d'éléments qui sont analytiques par rapport à l'état ω et satisfont*

$$(6) \quad \|\sigma_{i/2}^\omega(a_n) - a_n\| \rightarrow 0, \quad \|a_n^* a_n - 1\|, \|a_n a_n^* - 1\| \rightarrow 0 \quad \text{et} \quad \omega(a_n) \rightarrow 0.$$

On suppose que N_2 contient des suites $(b_n), (c_n)$ qui satisfont les mêmes conditions que (a_n) ainsi que la condition $\omega(c_n b_n^) \rightarrow 0$. Alors,*

a) N est un facteur plein.

b) Notons $\text{Aut}(N_i, \omega_i)$ le groupe d'automorphismes de N_i préservant l'état ω_i et $\pi : \text{Aut}(N) \rightarrow \text{Out}(N)$ l'application quotient. Alors, l'homomorphisme

$$(7) \quad \text{Aut}(N_1, \omega_1) \times \text{Aut}(N_2, \omega_2) \rightarrow \text{Out}(N) : (\alpha_1, \alpha_2) \mapsto \pi(\alpha_1 * \alpha_2)$$

est un homéomorphisme à image fermé.

c) *L'invariant $\tau(N)$ est la topologie la plus faible qui rend continues les applications $t \mapsto \sigma_t^{\omega_i}$ de \mathbb{R} dans $\text{Aut } N_i$ ($i = 1, 2$).*

Démonstration. — Soient (x_k) une suite d'unitaires dans N et $\alpha_k \in \text{Aut}(N_1, \omega_1)$, $\beta_k \in \text{Aut}(N_2, \omega_2)$ des suites d'automorphismes. Supposons que $\text{Ad}(x_k^*) \circ (\alpha_k * \beta_k) \rightarrow \text{id}$ dans $\text{Aut}(N)$. Il suffit de démontrer que $\|x_k - \omega(x_k)1\|_2 \rightarrow 0$. En effet, prenant $\alpha_k = \text{id}$ et $\beta_k = \text{id}$ pour tout k , on aura démontré que chaque suite centrale est triviale et que donc, N est un facteur plein. On aura également démontré que l'homomorphisme (7) est un homéomorphisme. Finalement c) résulte de b).

Choisissons $\varepsilon > 0$. Prenons n tel que $\mathcal{F}(a_n, b_n, c_n) < \varepsilon/2$. Si $k \rightarrow \infty$, on a

$$\|x_k a_n - (\alpha_k * \beta_k)(a_n) x_k\|_2 = \|(\text{id} - \text{Ad}(x_k^*) \circ (\alpha_k * \beta_k))(a_n)\|_2 \rightarrow 0$$

et on a le même résultat en remplaçant (a_n) par (b_n) ou (c_n) . Il découle du lemme 4.1 qu'il existe n_0 tel que pour tout $n \geq n_0$ on a $\|x_n - \omega(x_n)1\|_2 < \varepsilon$. \square

On peut appliquer la proposition 4.2 à un produit libre $(N_1, \omega_1) * ((N_2, \omega_2) * (N_3, \omega_3))$ si chacune des algèbres (N_i, ω_i) contient une suite (a_n) qui satisfait les conditions (6). En effet, comme b_n et c_n sont *-libres dans ce cas-là, on a automatiquement que $\omega(c_n b_n^*) \rightarrow 0$. En particulier, il découle de la proposition 2.3 qu'on peut appliquer la proposition 4.2 à un facteur d'Araki-Woods libre associé à une représentation orthogonale de \mathbb{R} qui est une somme directe de trois représentations, pourvu qu'on démontre que chaque facteur d'Araki-Woods libre contient une suite (a_n) qui satisfait les conditions (6).

LEMME 4.3. — *Soit (U_t) une représentation orthogonale de \mathbb{R} sur l'espace de Hilbert réel $H_{\mathbb{R}}$. Soit $\Gamma(H_{\mathbb{R}}, U_t)''$ l'algèbre de von Neumann associée à l'état quasi-libre libre φ_U . Soit (σ_t) le groupe modulaire de l'état φ_U . Alors il existe une suite d'unitaires (u_n) dans $\Gamma(H_{\mathbb{R}}, U_t)''$ qui sont analytiques par rapport à (σ_t) et satisfont*

$$\|\sigma_z(u_n) - u_n\| \rightarrow 0 \text{ uniformément sur des compacts de } \mathbb{C}, \text{ et } \varphi_U(u_n) \rightarrow 0.$$

Démonstration. — Si $U_t = \text{id}$ pour tout $t \in \mathbb{R}$, le lemme est trivial. On suppose donc que (U_t) est non-trivial. Soit A l'opérateur auto-adjoint strictement positif sur le complexifié H de $H_{\mathbb{R}}$ tel que $U_t = A^{it}$. Soit J l'anti-unitaire canonique de H et $T = JA^{-1/2}$ l'involution sur H associée à (U_t) . Comme $A \neq 1$ et $JAJ = A^{-1}$, on peut prendre $\lambda > 1$ dans le spectre de A . Notons χ_n la fonction indicatrice de l'intervalle $[\lambda - \frac{1}{n}, \lambda + \frac{1}{n}]$ et prenons des vecteurs unité ξ_n dans l'image de $\chi_n(A)$. Comme $JAJ = A^{-1}$, les vecteurs ξ_n et $J\xi_n$ seront orthonormaux pour n suffisamment grand. On définit les éléments $x_n \in \Gamma(H_{\mathbb{R}}, U_t)''$ par $x_n = \ell(\xi_n) + \ell(T\xi_n)^*$. Il est clair que x_n est analytique par rapport à (σ_t) et que $\|\sigma_z^{\omega}(x_n) - \lambda^{iz} x_n\| \rightarrow 0$ uniformément sur des compacts de \mathbb{C} .

Définissons l'opérateur $y_n = \ell(\xi_n) + \frac{1}{\sqrt{\lambda}} \ell(J\xi_n)^*$ dans $B(\mathcal{F}(H))$. Alors $\|x_n - y_n\| \rightarrow 0$ et d'après le théorème 2.5, l'opérateur $y_n^* y_n$ a une distribution sans atomes par rapport à l'état vectoriel du vide qui ne dépend pas de n . Il existe donc une fonction continue $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que $\langle \exp(ig(y_n^* y_n)) \Omega, \Omega \rangle = 0$ pour tout n .

Choisissons $\varepsilon > 0$ et $K \subset \mathbb{C}$ compact. Comme g peut être approximée par des polynômes uniformément sur le spectre de $y_n^* y_n$, on peut prendre un polynôme P avec des coefficients réels tel que $|\langle \exp(ip(y_n^* y_n)) \Omega, \Omega \rangle| < \varepsilon/2$ pour tout n . On sait que $\|\sigma_z^\omega(x_n^* x_n) - x_n^* x_n\| \rightarrow 0$ uniformément sur des compacts de \mathbb{C} et que $\|x_n - y_n\| \rightarrow 0$. Pour n suffisamment grand $u := \exp(iP(x_n^* x_n))$ est alors un unitaire dans $\Gamma(H_{\mathbb{R}}, U_t)''$ qui satisfait $\|\sigma_z^\omega(u) - u\| < \varepsilon$ pour tout $z \in K$ et $|\varphi_U(u)| < \varepsilon$. \square

RÉFÉRENCES

- [1] H. ARAKI – On quasifree states of CAR and Bogoliubov automorphisms. *Publ. Res. Inst. Math. Sci.* **6** (1970), 385–442.
- [2] H. ARAKI & E.J. WOODS – A classification of factors. *Publ. Res. Inst. Math. Sci. Ser. A* **4** (1968), 51–130.
- [3] L. BARNETT – Free product von Neumann algebras of type III. *Proc. Amer. Math. Soc.* **123** (2) (1995), 543–553.
- [4] F. COMBES – Les facteurs de von Neumann de type III (d’après A. Connes). *Séminaire Bourbaki, Exp. No. 461*, Lecture Notes in Math. **514** Springer, Berlin, 1976, 124–137.
- [5] A. CONNES – Classification of injective factors. *Ann. of Math.* **104** (1976), 73–115.
- [6] A. CONNES – Almost periodic states and factors of type III₁. *J. Funct. Anal.* **16** (1974), 415–445.
- [7] A. CONNES – Une classification des facteurs de type III. *Ann. Sci. École Norm. Sup.* **6** (1973), 133–252.
- [8] J. DIXMIER – Les C*-algèbres et leurs représentations. Gauthier-Villars, Paris, 1964.
- [9] É. GHYS & P. DE LA HARPE – Sur les groupes hyperboliques d’après Mikhael Gromov. *Progress in Mathematics* **83**, Birkhäuser, Boston, 1990.
- [10] U. HAAGERUP – Connes’ bicentralizer problem and uniqueness of the injective factor of type III₁. *Acta Math.* **158** (1987), 95–148.
- [11] F.J. MURRAY & J. VON NEUMANN – On rings of operators. IV. *Ann. Math.* (2) **44** (1943), 716–808.
- [12] N. OZAWA – Solid von Neumann algebras. À paraître dans *Acta Math.*
- [13] G. PISIER & D. SHLYAKHTENKO – Grothendieck’s theorem for operator spaces. *Invent. Math.* **150** (2002), 185–217.

- [14] S. POPA & D. SHLYAKHTENKO – Universal properties of $L(\mathbb{F}_\infty)$ in subfactor theory. *Acta Math.* **191** (2003), 225–257.
- [15] R.T. POWERS & E. STØRMER – Free states of the canonical anticommutation relations. *Comm. Math. Phys.* **16** (1970), 1–33.
- [16] F. RĂDULESCU – A one-parameter group of automorphisms of $L(\mathbb{F}_\infty) \otimes B(H)$ scaling the trace. *C. R. Acad. Sci. Paris Sér. I Math.* **314** (1992), 1027–1032.
- [17] F. RĂDULESCU – A type III_λ factor with core isomorphic to the von Neumann algebra of a free group, tensor $B(H)$. Dans *Recent advances in operator algebras (Orléans, 1992)*, Astérisque **232** (1995), 203–209.
- [18] M.A. RIEFFEL & A. VAN DAELE – A bounded operator approach to Tomita-Takesaki theory. *Pacific J. Math.* **69** (1977), 187–221.
- [19] D. SHALE & W.F. STINESPRING – States of the Clifford algebra. *Ann. of Math.* **80** (1964), 365–381.
- [20] D. SHLYAKHTENKO – On multiplicity and free absorption for free Araki-Woods factors. *Prépublication*.
- [21] D. SHLYAKHTENKO – On the Classification of Full Factors of Type III. À paraître dans *Trans. AMS*.
- [22] D. SHLYAKHTENKO – Prime type III factors. *Proc. Natl. Acad. Sci. USA* **97** (2000), 12439–12441.
- [23] D. SHLYAKHTENKO – A -valued semicircular systems. *J. Funct. Anal.* **166** (1999), 1–47.
- [24] D. SHLYAKHTENKO – Some applications of freeness with amalgamation. *J. reine angew. Math.* **500** (1998), 191–212.
- [25] D. SHLYAKHTENKO – Free quasi-free states. *Pacific J. Math.* **177** (1997), 329–368.
- [26] G. SKANDALIS – Algèbres de von Neumann de groupes libres et probabilités non commutatives (d’après Voiculescu, etc.). Dans *Séminaire Bourbaki, Vol. 1992/93*. Astérisque **216** (1993), 87–102.
- [27] Ș. STRĂTILĂ & L. ZSIDÓ – Lectures on von Neumann algebras. Abacus Press, Tunbridge Wells, 1979.
- [28] S. VAES – Strictly outer actions of groups and quantum groups. À paraître dans *J. reine angew. Math.*
- [29] D.V. VOICULESCU – The analogues of entropy and of Fisher’s information measure in free probability theory. III. *Geom. Funct. Anal.* **6** (1996), 172–199.

- [30] D.V. VOICULESCU – The analogues of entropy and of Fisher’s information measure in free probability theory. II. *Invent. Math.* **118** (1994), 411–440.
- [31] D.V. VOICULESCU, K.J. DYKEMA & A. NICA – Free random variables. *CRM Monograph Series 1*, American Mathematical Society, Providence, 1992.
- [32] D.V. VOICULESCU – Symmetries of some reduced free product C^* -algebras. Dans *Operator algebras and their connections with topology and ergodic theory*, Lecture Notes in Math. **1132**, Springer, Berlin, 1985, 556–588.

Stefaan VAES

CNRS, Institut de Mathématiques de Jussieu
Algèbres d’Opérateurs, Plateau 7E
175 rue du Chevaleret
75013 Paris
E-mail : vaes@math.jussieu.fr

MÉTRIQUES KÄHLÉRIENNES À COURBURE SCALAIRE CONSTANTE : UNICITÉ, STABILITÉ

par Olivier BIQUARD

Une surface de Riemann compacte admet une métrique à courbure constante, unique à l'action près des automorphismes holomorphes. La recherche d'un phénomène analogue en dimension supérieure est une question centrale de la géométrie différentielle complexe. Plus précisément, il s'agit, étant donnée une variété kählérienne compacte, de trouver dans chaque classe de Kähler une métrique « canonique ». La notion la plus naturelle, introduite par Calabi, est celle de métrique extrémale. Les métriques kählériennes à courbure scalaire constante sont extrémales, et la réciproque est souvent vraie (en particulier en l'absence de champ de vecteurs holomorphe).

La question de l'existence des métriques kählériennes à courbure scalaire constante est très difficile, et peu de résultats sont connus, en dehors du cas où la classe canonique est un multiple de la classe de Kähler : le problème se réduit alors à l'existence d'une métrique Kähler-Einstein, pour lequel on renvoie à l'excellent exposé n° 830 de J.-P. Bourguignon et aux références qu'il contient. Rappelons simplement que ce problème est complètement résolu dans les cas $c_1 < 0$ (Aubin [2], Yau [66]) et $c_1 = 0$ (Yau [66, 67]), mais le cas Fano ($c_1 > 0$), demeure ouvert en dépit de nombreux résultats, en particulier de Tian, voir notamment [56, 58, 60]. Yau [68] a conjecturé que l'existence d'une métrique Kähler-Einstein dans le cas Fano est liée à une forme de stabilité algébrique de la variété, au sens de la théorie géométrique des invariants. Cette conjecture a été confirmée par Tian, qui a montré que l'existence d'une métrique Kähler-Einstein implique une notion de stabilité qu'il appelle K-stabilité [60]; ce travail l'a mené à formuler une « conjecture de Hitchin-Kobayashi » pour les variétés, liant stabilité et existence de métriques kählériennes à courbure scalaire constante. Les travaux de Donaldson, puis de Mabuchi, Chen et Tian, ont permis d'avancer de manière substantielle dans cette direction, et en particulier d'obtenir des résultats généraux d'unicité et de stabilité des métriques kählériennes à courbure scalaire constante.

On désignera par $\text{Aut}(M)$ le groupe des automorphismes holomorphes de la variété complexe compacte M , et par $\text{Aut}^0(M)$ la composante connexe de l'identité. Si (M, L) est une variété kählérienne polarisée (la classe de Kähler est $c_1(L)$), le groupe des automorphismes holomorphes de L modulo les automorphismes triviaux \mathbb{C}^* sera noté $\text{Aut}(M, L)$, c'est un sous-groupe de $\text{Aut}(M)$.

THÉORÈME 0.1 (Donaldson [18]). — *Soit (M, L) une variété complexe compacte polarisée, à groupe d'automorphismes $\text{Aut}(M, L)$ discret. Si la classe de Kähler $c_1(L)$ admet une métrique kählérienne à courbure scalaire constante, alors :*

1. *la métrique à courbure scalaire constante est unique dans la classe de Kähler ;*
2. *pour k assez grand, les plongements projectifs de M dans $PH^0(M, L^k)$ sont stables au sens de Chow-Mumford (i.e. (M, L) est asymptotiquement stable au sens de Chow-Mumford).*

La force du second énoncé se mesure au fait que la stabilité asymptotique d'une variété algébrique polarisée est une propriété notoirement difficile à vérifier.

À noter que dans le cas Kähler-Einstein Fano, l'unicité de la métrique Kähler-Einstein est aussi un problème délicat, résolu antérieurement par Bando et Mabuchi [3].

Le théorème a été étendu par Mabuchi, et Chen et Tian. Ces derniers aboutissent à l'énoncé le plus général suivant.

THÉORÈME 0.2. — *Sur une variété complexe compacte M , deux métriques kählériennes extrémales dans la même classe de Kähler diffèrent par un automorphisme holomorphe dans $\text{Aut}^0(M)$.*

Le cas d'une variété polarisée, à groupe d'automorphismes non trivial, est traité par Mabuchi [38, 39, 40, 41]. Il montre en outre une forme modifiée de stabilité au sens de Chow, tenant compte de l'action du centre de $\text{Aut}^0(M, L)$ sur $H^0(M, L^k)$, voir section 3.1.2.

Le théorème d'unicité définitif est montré par Chen et Tian [13] qui suppriment la condition que la classe de Kähler soit entière.

Les liens avec la K-(semi)stabilité, ainsi que la conjecture liant K-stabilité et existence de métriques kählériennes à courbure scalaire constante, seront détaillés dans la section 3.2.

Donnons un bref aperçu de la méthode de Donaldson : dans [16] il interprète le problème des métriques kählériennes à courbure scalaire constante comme un analogue de dimension infinie d'un problème d'application moment de l'action hamiltonienne d'un groupe compact G sur une variété symplectique. Le rôle de l'espace symétrique de type non compact $G^{\mathbb{C}}/G$ est joué par l'espace des potentiels de Kähler, et l'unicité résulterait du formalisme général des applications moment si, dans l'espace des potentiels de Kähler, deux points pouvaient toujours être joints par une géodésique [17]. Dans [18], Donaldson contourne la difficulté par une méthode de quantification consistant à approximer l'espace des potentiels de Kähler par les espaces symétriques $SL(N_k+1)/SU(N_k+1)$ des métriques de Fubini-Study sur les espaces projectifs $P^{N_k} = PH^0(M, L^k)^*$ dans lesquels se plonge M (le rôle de la constante de Planck étant joué par $1/k$) ; la métrique kählérienne à courbure scalaire constante s'approche dans chaque projectif par une métrique « équilibrée », dont l'existence est équivalente à la stabilité au sens de Chow (Zhang [70], Luo [35]).

Cet exposé a pour but d'expliquer la méthode utilisée par Donaldson, dont les principes sont utilisés aussi par Mabuchi. En revanche, Chen et Tian reviennent au programme initial en travaillant directement sur l'espace de dimension infinie des potentiels de Kähler, ce qui explique qu'ils n'ont plus besoin d'une polarisation ; la démonstration passe par des résultats nouveaux de régularité de solutions d'une équation de Monge-Ampère complexe homogène, que nous n'aborderons pas dans ce séminaire.

Dans la première section, nous exposons quelques généralités sur les métriques kählériennes, avant de passer au schéma formel posant le problème sous forme symplectique. Dans la seconde section, nous donnons la démonstration proprement dite du théorème, via la construction des métriques équilibrées. Enfin, dans la troisième section, nous effleurons diverses notions de stabilité des variétés algébriques, la conjecture sur le lien avec l'existence de métriques kählériennes à courbure scalaire constante, ainsi que certains développements très récents.

Remerciements. Je remercie Paul Gauduchon pour sa précise relecture du manuscrit.

1. GÉOMÉTRIE KÄHLÉRIENNE ET COURBURE SCALAIRE

1.1. Préliminaires

Ici, nous introduisons les métriques extrémales. Outre les articles fondateurs de Calabi [7, 8], d'excellentes références sur le sujet sont [4, 25, 62].

1.1.1. Métriques extrémales. — Soit M^{2n} une variété complexe, dont on notera la structure complexe J . Une forme de Kähler est une $(1,1)$ -forme fermée, telle que la formule $g(X, Y) = \omega(X, JY)$ définisse une métrique riemannienne. La connexion de Levi-Civita induit une connexion sur le fibré canonique $K_M = \Lambda^n \Omega_M^1$, dont la courbure s'écrit $i\rho_\omega$ pour une $(1,1)$ -forme réelle fermée ρ_ω appelée forme de Ricci. Elle est reliée au tenseur de Ricci via la structure complexe : $\rho_\omega(X, Y) = Ric(JX, Y)$. La forme $\frac{\rho_\omega}{2\pi}$ représente la classe de cohomologie $c_1(M)$, et la courbure scalaire de la métrique g est

$$s_\omega = 2\Lambda\rho_\omega$$

où Λ est l'opérateur de contraction par la forme de Kähler⁽¹⁾.

Pour fixer les notations, si, dans des coordonnées holomorphes locales (z^j) , la forme de Kähler s'écrit

$$\omega = \frac{i}{2} g_{j\bar{k}} dz^j \wedge d\bar{z}^k,$$

alors on obtient les formules

$$\rho_\omega = -i\partial\bar{\partial} \log \det(g_{j\bar{k}}), \quad s_\omega = -4g^{\ell\bar{m}} \frac{\partial^2}{\partial z^\ell \partial \bar{z}^m} \log \det(g_{j\bar{k}}).$$

⁽¹⁾La normalisation de la courbure scalaire en géométrie kählérienne fluctue suivant les auteurs, on trouve souvent le choix $\frac{1}{4}s_\omega$ qui simplifie certaines formules ; on a préféré ici s'en tenir strictement à la définition provenant de la géométrie riemannienne. Cela peut expliquer, pour certaines formules, la divergence entre ce séminaire et certains des articles cités.

La forme volume est $d\mu_\omega = \frac{\omega^n}{n!}$, le volume total $V = \int_M d\mu_\omega$ ne dépend que de la classe de cohomologie $\Omega = [\omega]$, et la moyenne \bar{s}_ω de la courbure scalaire est ainsi déterminée par la topologie :

$$\bar{s}_\omega = \frac{1}{V} \int_M s_\omega d\mu_\omega = \frac{1}{V} \int_M 2\rho_\omega \wedge \frac{\omega^{n-1}}{(n-1)!} = 4\pi n \frac{c_1(M)\Omega^{n-1}}{\Omega^n}.$$

Les identités kählériennes, impliquant l’opérateur $d^C = J^{-1}dJ$ sur les formes différentielles, permettent de retrouver l’identité de Bianchi :

$$d^* \rho_\omega = [\Lambda, d^C] \rho_\omega = -d^C \Lambda \rho_\omega = -\frac{1}{2} d^C s_\omega.$$

L’équation « s_ω constante » est donc équivalente à « ρ_ω harmonique » ; en particulier, si $c_1(M)$ est proportionnelle à la classe de Kähler (dont le représentant harmonique est justement ω), l’équation est équivalente à demander que ρ_ω soit proportionnelle à ω , c’est-à-dire que ω soit Kähler-Einstein.

La fonctionnelle de Calabi [7] est définie sur l’espace des formes de Kähler dans la classe de cohomologie fixée $\Omega \in H^2(M, \mathbb{R})$, par

$$\mathcal{C}(\omega) = \int_M s_\omega^2 d\mu_\omega.$$

Les points critiques de \mathcal{C} , appelés métriques extrémales, sont caractérisés par la condition que le champ de vecteurs $K = \sharp ds_\omega$ (où $\sharp : \Omega^1 \rightarrow T$ est la dualité symplectique, définie par $(\sharp\alpha) \lrcorner \omega = \alpha$) soit holomorphe. Notant

$$(1) \quad \mathcal{D} = 2\bar{\partial}\sharp\bar{\partial}$$

l’opérateur de Lichnerowicz, l’équation s’écrit donc $\mathcal{D}s_\omega = 0$. Elle est bien entendu vérifiée si s_ω est constante.

1.1.2. Le groupe d’automorphismes. — Un rôle important est joué ici par le groupe des automorphismes de M ou de (M, L) . Supposons donc donnée sur le fibré holomorphe en droites complexes L une métrique hermitienne, à courbure $F_L = -i\omega$. Notons également ξ le champ de vecteurs tautologique sur L . Un champ de vecteurs complexe sur L se décompose en

$$\hat{v} = \tilde{v} + f\xi,$$

où \tilde{v} est horizontal pour la connexion de L , et f est une fonction à valeurs complexes. On vérifie facilement que \hat{v} est holomorphe aux conditions suivantes :

1. \tilde{v} est le remonté horizontal d’un champ de vecteurs holomorphe v sur M ;
2. la fonction f satisfait $\sharp\bar{\partial}f = v$ (et donc $\mathcal{D}f = 0$).

On en déduit immédiatement :

LEMME 1.1. — *L’algèbre de Lie du groupe $\text{Aut}(M, L)$ est isomorphe à l’espace des solutions $f \in C_0^\infty(M, \mathbb{C})$ de l’équation $\mathcal{D}f = 0$.*

Une autre manière d'énoncer le lemme est de dire que l'algèbre de Lie de $\text{Aut}(M, L)$ s'identifie aux champs de vecteurs holomorphes, hamiltoniens-complexes (de la forme $\sharp\bar{\partial}f$ pour f une fonction complexe).

Le groupe d'automorphismes est une source importante d'obstructions à l'existence de métriques kählériennes à courbure scalaire constante. La plus ancienne est l'obstruction de Matsushima-Lichnerowicz [42, 33] : *si M admet une métrique kählérienne à courbure scalaire constante, alors l'algèbre de Lie des champs de vecteurs holomorphes sur M est réductive*. Plus précisément, cette algèbre se décompose en $\mathfrak{a} \oplus \mathfrak{h}$, où \mathfrak{a} est une algèbre de Lie complexe abélienne (constituée des champs de vecteurs parallèles), et \mathfrak{h} s'identifie aux solutions complexes de l'équation $\mathcal{D}f = 0$ (voir le lemme 1.1 dans le cas polarisé) ; les solutions réelles engendrent le groupe d'isométries de M .

Un autre invariant provenant du groupe d'automorphismes est le *caractère de Futaki*, qui sera introduit section 1.3.4.

1.1.3. Exemple : les surfaces complexes réglées. — Nous n'aborderons pas ici les exemples provenant du problème des métriques Kähler-Einstein (voir [5]).

Burns et De Bartolomeis [6] furent sans doute les premiers à détecter un lien entre stabilité algébrique et existence de métriques kählériennes à courbure scalaire nulle : sur une surface complexe réglée $S = PE$, où E est un fibré holomorphe de rang 2 sur une surface de Riemann Σ , à genre supérieur ou égal à 2, ils ont montré que S admet une métrique kählérienne à courbure scalaire nulle si et seulement si E provient d'une représentation du $\pi_1(\Sigma)$ dans PU_2 , c'est-à-dire, par le théorème de Narasimhan et Seshadri, est (poly-)stable. La métrique kählérienne à courbure scalaire nulle dans ce cas est localement symétrique (quotient du produit du disque hyperbolique et de la droite projective P^1).

Ce résultat a été étendu par LeBrun, en utilisant la théorie de Seiberg-Witten, aux métriques kählériennes à courbure scalaire constante négative sur les surfaces réglées [31].

Le problème de construction de métriques kählériennes à courbure scalaire constante est très difficile (équation d'ordre 4 sur le potentiel de Kähler), et très peu de résultats sont connus. Une exception notable existe en dimension 4 : les métriques kählériennes à courbure scalaire nulle, invariantes sous l'action d'un cercle, se décrivent explicitement en termes de fonctions harmoniques sur l'espace hyperbolique réel de dimension 3 (ansatz hyperbolique de LeBrun [30]). Parmi les applications de cet ansatz, citons les éclatements de surfaces réglées admettant l'action holomorphe d'un cercle [32] : les éclatements peuvent être codés en termes d'une structure parabolique sur E (au sens de Mehta et Seshadri), et LeBrun et Singer montrent que l'existence d'une métrique kählérienne à courbure scalaire nulle sur ces surfaces réglées éclatées (S^1 -invariantes) est équivalente à la stabilité du fibré parabolique associé. Une question intéressante est la généralisation éventuelle à toutes les surfaces réglées (même sans symétrie).

Enfin, toujours en dimension 4, les métriques kählériennes à courbure scalaire nulle peuvent être obtenues par une méthode twistorielle, qui permet de faire des recollements,

voir par exemple [29]. Plus récemment est apparue une méthode de désingularisation [49], s'appuyant sur la construction de métriques kählériennes à courbure scalaire nulle sur les désingularisations de certains quotients de \mathbb{C}^2 par des groupes finis [10] : par exemple, on peut construire une métrique kählérienne à courbure scalaire nulle sur une surface obtenue à partir de P^2 en éclatant 10 points—le nombre minimal de points nécessaire.

1.2. Brève revue du quotient kählérien

On fait ici un très bref rappel sur la théorie des invariants et le quotient symplectique, consulter [23, section 6.5] et [44].

1.2.1. Réduction symplectique. — Soit (\mathcal{X}, ϖ) une variété symplectique, et G un groupe d'algèbre de Lie \mathfrak{g} . Soit une action hamiltonienne de G sur (\mathcal{X}, ϖ) , c'est-à-dire qu'il existe une application $\mu : X \rightarrow \mathfrak{g}^*$, équivariante pour l'action de G , et satisfaisant pour tout $\xi \in \mathfrak{g}$,

$$d\langle \mu, \xi \rangle = v_\xi \lrcorner \varpi$$

où v_ξ est le champ de vecteurs sur \mathcal{X} induit par l'action infinitésimale de ξ . Une telle application est appelée application moment, elle est unique à l'addition près d'un élément de $(\mathfrak{g}^*)^G$.

Si 0 est une valeur régulière de μ , le théorème de Marsden-Weinstein indique que le quotient $\mu^{-1}(0)/G$ (appelé quotient symplectique de \mathcal{X} par G) admet une forme symplectique, qui, tirée en arrière sur $\mu^{-1}(0)$, coïncide avec ϖ .

Supposons que la forme symplectique provienne de la courbure d'un fibré en droites complexes L sur \mathcal{X} , donc $\varpi = iF_L \in 2\pi c_1(L)$. Le choix d'une application moment permet de remonter l'action de G à L , en faisant agir infinitésimalement $\xi \in \mathfrak{g}$ par

$$\hat{v}_\xi = \tilde{v}_\xi + \langle \mu(x), \xi \rangle \frac{d}{d\theta}$$

où \tilde{v}_ξ est le remonté horizontal de v_ξ sur L grâce à la connexion.

Si en outre \mathcal{X} est kählérienne et G est compact (préservant un produit scalaire sur \mathfrak{g} permettant de l'identifier avec \mathfrak{g}^*), alors le quotient symplectique est aussi kählérien. Il a en outre des liens profonds avec l'action complexifiée de $G^{\mathbb{C}}$ sur \mathcal{X} . L'action de G sur L se complexifie aussi, et on obtient des orbites de $G^{\mathbb{C}}$ dans L au-dessus d'orbites dans \mathcal{X} . Soit Z la fonction G -invariante sur L , introduite par Kempf et Ness [28], et définie par $Z(z) = -\log |z|^2$. La fonction Z est un potentiel pour la forme $\pi^*\omega$, où π est la projection $\pi : L - \{\text{section nulle}\} \rightarrow M$, c'est-à-dire $i\partial\bar{\partial}Z = \pi^*\omega$. Fixons un point base $p \in L$, au-dessus d'un point $x \in \mathcal{X}$, et $\xi \in \mathfrak{g}$: un calcul facile fournit les formules

$$\begin{aligned} \frac{d}{dt}Z(e^{it\xi}p) &= \langle \mu(e^{it\xi}x), \xi \rangle, \\ \frac{d^2}{dt^2}Z(e^{it\xi}p) &= |v_\xi(e^{it\xi}x)|^2. \end{aligned}$$

On voit donc que

- $\mu(x) = 0$ si et seulement si Z admet un point critique en p ;

- Z est une fonction convexe, en le sens suivant : compte tenu du choix de p , la fonction Z se tire en arrière en une fonction sur l'espace symétrique $\mathcal{Q} = G^{\mathbb{C}}/G$, que nous appellerons encore Z ; comme dans tout espace symétrique, les géodésiques sont données par l'action infinitésimale de \mathfrak{ig} , donc la seconde identité indique que Z est convexe sur \mathcal{Q} .

On déduit immédiatement le résultat classique d'unicité : dans \mathcal{X} , une $G^{\mathbb{C}}$ -orbite coupe $\mu^{-1}(0)$ en au plus une G -orbite. En outre, en notant G_x le groupe d'isotropie, on a l'isomorphisme des groupes discrets $G_x^{\mathbb{C}}/(G_x^{\mathbb{C}})^0 = G_x/G_x^0$.

L'existence est liée à la stabilité de la manière suivante : il y a équivalence entre

1. il existe un unique $g \in G^{\mathbb{C}}$ modulo G tel que $\mu(gx) = 0$;
2. la fonctionnelle Z sur $G^{\mathbb{C}}/G$, correspondant à x , est propre ;
3. l'orbite complexe $G^{\mathbb{C}}p$ dans L est fermée, et le groupe d'isotropie G_x est fini.

La troisième condition est la condition que x soit un point *stable*, au sens de la théorie géométrique des invariants (on dit parfois *proprement stable*), et la seconde condition peut être vue comme une condition de stabilité analytique.

La condition sur le groupe d'isotropie est parfois supprimée, et on parle alors de point faiblement stable, ou *polystable*. L'existence d'un zéro de μ dans l'orbite complexe reste alors assurée, mais l'unicité de g ne l'est que modulo $(G_x^{\mathbb{C}})^0$.

Enfin, la semistabilité est définie en demandant que l'orbite complexe $G^{\mathbb{C}}p$ ne contienne pas 0 dans son adhérence.

1.2.2. La correspondance de Hitchin-Kobayashi. — Si cette théorie du quotient kählerien est bien établie en dimension finie, tel n'est pas le cas en dimension infinie. Cependant, en dimension infinie existe un problème qui sert de paradigme pour celui des métriques kähleriennes à courbure scalaire constante. Soit une variété kählerienne compacte (M^{2n}, ω) , et E un fibré holomorphe sur X , de rang r (pour simplifier, on supposera $\deg_{\omega}(E) = c_1(E)[\omega]^{n-1} = 0$). Le problème est de trouver une métrique hermitienne h sur E , de courbure F_h , satisfaisant l'équation de Hermite-Einstein :

$$\Lambda F_h = 0.$$

La conjecture, posée par Hitchin et Kobayashi, fut résolue par Donaldson [15], et Uhlenbeck et Yau [64] : si E est indécomposable, l'existence d'une telle métrique est équivalente à la stabilité du fibré E , à savoir, pour tout sous-faisceau strict E' de E , on a $\deg_{\omega}(E') < 0$. Si E est décomposable, la bonne condition est la polystabilité de E , à savoir, E est somme directe de sous-fibrés stables de degré nul.

Ce problème s'interprète en termes d'application moment de la manière suivante. Au lieu de faire varier la métrique, on fixe une métrique hermitienne h_0 sur E , et on considère l'espace \mathcal{X} des opérateurs $\bar{\partial}$ sur E , ou de manière équivalente, des connexions unitaires A . Il s'agit d'un espace affine, de direction l'espace des sections du fibré $\Omega^{0,1} \otimes \text{End}(E)$; la norme L^2 le munit d'une structure kählerienne (plate). Le groupe de jauge G est défini comme le groupe des transformations unitaires du fibré E (il s'agit donc des sections d'un fibré sur M de fibre $U(r)$). Son complexifié $G^{\mathbb{C}}$ est le groupe des

transformations complexes de E , il agit sur \mathcal{X} par $g(\bar{\partial}^E) = g \circ \bar{\partial}^E \circ g^{-1}$, et sa partie unitaire G en préserve la structure kählérienne.

Le problème de Hermite-Einstein s’interprète en termes de l’action de $G^{\mathbb{C}}$ sur \mathcal{X} : en effet, la métrique $h = g^*h_0g$ sur E est d’Hermite-Einstein si et seulement si la métrique h_0 est Hermite-Einstein sur $g(E)$. Notons que la paramétrisation des métriques h par g^*h_0g n’est rien d’autre que l’identification de l’espace $\mathcal{Q} = G^{\mathbb{C}}/G$ à l’espace des métriques hermitiennes sur E .

Un calcul facile montre que $A \rightarrow \Lambda F_A$ est une application moment pour l’action de G sur \mathcal{X} . L’unicité de la métrique d’Hermite-Einstein peut être déduite du formalisme général de l’application moment, avec la fonctionnelle Z sur \mathcal{Q} définie plus haut qui n’est autre que la fonctionnelle de Donaldson. L’existence de la métrique d’Hermite-Einstein est beaucoup plus difficile, mais la relation entre la propriété de la fonctionnelle de Donaldson et la stabilité du fibré est au cœur de la démonstration, voir par exemple dans [55].

1.3. Le point de vue symplectique

Revenons au problème des métriques kählériennes à courbure scalaire constante. Jusqu’à la fin de cette section, nous décrivons le point de vue symplectique adopté par Donaldson [16, 17].

1.3.1. La courbure scalaire comme application moment. — Le caractère symplectique du problème se voit en changeant de point de vue : au lieu de fixer la structure complexe de la variété et de faire varier la forme de Kähler, on fixe une variété symplectique compacte (M^{2n}, ω) , et on considère l’ensemble \mathcal{J} des structures presque-complexes J compatibles à ω , c’est-à-dire satisfaisant $\omega(JX, JY) = \omega(X, Y)$ et $g(X, Y) = \omega(X, JY)$ est une métrique riemannienne.

On peut voir \mathcal{J} comme l’espace des sections d’un fibré sur M de fibre l’espace hermitien symétrique $Sp(2n)/U(n)$. La structure complexe de $Sp(2n)/U(n)$, et sa métrique couplée à la forme volume $d\mu_\omega$ de M , donnent formellement à \mathcal{J} une structure de variété kählérienne de dimension infinie, dont on notera la forme de Kähler ϖ . Plus précisément, une structure presque-complexe étant fixée, on décrit les autres structures presque-complexes par leur espace $\Omega^{1,0}$, paramétré comme le graphe d’un $\phi \in \text{Hom}(\Omega^{1,0}, \Omega^{0,1}) = \Omega^{0,1} \otimes T^{1,0}$. La compatibilité à ω s’écrit alors $\phi \lrcorner \omega = 0$, où l’opération \lrcorner doit être comprise comme la composition de la contraction et du produit extérieur :

$$(\Omega^{0,1} \otimes T^{1,0}) \otimes \Omega^{1,1} \rightarrow \Omega^{0,1} \otimes \Omega^{0,1} \rightarrow \Omega^{0,2}.$$

Le groupe G des symplectomorphismes hamiltoniens de M agit sur \mathcal{J} par $\phi(J) = \phi_*J\phi_*^{-1}$. Cette action préserve la structure kählérienne. L’algèbre de Lie \mathfrak{g} s’identifie à l’espace $C_0^\infty(M)$ des fonctions d’intégrale nulle, une telle fonction H définissant le champ de vecteurs hamiltonien $X_H = \sharp dH$ sur M .

Enfin, la définition de la courbure scalaire peut être étendue à ce cadre presque-complexe : l’opérateur $\bar{\partial} : \Omega^{1,0}M \rightarrow \Omega^{1,1}M$ s’étend en une connexion hermitienne sur

$\Omega^{1,0}M$ (connexion de Chern), et donc sur le fibré canonique K_M . Sa courbure s'écrit $i\rho_J$, d'où on déduit la courbure scalaire $s_J = 2\Lambda\rho_J$.

La courbure scalaire définit un élément de \mathfrak{g}^* par l'application

$$H \longrightarrow \int_M s_J H d\mu_\omega.$$

LEMME 1.2 (Donaldson). — *L'application $J \rightarrow \frac{1}{4}s_J$ est une application moment pour l'action de G sur \mathcal{J} . En particulier, le lieu d'annulation de l'application moment est exactement l'espace des structures presque-complexes à courbure scalaire constante.*

Démonstration. — Le tenseur de Nijenhuis $N \in \Omega^{0,2} \otimes T^{1,0}$ de la structure presque-complexe J est défini par $\bar{\partial}^2 f = N \lrcorner \partial f$ pour toute fonction f . Il intervient dans le calcul de l'action infinitésimale d'un champ de vecteurs X sur J :

$$(2) \quad \mathcal{L}_X J = \bar{\partial} X^{1,0} - X^{0,1} \lrcorner N$$

où le résultat est vu comme une section de $\Omega^{0,1} \otimes T^{1,0}$.

Fixons une fonction $H \in \mathfrak{g} = C_0^\infty(M)$. Le lemme signifie qu'on a sur \mathcal{J} l'égalité

$$d\langle s_J, H \rangle = 4(\mathcal{L}_{X_H} J) \lrcorner \varpi,$$

c'est-à-dire, pour toute variation infinitésimale $\phi \in T_J \mathcal{J}$,

$$(3) \quad \int_M d_J s(\phi) H \omega^n = 4 \int_M \operatorname{Re} \langle i(\bar{\partial} \# \bar{\partial} H - (\# \partial H) \lrcorner N), \phi \rangle \omega^n.$$

On renvoie à [16] pour les détails du calcul de la différentielle de la courbure scalaire dans ce contexte. □

1.3.2. *Complexification.* — Le groupe de symplectomorphismes G n'admet pas de complexification. Cependant, \mathcal{J} étant une variété complexe, l'action infinitésimale de \mathfrak{g} peut être complexifiée. La fonction $iH \in iC_0^\infty(M)$ agira donc sur J par $J\mathcal{L}_{X_H} J$. La distribution de \mathcal{J} , engendrée en J par les $\mathcal{L}_{X_H} J$ et $J\mathcal{L}_{X_H} J$, est involutive, et ses feuilles maximales jouent le rôle des orbites du groupe complexe manquant.

Précisons cette discussion quand on se restreint aux structures complexes intégrables $\mathcal{J}_{int} \subset \mathcal{J}$. Si J est intégrable, alors par (2), on a

$$(4) \quad J\mathcal{L}_{X_H} J = \bar{\partial}(iX_H^{1,0}) = \mathcal{L}_{JX_H} J,$$

donc l'action complexifiée infinitésimale agit par difféomorphismes (infinitésimaux) sur J . Un point de vue équivalent consiste à fixer J et à modifier plutôt la forme de Kähler ω par $-\mathcal{L}_{JX_H} \omega = -dd^C H$.

Introduisons à présent l'espace des formes de Kähler dans la même classe que ω :

$$\mathcal{K} = \{\omega_\varphi = \omega + dd^C \varphi, \omega_\varphi > 0\}.$$

Par le lemme de Moser, on peut choisir, pour chaque $\omega_\varphi \in \mathcal{K}$, un difféomorphisme F_φ , dépendant de manière régulière de φ , tel que $F_\varphi^* \omega_\varphi = \omega$. On montre alors que la feuille maximale de la distribution involutive de \mathcal{J} passant par J est l'image de l'application⁽²⁾

$$\mathcal{K} \times G \rightarrow \mathcal{J}, \quad (\omega_\varphi, \sigma) \rightarrow \sigma^* F_\varphi^* J.$$

Il apparaît clairement que, $J \in \mathcal{J}_{int}$ étant fixé, le rôle de l'espace symétrique $G^{\mathbb{C}}/G$ est joué par l'espace des formes de Kähler \mathcal{K} .

Remarque 1.3. — Le calcul de la variation de la courbure scalaire s_{ω_φ} par rapport à φ (voir par exemple [25]), est essentiellement équivalent au calcul montrant que s_J est une application moment, le lien entre les deux points de vue étant donné par l'application d'un difféomorphisme infinitésimal, grâce à (4). Plus précisément, l'action infinitésimale de $-JX_\varphi$ sur J (ω restant fixe) mène à une variation de J par $\phi = -i\bar{\partial}X_\varphi^{1,0} = -\frac{i}{2}\mathcal{D}\varphi$ (\mathcal{D} l'opérateur de Lichnerowicz défini en (1)), et, par (3), à une variation de s_J par $-2\text{Re}(i\mathcal{D}^*\phi) = -\mathcal{D}^*\mathcal{D}\varphi$; puis on revient à J en faisant agir le champ de vecteurs JX_φ , d'où une contribution supplémentaire $\mathcal{L}_{JX_\varphi}s = \langle ds, d\varphi \rangle$, ce qui donne la variation complète, pour une variation $\dot{\omega} = dd^C\varphi$:

$$(5) \quad \dot{s} = -\mathcal{D}^*\mathcal{D}\varphi + \langle ds, d\varphi \rangle.$$

1.3.3. *L'espace des métriques kählériennes.* — Même si le complexifié du groupe des symplectomorphismes n'existe pas, Donaldson [17] a observé que l'espace \mathcal{K} des formes de Kähler dans la classe Ω a toutes les propriétés souhaitables pour $G^{\mathbb{C}}/G$. En effet, \mathcal{K} est muni d'une métrique riemannienne naturelle, la métrique de Mabuchi [37] définie en $\omega \in \mathcal{K}$ par

$$\langle \varphi_1, \varphi_2 \rangle = \frac{1}{V} \int_M \varphi_1 \varphi_2 d\mu_\omega$$

qui en fait formellement un espace symétrique de dimension infinie, à courbure négative.

La fonctionnelle Z de la section 1.2 admet un analogue sur \mathcal{K} , à savoir la K-énergie de Mabuchi [36] : on vérifie que la 1-forme σ sur \mathcal{K} définie en ω par ($\varphi \in C_0^\infty(M)$)

$$(6) \quad \sigma_\omega(\varphi) = - \int_M \varphi s_\omega d\mu_\omega$$

est fermée, et le choix d'un point base $\omega_0 \in \mathcal{K}$ permet alors de définir sa primitive par

$$\mathbf{E}_{\omega_0}(\omega) = - \int_0^1 dt \int_M \dot{\varphi}_t (s_{\omega_t} - \bar{s}) d\mu_{\omega_t},$$

où $\omega_t = \omega_0 + dd^C\varphi_t$ est un chemin liant ω_0 à $\omega = \omega_1$ dans \mathcal{K} , et $\bar{s} = 4\pi n \frac{c_1(M)\Omega^{n-1}}{\Omega^n}$.

La K-énergie satisfait les propriétés formelles de l'application moment expliquées dans la section 1.2 :

- les points critiques de \mathbf{E} sont exactement les métriques à courbure scalaire constante;

⁽²⁾L'image, telle que nous la décrivons, n'est bien définie que pour $H^1(M, \mathbb{R}) = 0$.

- \mathbf{E} est convexe le long des géodésiques de \mathcal{K} : le long d’une géodésique $\omega_t = \omega_0 + dd^C \varphi_t$, on a $\frac{d^2}{dt^2} \mathbf{E}(\omega_t) = \int_M |\mathcal{D}_{\omega_t} \dot{\varphi}_t|^2 d\mu_{\omega_t}$. De là se déduit facilement que si une géodésique de \mathcal{K} relie deux métriques kählériennes à courbure scalaire constante, alors celles-ci diffèrent par un élément du groupe $\text{Aut}^0(M)$.

Cependant, l’existence de géodésiques entre deux points quelconques de cet espace symétrique de dimension infinie est un problème très difficile, qui se ramène à une équation de Monge-Ampère complexe homogène : plus précisément, soit la surface de Riemann à bord $\Sigma = [0, 1] \times S^1$, alors le chemin de métriques de Kähler $\omega_t = \omega + i\partial\bar{\partial}\varphi_t$ est une géodésique si et seulement si sur $M \times \Sigma$ est satisfaite l’équation

$$(\omega + i\partial\bar{\partial}\Phi)^{n+1} = 0, \quad \text{où } \Phi(x, t, \theta) = \varphi_t(x).$$

C’est l’étude de cette équation qui avait amené Semmes à redécouvrir la métrique de Mabuchi [54].

Chen [12] montre l’existence de solutions $C^{1,1}$, mais leur régularité est un problème très difficile, qu’il ne résout que sur les variétés à $c_1(M)$ négatif ; l’unicité des métriques kählériennes à courbure scalaire constante dans ce cas s’en déduit. Après les progrès de [9, 19], la solution définitive semble donnée dans les travaux récents de Chen et Tian [13], conduisant à l’énoncé le plus général du théorème 0.2.

1.3.4. Le caractère de Futaki. — La version infinitésimale de la K-énergie de Mabuchi permet de définir un invariant classique des champs de vecteurs holomorphes par rapport à une classe de Kähler, l’invariant de Futaki ([24], voir aussi [8] ; on suit ici [5, 4.6]).

La 1-forme σ sur \mathcal{K} définie par (6) est invariante sous l’action de $\text{Aut}^0(M)$ agissant sur \mathcal{K} . Il est facile de voir que le champ de vecteurs hamiltonien-complexe $v = \sharp\bar{\partial}f$ agit en ω par $dd^C\varphi_v$, avec $\varphi_v = \text{Im}(f)$. D’un autre côté, par invariance, et puisque σ est fermée,

$$0 = \mathcal{L}_v\sigma = d(v \lrcorner \sigma) = -d \int_M \varphi_v(s_\omega - \bar{s}_\omega) d\mu_\omega.$$

On déduit que la quantité

$$\mathcal{F}(v) = - \int \varphi_v(s_\omega - \bar{s}_\omega) d\mu_\omega$$

ne dépend pas de la métrique $\omega \in \mathcal{K}$ choisie, donc ne dépend que du champ de vecteurs v : c’est l’invariant de Futaki de v .

Son annulation est évidemment une condition nécessaire à l’existence d’une métrique kählérienne à courbure scalaire constante dans la classe de Kähler. En outre, en évaluant \mathcal{F} sur le champ de vecteurs $i\sharp\bar{\partial}s_\omega$, on voit qu’une métrique extrémale est à courbure scalaire constante si et seulement si le caractère de Futaki est nul.

2. LA QUANTIFICATION DE DONALDSON

Dans cette section, nous détaillons la méthode de Donaldson pour montrer le théorème 0.1, et notamment son idée de « quantification ».

2.1. L'approximation de dimension finie

2.1.1. Métrique équilibrée. — Soit (M, L) une variété kählérienne munie d'un fibré très ample L , avec forme de Kähler $\omega \in 2\pi c_1(L)$, provenant d'une métrique sur L à courbure $F_L = -i\omega$ (la métrique de L est donc fixée à une constante multiplicative près). Nous avons un plongement de Kodaira de M dans le projectif $P^N = PE^*$, où $E = H^0(M, L)$ est de dimension $N + 1$.

Le choix d'une métrique hermitienne sur E permet de définir sur P^N la métrique de Fubini-Study

$$\omega_{FS} = iF_{\mathcal{O}(1)} = i\partial\bar{\partial} \log \sum_0^N |z^\alpha|^2.$$

Définissons alors la matrice $M = (M_{\alpha\beta}) \in \mathfrak{u}(N + 1)$ par

$$M_{\alpha\beta} = i \int_M \frac{\bar{z}^\alpha z^\beta}{|z|^2} d\mu_{FS}.$$

DÉFINITION 2.1. — *Le plongement $i : M \hookrightarrow P^N$ est équilibré si $(M_{\alpha\beta}) = \lambda(\delta_{\alpha\beta})$, c'est-à-dire si la projection de M sur $\mathfrak{su}(N + 1)$ est nulle.*

La variété polarisée (M, L) est équilibrée s'il existe sur $H^0(M, L)$ une métrique telle que le plongement de Kodaira de M dans $PH^0(M, L)^$ soit équilibré.*

D'un autre côté, étant donnée la métrique kählérienne ω sur M , la norme L^2 fournit une métrique sur $H^0(M, L)$; si (s_α) en est une base orthonormale, définissons une fonction sur M par la formule

$$\rho(\omega) = \sum_0^N |s_\alpha|^2;$$

cette fonction ne dépend ni du choix de base orthonormale (s_α) , ni de la multiplication de la métrique de L par une constante, donc est intrinsèquement attachée à ω . Il est facile de voir que ρ permet de calculer la différence entre ω et la métrique de Fubini-Study, tirée en arrière par le plongement de Kodaira :

$$\omega = i^* \omega_{FS} - i\partial\bar{\partial} \log \rho.$$

On en déduit immédiatement l'équivalence entre :

- le plongement $i : M \hookrightarrow P^N$ est équilibré pour une métrique de Fubini-Study ω_{FS} sur P^N , et $\omega = i^* \omega_{FS}$;
- la fonction $\rho(\omega)$ est constante.

Le problème de trouver un plongement équilibré admet une interprétation en termes d'application moment. Considérons l'espace \mathcal{B} des bases de E ; une base $s = (s_0, \dots, s_N) \in \mathcal{B}$ définit une métrique H sur E , donc une métrique de Fubini-Study sur PE^* , qui, tirée en arrière, fournit une métrique $\omega(s)$ sur M ; une manière équivalente de définir $\omega(s)$ est de fixer la métrique $h = FS(H)$ de L en décidant que

$$(7) \quad \sum_0^N |s_\alpha|_h^2 = 1,$$

puis de prendre $\omega(s) = iF_h$. La métrique h de L et ω_s permettent de définir

$$\mu_{SU}(s) = \pi_{SU}[i(s_\alpha, s_\beta)_h] \in SU(N + 1).$$

THÉORÈME 2.2. — *L'espace $\mathcal{B}/\text{Aut}^0(M, L)$ a une structure kählérienne pour laquelle l'action de $SU(N + 1)$ est hamiltonienne, avec application moment μ_{SU} .*

Cette structure kählérienne est construite par Donaldson [18] par un quotient de dimension infinie. Phong et Sturm [48] en donnent une construction de dimension finie, basée sur des idées de Zhang [70], utilisant des techniques de type « métriques de Quillen ». Plutôt que de donner les détails de cette construction, nous décrivons ci-après une approche élémentaire, suffisante pour les résultats démontrés dans cet article, consistant à expliciter dans cette situation la fonctionnelle Z de la section 1.2.

2.1.2. La fonctionnelle convexe Z . — Plusieurs points de vue existent, notamment celui de Zhang qui utilise la « norme de Chow », mais nous suivons ici plutôt [22, 47].

L'espace $\tilde{\mathcal{K}}$ des métriques hermitiennes sur L s'identifie à l'espace des potentiels de Kähler : si deux métriques h et h_0 sont liées par $h = e^{-\varphi}h_0$, alors les formes de Kähler diffèrent par $i\partial\bar{\partial}\varphi$. On définit sur $\tilde{\mathcal{K}}$, à une constante près, la fonctionnelle classique \mathbf{I} par sa différentielle en $\omega \in \mathcal{K}$:

$$d_\omega \mathbf{I}(\varphi) = - \int_M \varphi d\mu_\omega.$$

Cette formule définit une 1-forme fermée sur $\tilde{\mathcal{K}}$, dont la primitive fournit, à une constante près, la fonctionnelle \mathbf{I} .

Fixons un déterminant de E . À une métrique hermitienne H sur E , on associe [22]

$$Z(H) = -\mathbf{I}(FS(H)) + \frac{V}{N + 1} \log \det H,$$

où $FS(H)$ est la métrique induite sur L par (7); le deuxième terme dans Z a pour seul but de rendre Z invariante par homothétie : $Z(aH) = Z(H)$.

LEMME 2.3. — *Soit $\xi \in \mathfrak{su}(N + 1)$; on a en $t = 0$ les dérivées suivantes de $f(t) = Z(e^{it\xi}H)$:*

$$\begin{aligned} f'(0) &= \sum i\xi_{\alpha\beta}(s_\alpha, s_\beta)_{FS(H)}, \\ f''(0) &= \int_M |\pi_{\mathcal{N}}(v_\xi)|^2 d\mu_{FS(H)}, \end{aligned}$$

où v_ξ est le champ de vecteurs engendré par l'action infinitésimale de ξ , et $\pi_{\mathcal{N}}$ représente la projection sur le fibré normal de $M \hookrightarrow P^N$.

De la dérivée première, on déduit la caractérisation des métriques équilibrées comme points critiques de Z ; de la seconde, le corollaire suivant, qui peut aussi être vu directement comme conséquence du théorème 2.2.

COROLLAIRE 2.4. — *Si $\text{Aut}(M, L)$ est discret, alors une métrique équilibrée pour (M, L) est unique (à homothétie près).*

Démonstration. — Nous donnons une idée de la démonstration du lemme 2.3, notamment la seconde formule qui a une certaine importance dans la suite. On suit [22, proposition 1].

Notons $h_0 = FS(H)$ la métrique induite sur L ; dans ce calcul, toutes les normes des sections seront écrites par rapport à h_0 . En diagonalisant ξ dans une base orthonormale (s_α) , on peut écrire

$$H_t = e^{i\xi t} H = \text{diag}(e^{\lambda_\alpha t}).$$

On a la base H_t -orthonormale $s_\alpha(t) = e^{-\frac{\lambda_\alpha}{2}t} s_\alpha$, et donc $h_t = FS(H_t) = e^{-\varphi_t} h_0$ est défini par

$$\varphi_t = \log \sum e^{-\lambda_\alpha t} |s_\alpha|^2,$$

d'où en particulier à $t = 0$:

$$\begin{aligned} \dot{\varphi} &= - \sum \lambda_\alpha |s_\alpha|^2, \\ \ddot{\varphi} &= \sum \lambda_\alpha^2 |s_\alpha|^2 - \left(\sum \lambda_\alpha |s_\alpha|^2 \right)^2. \end{aligned}$$

Il en résulte immédiatement

$$f'(0) = \int_M \dot{\varphi} d\mu_\omega = \int_M \sum -\lambda_\alpha |s_\alpha|^2 d\mu_\omega,$$

d'où la première formule. Pour calculer la dérivée seconde, on s'appuie sur la formule générale

$$\frac{d^2 \mathbf{I}}{dt^2} = \int_M (-\ddot{\varphi} + \dot{\varphi} \partial^* \partial \dot{\varphi}) d\mu_\omega$$

pour arriver à

$$f''(0) = \int_M \left(\ddot{\varphi} - \frac{1}{2} |d\dot{\varphi}|^2 \right) d\mu_\omega.$$

Voyons P^N comme le quotient de la sphère $S^{2N+1} \subset \mathbb{C}^{N+1}$ par l'action du cercle (en particulier sa métrique est induite de celle de la sphère). L'action infinitésimale de ξ induit un champ de vecteurs $\tilde{v}_\xi = \sum \lambda_\alpha z_\alpha \frac{\partial}{\partial z_\alpha}$ sur \mathbb{C}^{N+1} , se projetant sur v_ξ dans P^N . On a donc, sur la sphère $S^{2N+1} = \{ \sum |z_\alpha|^2 = 1 \}$,

$$|v_\xi|^2 = \sum \lambda^2 |z_\alpha|^2 - \left(\sum \lambda_\alpha |z_\alpha|^2 \right)^2.$$

Puisque $\sum |s_\alpha|^2 = 1$, il est clair que sur l'image de M dans P^N , on a

$$-\ddot{\varphi} = |v_\xi|^2.$$

D'un autre côté, on voit que $\tilde{v}_\xi = \#\bar{\partial} \sum \lambda_\alpha |z_\alpha|^2$, donc la projection de v_ξ sur $TM \subset TP^N$ n'est autre que

$$(8) \quad \pi_{TM}(v_\xi) = \#\bar{\partial} \sum \lambda_\alpha |s_\alpha|^2 = -\#\bar{\partial}\dot{\varphi},$$

dont la norme est $|\bar{\partial}\dot{\varphi}|^2 = \frac{1}{2}|d\dot{\varphi}|^2$. Finalement,

$$f''(0) = \int_M (|v_\xi|^2 - |\pi_{TM}(v_\xi)|^2) d\mu_\omega = \int_M |\pi_{\mathcal{N}} v_\xi|^2 d\mu_\omega.$$

□

2.1.3. Résultats. — Supposons la variété polarisée (M, L) donnée, avec $\text{Aut}(M, L)$ discret. Si (M, L^k) est équilibrée, alors il existe une unique métrique de Fubini-Study équilibrée sur P^{N_k} , que nous noterons $\tilde{\omega}_k$. La métrique tirée en arrière par le plongement $i_k : M \hookrightarrow P^{N_k}$,

$$\omega_k = \frac{1}{k} i_k^* \tilde{\omega}_k,$$

est dans la classe de cohomologie fixée $2\pi c_1(L)$.

THÉORÈME 2.5. — *Supposons que $\text{Aut}(M, L)$ soit discret, alors :*

1. *s'il existe ω_∞ à courbure scalaire constante, alors pour k suffisamment grand, (M, L^k) est équilibrée, et les métriques ω_k induites satisfont $\omega_k \rightarrow \omega_\infty$;*
2. *si (M, L^k) est équilibrée, induisant la métrique ω_k sur M , et $\omega_k \rightarrow \omega_\infty$, alors ω_∞ est à courbure scalaire constante.*

L'unicité dans le théorème 0.1, est une conséquence du premier énoncé : l'unicité des métriques équilibrées ω_k implique l'unicité de la limite à courbure scalaire constante ω_∞ .

Le second énoncé est une réciproque élémentaire : si les métriques équilibrées convergent, c'est vers une métrique à courbure scalaire constante (mais montrer a priori la convergence semble très difficile).

Dans le cas où le groupe d'automorphismes n'est pas discret, un résultat similaire est montré par Mabuchi, voir section 3.1.2.

La suite de cette section est consacrée à la démonstration du théorème 2.5.

2.2. Noyau de Bergmann

Soit (M^{2n}, L) une variété kählérienne polarisée. Dans cette partie il sera plus commode de normaliser la forme de Kähler ω dans $c_1(L)$; elle provient ainsi d'une métrique sur L , de courbure $F_L = -2\pi i\omega$ (la métrique de L est fixée à une constante multiplicative près). Si l'entier k est assez grand, le plongement de Kodaira envoie M dans l'espace projectif $P^{N_k} = PH^0(M, L^k)^*$. La dimension $N_k + 1 = h^0(M, L^k)$ est donnée par le théorème de Riemann-Roch, pour k assez grand :

$$(9) \quad N_k + 1 = \chi(L^k) = \int_M e^{kc_1(L)} Td(M) = a_0 k^n + a_1 k^{n-1} + \dots + a_n,$$

avec notamment

$$(10) \quad a_0 = \frac{c_1(L)^n}{n!}, \quad a_1 = \frac{1}{2(n-1)!} c_1(L)^{n-1} c_1(M).$$

Notons $\rho_k(\omega) = \sum_0^{N_k} |s_\alpha|^2$ pour une base orthonormale (s_α) de $H^0(M, L^k)$. Ainsi,

$$\int_M \rho_k(\omega) d\mu_\omega = h^0(M, L^k) = a_0 k^n + a_1 k^{n-1} + \dots,$$

avec les coefficients a_i s’exprimant en termes d’intégrales impliquant la courbure de ω . L’étude du noyau $\rho_k(\omega)$ peut donc être vue comme une géométrisation du théorème de Riemann-Roch. Les résultats établis par Tian [57], Catlin [11], Zelditch [69], Ruan [53] et Lu [34] en donnent un développement asymptotique précis quand k tend vers l’infini, voir aussi l’article récent [14].

THÉORÈME 2.6. — *La forme ω étant fixée, on a quand k tend vers l’infini un développement asymptotique*

$$\rho_k(\omega) = A_0(\omega)k^n + A_1(\omega)k^{n-1} + \dots,$$

où les $A_i(\omega)$ sont des fonctions sur M , localement définies par ω . En particulier, on a $A_0 = 1$ et $A_1(\omega) = \frac{s_\omega}{8\pi}$.

Le développement s’entend au sens suivant : pour tous $r, N \geq 0$, on a

$$\left\| \rho_k(\omega) - \sum_0^N A_i(\omega)k^{n-i} \right\|_{C^r(M)} \leq K_{r,N,\omega} k^{n-N-1}.$$

De plus, à r et N fixés, la constante $K_{r,N,\omega}$ peut être prise uniforme par rapport à un ensemble de métriques borné dans C^s , où s dépend de r et N .

Expliquons rapidement les raisons de l’existence d’un tel développement (Catlin, Zelditch) : le domaine $D = \{|z| \leq 1\} \subset L^*$ est strictement pseudoconvexe. Une section holomorphe s de L^k donne une fonction holomorphe \hat{s} sur D , et donc holomorphe CR sur ∂D , par $\hat{s}(x) = \langle x^k, s(p(x)) \rangle$, où p est la projection $L^k \rightarrow M$. En outre, ce procédé fournit toutes les fonctions holomorphes sur ∂D de poids k par rapport à l’action du cercle. On peut exprimer cela de la façon suivante : l’espace de Hardy $H^2(\partial D)$ des fonctions holomorphes L^2 sur le bord ∂D admet une décomposition de Fourier $H^2(\partial D) = \sum_k H_k^2(\partial D)$, avec une identification $H_k^2(\partial D) = H^0(M, L^k)$. Le projecteur de Szegő $L^2(\partial D) \rightarrow H^2(\partial D)$ admet un noyau $\pi(x, y)$, et la fonction $\rho_k(\omega)$ peut être vue comme le k -ième coefficient de Fourier de $\pi(x, x)$. L’existence du développement asymptotique provient alors de résultats généraux de Boutet de Monvel et Sjöstrand sur le noyau de Szegő [43]. Le calcul précis des premiers coefficients est fait par Lu.

2.3. Solution formelle

Partons d'une métrique kählérienne ω_0 dans la classe de Kähler $c_1(L)$. On souhaite construire une métrique équilibrée pour le plongement de M dans $PH^0(M, L^k)^*$. Posons $q = \frac{1}{k}$, $\chi(q) = \chi(L^k)$ et $B(q, \omega) = q^n \rho_k(\omega)$; on a donc par le théorème 2.6 un développement asymptotique

$$(11) \quad B(q, \omega) = 1 + \frac{s_\omega}{8\pi}q + f_2q^2 + \dots$$

avec

$$\frac{1}{V} \int_M B(q, \omega) d\mu_\omega = \frac{q^n n!}{c_1(L)^n} \chi(q) = 1 + \frac{n}{2} \frac{c_1(L)^{n-1} c_1(M)}{c_1(L)^n} q + \dots$$

Il est utile de poser⁽³⁾

$$(12) \quad \begin{aligned} \tilde{B}(q, \omega) &= \frac{1}{q} (B(q, \omega) - \frac{q^n}{V} \chi(q)) \\ &= \frac{s_\omega}{8\pi} - \frac{n}{2} \frac{c_1(L)^{n-1} c_1(M)}{c_1(L)^n} + O(q) \\ &= \frac{1}{8\pi} (s_\omega - \bar{s}_\omega) + O(q) \end{aligned}$$

avec : $\tilde{B}(q, \omega) = 0$ si et seulement si ω est équilibrée pour (M, L^k) .

Supposons à présent la courbure scalaire constante : $s_{\omega_0} = \bar{s}_{\omega_0}$. Considérons une variation infinitésimale $\dot{\omega} = dd^C \varphi$. En $q = 0$, on peut écrire d'après (5)

$$\frac{\partial \tilde{B}}{\partial \omega}(\varphi) = -\frac{1}{8\pi} \mathcal{D}^* \mathcal{D} \varphi.$$

Par le lemme 1.1, si $\text{Aut}(M, L)$ est discret, l'opérateur $\mathcal{D}^* \mathcal{D}$ n'a pas de noyau; autoadjoint, il est donc inversible. À partir du développement (11), le théorème des fonctions implicites fabrique alors une solution

$$\omega(q) = \omega + dd^C(\eta_1 q + \eta_2 q^2 + \dots)$$

au problème $\tilde{B}(q, \omega(q)) = 0$. Bien entendu, la variable q ne prend qu'un ensemble discret de valeurs, et la solution est à entendre au sens des séries formelles. En tronquant à l'ordre A , on obtient une solution approchée $\omega_A(q)$ telle que

$$(13) \quad \|\tilde{B}(q, \omega_A(q))\|_{C^{r+2}} \leq C_A q^{A+1}.$$

Digression : démonstration du second point du théorème 2.5. — Si on a pour les $q = 1/k$ une suite de métriques équilibrées ω_q qui converge vers une limite ω_0 , alors, par (12),

$$0 = \tilde{B}(q, \omega_q) = \frac{1}{8\pi} (s_{\omega_q} - \bar{s}) + O(q)$$

avec, d'après le théorème 2.6, le $O(q)$ uniforme par rapport à un ensemble borné de métriques; en prenant la limite quand q tend vers 0, on obtient $s_{\omega_0} = \bar{s}$. \square

⁽³⁾Je tiens cette présentation du problème de T. Mabuchi.

2.4. Point de vue dual

Après la construction d’une solution approchée $\omega_A(q)$, l’idée est de changer de point de vue pour regarder le problème, non plus du point de vue d’une forme de Kähler vérifiant $\tilde{B}(q, \omega) = 0$, mais d’une métrique équilibrée sur $H^0(M, L^k)$ (rappelons $q = 1/k$). L’avantage est de passer d’un problème en dimension infinie à un autre en dimension finie, qui sera résolu par la méthode suivante.

2.4.1. Un résultat général. — Soit un groupe compact G agissant de manière hamiltonienne sur une variété kählérienne \mathcal{X} , avec stabilisateurs discrets. Au point x , l’action infinitésimale $\xi \rightarrow v_\xi(x)$ fournit une application injective $\sigma_x : \mathfrak{g} \rightarrow T_x \mathcal{X}$. L’opérateur $Q_x = \sigma_x^* \sigma_x$ sur \mathfrak{g} est inversible, et on notera

$$(14) \quad \Lambda_x = \|Q_x^{-1}\|_{op},$$

où l’on a choisi un produit scalaire invariant sur \mathfrak{g} . Autrement dit, l’inégalité $\Lambda_x \leq \lambda$ est équivalente à

$$(15) \quad |\xi|^2 \leq \lambda |v_\xi(x)|^2$$

pour tout $\xi \in \mathfrak{g}$.

LEMME 2.7. — *Dans cette situation, si $x_0 \in \mathcal{X}$, et s’il existe deux nombres λ et δ tels que*

1. $\lambda |\mu(x_0)| < \delta$,
2. $\Lambda_{e^{i\xi}x_0} \leq \lambda$ dès que $|\xi| \leq \delta$,

alors il existe $\eta \in \mathfrak{g}$ avec $|\eta| \leq \delta$ et $\mu(e^{i\eta}x_0) = 0$.

Le lemme, sans difficulté, est démontré en suivant le flot du gradient de la fonction $|\mu|^2$ à partir de x_0 [18, proposition 17].

2.4.2. Estimation de la seconde forme fondamentale. — Repartons de la solution tronquée $\omega_A(q)$. L’estimation (13) indique qu’on a pour chaque $k = \frac{1}{q}$ une métrique presque équilibrée. En vue du lemme précédent, on peut fabriquer à partir de cette solution approchée une solution exacte à condition de contrôler de manière uniforme par rapport à k les normes concernées. C’est là le point le plus technique de [18], que nous détaillons maintenant.

Pour pouvoir travailler de manière uniforme, il faut fixer des métriques de référence pour chaque k . On fixe la métrique kählérienne à courbure scalaire constante ω_0 sur M , et pour chaque k la métrique $\tilde{\omega}_0 = k\omega_0$.

La métrique de référence fixée, on dit qu’une métrique $\tilde{\omega}$ est à géométrie R -bornée, si

$$\tilde{\omega} > \frac{1}{R} \tilde{\omega}_0, \\ \|\tilde{\omega} - \tilde{\omega}_0\|_{C^r(\tilde{\omega}_0)} < R.$$

Cette définition donne la classe de métriques sur laquelle les estimations seront faites pour chaque k .

On dira de plus qu'une base $s = (s_\alpha)$ de $H^0(M, L^k)$ est à géométrie R -bornée, si la métrique de Fubini-Study qu'elle induit sur M est à géométrie R -bornée.

Le cœur de la démonstration consiste maintenant en une estimation uniforme de la constante Λ définie en 2.4.1, pour le problème des métriques équilibrées expliqué dans la section 2.1.1, dont nous reprenons à présent les notations. Nous suivrons ici, non pas l'approche initiale de Donaldson, mais celle de Phong et Sturm [48, théorème 2] :

THÉORÈME 2.8. — *Si $\text{Aut}(M, L)$ est discret, alors pour tout $R > 1$ il existe C et $\varepsilon < \frac{1}{10}$ tels que, si une base $s = (s_\alpha)$ de $H^0(M, L^k)$ est à géométrie R -bornée, et $\|\mu_{SV}(s)\|_{op} < \varepsilon$, alors*

$$\Lambda_s \leq Ck^2.$$

Démonstration. — Vu le lemme 2.3 et (15), l'estimation à montrer est simplement, pour tout $\xi \in \mathfrak{su}(N_k + 1)$,

$$(16) \quad |\xi|^2 \leq Ck^2 \int_M |\pi_{\mathcal{N}}(v_\xi)|^2 d\mu_{\tilde{\omega}}.$$

Comme dans la démonstration du lemme 2.3, diagonalisons dans une base $s = (s_\alpha)$, donc $i\xi = \text{diag}(\lambda_\alpha)$, et définissons la fonction réelle f sur M par

$$f = \sum \lambda_\alpha |s_\alpha|^2.$$

L'inégalité de Poincaré sur M ,

$$\int_M |f|^2 d\mu_\omega \leq c \left(\int_M |\bar{\partial}f|^2 d\mu_\omega + \left(\int_M f d\mu_\omega \right)^2 \right),$$

devient après le changement d'échelle $\tilde{\omega} = k\omega$,

$$(17) \quad \int_M |f|^2 d\mu_{\tilde{\omega}} \leq c \left(k \int_M |\bar{\partial}f|^2 d\mu_{\tilde{\omega}} + k^{-n} \left(\int_M f d\mu_{\tilde{\omega}} \right)^2 \right).$$

Rappelons (8), à savoir $\pi_{TM}(v_\xi) = \sharp \bar{\partial}f$, donc

$$\int_M |f|^2 d\mu_{\tilde{\omega}} \leq c \left(k \int_M |\pi_{TM}(v_\xi)|^2 + k^{-n} \left(\int_M f d\mu_{\tilde{\omega}} \right)^2 \right).$$

D'un autre côté,

$$\begin{aligned} \int_M |v_\xi|^2 &= \int_M \left(\sum \lambda_\alpha^2 |s_\alpha|^2 - \left(\sum \lambda_\alpha |s_\alpha|^2 \right)^2 \right) d\mu_{\tilde{\omega}} \\ &\geq \int_M \left(\sum \lambda_\alpha^2 |s_\alpha|^2 \right) d\mu_{\tilde{\omega}} - c \left(k \int_M |\pi_{TM}(v_\xi)|^2 + k^{-n} \left(\int_M f d\mu_{\tilde{\omega}} \right)^2 \right). \end{aligned}$$

Il en résulte

$$(18) \quad (ck + 1) \int_M |v_\xi|^2 \geq \int_M \left(\sum \lambda_\alpha^2 |s_\alpha|^2 \right) d\mu_{\tilde{\omega}} - ck^{-n} \left(\int_M \sum \lambda_\alpha |s_\alpha|^2 d\mu_{\tilde{\omega}} \right)^2$$

$$(19) \quad \geq c' \sum \lambda_\alpha^2$$

si k est assez grand et la métrique suffisamment équilibrée (alors les $\int_M |s_\alpha|^2 d\mu_{\tilde{\omega}}$ sont proches de 1).

Il est clair que le théorème est alors une conséquence de (19) et de l'estimation :

$$(20) \quad \int_M |\pi_{TM} v_\xi|^2 d\mu_{\tilde{\omega}} \leq ck \int_M |\pi_{\mathcal{N}} v_\xi|^2 d\mu_{\tilde{\omega}},$$

que nous établissons maintenant. Puisque $\text{Aut}(M, L)$ est discret, on a une inégalité de Poincaré, qui, après changement d'échelle comme dans (17), s'écrit

$$(21) \quad \int_M |\pi_{TM} v_\xi|^2 d\mu_{\tilde{\omega}} \leq ck \int_M |\bar{\partial} \pi_{TM} v_\xi|^2 d\mu_{\tilde{\omega}}.$$

Comme v_ξ est holomorphe, on a

$$(22) \quad \bar{\partial} \pi_{TM} v_\xi = -\bar{\partial} \pi_{\mathcal{N}} v_\xi = -\alpha \pi_{\mathcal{N}} v_\xi,$$

où $\alpha \in \Omega^{0,1}(\text{Hom}(\mathcal{N}, TM))$ est la seconde forme fondamentale de M dans P^{N_k} , définie par l'opérateur $\bar{\partial}$ du fibré $TP^{N_k} = TM \oplus \mathcal{N}$ le long de M :

$$\bar{\partial}_{TP^{N_k}} = \begin{pmatrix} \bar{\partial}_{TM} & \alpha \\ 0 & \bar{\partial}_{\mathcal{N}} \end{pmatrix}.$$

L'estimation souhaitée (20) est alors une conséquence immédiate de (21), (22) et du fait suivant.

FAIT 2.9. — *Sous les hypothèses du théorème, la seconde forme fondamentale α de M dans TP^{N_k} est bornée par une constante ne dépendant que de R .*

La démonstration du fait vient de la formule bien connue suivant laquelle « la courbure décroît dans les sous-fibrés » :

$$F_{TP^{N_k}}|_{TM} - F_{TM} = -\alpha \wedge \alpha^*,$$

d'où on déduit

$$|\alpha|^2 \leq |F_{TP^{N_k}}| + |F_{TM}|.$$

La courbure de TP^{N_k} est bornée (courbure de la métrique de Fubini-Study), comme l'est celle de TM grâce à l'hypothèse que $\tilde{\omega}$ reste à géométrie R -bornée. \square

2.4.3. Fin de la démonstration du théorème 2.5. — Rassemblons maintenant brièvement les éléments de la démonstration, en omettant cependant la fastidieuse vérification de l'uniformité convenable des estimations que nous écrirons. La solution tronquée à un ordre A fixé à l'avance, $\omega_A(q)$, est presque équilibrée : avec un peu de travail, pour une base s de $H^0(M, L^k)$ définie comme en (7), l'estimation (13) donne

$$\|\mu_{SU}(s)\|_{op} = O(q^{A+2}).$$

La norme figurant dans les hypothèses du lemme 2.7 lui est liée par

$$|\mu_{SU}(s)| \leq \sqrt{N_k + 1} \|\mu_{SU}(s)\|_{op},$$

avec $N_k = O(q^{-n})$, donc

$$|\mu_{SU}(s)| = O(q^{A+2-\frac{n}{2}}).$$

Par le théorème 2.8, on peut choisir la constante λ dans le lemme 2.7 en $O(q^{-2})$, et donc

$$\lambda|\mu_{SU}(s)| = O(q^{A-\frac{n}{2}}).$$

Si on a choisi initialement $A > \frac{n}{2}$, le lemme 2.7 fournit pour q assez petit une solution s_q à $\mu_{SU}(s_q) = 0$ et $|s - s_q| = O(q^{A-\frac{n}{2}})$. Revenant aux formes de Kähler, la métrique équilibrée $\tilde{\omega}_q$ satisfait

$$\|\tilde{\omega}_q - \tilde{\omega}_0\|_{C^r(\tilde{\omega}_0)} = O(q^{A-\frac{n}{2}}),$$

et par changement d'échelle :

$$\|\omega_q - \omega_0\|_{C^r(\omega_0)} = O(q^{A-r-\frac{n}{2}}).$$

Si on a choisi au début $A > r + \frac{n}{2}$, on déduit la convergence dans C^r des métriques équilibrées ω_q vers ω_0 . Par unicité des métriques équilibrées, la convergence est valable dans C^∞ . \square

3. STABILITÉ ET CONJECTURES

Il existe plusieurs notions de stabilité des variétés algébriques. L'idée générale est de considérer la stabilité d'une variété dans P^N comme la stabilité par rapport à l'action de $SL(N+1)$ sur le schéma de Hilbert des variétés de même polynôme de Hilbert. Chaque choix d'un plongement projectif de celui-ci mène à une notion différente de stabilité : Hilbert-Mumford, Chow-Mumford, K-stabilité. Nous nous concentrerons ici sur les deux dernières.

3.1. Stabilité de Chow-Mumford

3.1.1. Stabilité et métriques équilibrées. — Soit une variété projective $M^n \subset P^N = PE^*$, de degré d . Alors il existe un point $\hat{M} \in W = (\text{Sym}^d E)^{\otimes(n+1)}$, tel que $[\hat{M}] \in PW$ soit le point de Chow du cycle $M \subset P^N$, voir la construction dans [45, 1.16], qui généralise les deux cas simples suivants :

- si M est une hypersurface, \hat{M} est l'équation de M ;
- si M est linéaire, \hat{M} est obtenu par les coordonnées de Plücker.

Le groupe $SL(N+1) = SL(E)$ agit sur W .

DÉFINITION 3.1 (Mumford). — *La variété $M^n \subset P^N$ est stable (au sens de Chow) si $SL(N+1)\hat{M}$ est fermée dans W , et le stabilisateur de \hat{M} est fini ; elle est semistable si 0 n'est pas dans l'adhérence de $SL(N+1)\hat{M}$.*

La variété polarisée (M, L) est asymptotiquement stable si pour k assez grand, l'image $M_k \subset P^{N_k} = PH^0(M, L^k)^$ de M par les plongements de Kodaira est stable.*

Par exemple, les courbes asymptotiquement stables sont exactement les courbes stables au sens de Deligne-Mumford.

Comme indiqué section 1.2.1, on peut définir aussi la polystabilité en supprimant l'hypothèse de finitude du stabilisateur. C'est la polystabilité qui a une interprétation en termes de géométrie différentielle—les métriques équilibrées de la section 2.1 :

THÉORÈME 3.2 (Zhang [70]). — *La variété (M, L) est équilibrée si et seulement si le plongement $M \hookrightarrow PH^0(M, L)^*$ est polystable.*

Zhang énonce seulement stable \Rightarrow équilibré \Rightarrow semistable, mais son argument montre en réalité le théorème écrit ci-dessus, voir aussi [47]. La stabilité est équivalente à l'existence d'une *unique* métrique équilibrée. La fonctionnelle Z de la section 2.1.2, définie différemment, joue un rôle crucial dans la démonstration.

Joint au théorème 2.5, l'énoncé précédent implique immédiatement la partie relative à la stabilité du théorème 0.1 :

COROLLAIRE 3.3. — *Si (M, L) admet dans la classe $c_1(L)$ une métrique kählérienne à courbure scalaire constante, et $\text{Aut}(M, L)$ est discret, alors (M, L) est asymptotiquement stable.*

Comme indiqué dans l'introduction, tester la stabilité d'une variété algébrique est très difficile, et peu de résultats sont connus (Mumford, Gieseker [26], Viehweg [65]). Un corollaire du résultat est par exemple que toutes les variétés à $c_1 < 0$ sont asymptotiquement stables.

3.1.2. Stabilité relative. — Le cas où le groupe $\text{Aut}^0(M, L)$ n'est pas trivial est étudié par Mabuchi : en réalité, si $\text{Aut}^0(M, L)$ est semi-simple, les résultats demeurent identiques (il suffit de se restreindre aux fonctions invariantes sous le groupe d'isométries de la métrique kählérienne à courbure scalaire constante). Le cas délicat provient de la présence d'un centre Z dans $\text{Aut}^0(M, L)$. Plus généralement, si un tore complexe T agit, les espaces $E_k = H^0(M, L^k)$ se décomposent sous l'action de T ,

$$E_k = \bigoplus_1^{\nu_k} E_k^{\chi_j},$$

où les χ_j sont des caractères de T . On définit alors le groupe

$$G_k = \times_1^{\nu_k} SL(E_k^{\chi_j}).$$

On regarde l'image M_k de M dans PE_k^* : la *stabilité relative* à T de (M, L^k) est définie comme dans la définition 3.1, en regardant l'action de G_k sur \hat{M}_k au lieu de celle de $SL(E_k)$.

THÉORÈME 3.4 (Mabuchi [41]). — *Si (M, L) admet une métrique extrémale, alors (M, L) est asymptotiquement stable relativement au centre de $\text{Aut}^0(M, L)$.*

Le principe de la démonstration —la méthode de quantification—, reste le même, mais Mabuchi montre directement la stabilité à partir des solutions approchées, sans recourir à la construction de métriques équilibrées. Cependant, il montre aussi que la stabilité relativement à un tore complexe T est équivalente à l'existence de « métriques équilibrées relativement à T », voir les détails dans [38].

L'unicité de la métrique extrémale dans $c_1(L)$, modulo $\text{Aut}^0(M, L)$, est déduite dans [39]. Enfin, il semble qu'il y ait des obstructions à la stabilité asymptotique (absolue) de (M, L) , pourvue d'une métrique extrémale, en présence d'un centre de $\text{Aut}^0(M, L)$, voir [40].

3.2. K-stabilité

3.2.1. Configuration test. — Le critère de Hilbert-Mumford dit que la stabilité au sens de la théorie géométrique des invariants peut être testée sur les sous-groupes à un paramètre : l'orbite $G^{\mathbb{C}}x$ est fermée si tel est le cas pour tous les sous-groupes à un paramètre $\mathbb{C}^* \subset G^{\mathbb{C}}$. Pour un tel sous-groupe, on a une limite

$$x_0 = \lim_{\lambda \rightarrow 0} \lambda \cdot x$$

qui est un point fixe de l'action de \mathbb{C}^* ; la question de la stabilité se réduit alors à la question de l'action de \mathbb{C}^* sur la droite au-dessus de x_0 , qui a un poids $\rho \in \mathbb{Z}$ (action par λ^ρ) : si le poids ρ est toujours strictement négatif, alors x est stable (semistable s'il est toujours négatif ou nul).

L'idée de sous-groupe à un paramètre de l'action de $SL(N+1)$ sur le schéma de Hilbert mène à la définition suivante [20].

DÉFINITION 3.5. — *Une configuration test d'exposant r pour la variété polarisée (M, L) est la donnée de :*

1. *un schéma \mathcal{M} avec une action de \mathbb{C}^* ,*
2. *un fibré en droites ample $\mathcal{L} \rightarrow \mathcal{M}$, équivariant pour l'action de \mathbb{C}^* ,*
3. *une application plate, propre, \mathbb{C}^* -équivariante, $\pi : \mathcal{M} \rightarrow \mathbb{C}$, avec \mathbb{C}^* agissant sur \mathbb{C} de la manière standard,*

telle que les fibres $(\mathcal{M}_t, \mathcal{L}_t)$ soient isomorphes à (M, L^r) pour $t \neq 0$.

Une telle configuration est dite produit si $\mathcal{M} \simeq M \times \mathbb{C}$, et triviale si en outre \mathbb{C}^* agit seulement sur le facteur \mathbb{C} .

3.2.2. Invariant de Futaki et K-stabilité. — Si \mathbb{C}^* agit sur un schéma projectif X muni d'un fibré ample L , alors pour tout k , il agit aussi sur l'espace vectoriel

$$E_k = H^0(X, L^k).$$

Soit d_k sa dimension et w_k le poids de l'action induite sur $\Lambda^{d_k} E_k$. Pour k assez grand, d_k et w_k peuvent être calculés par le théorème de Riemann-Roch (équivariant), et sont

donc des polynômes en k , de degrés respectifs n et $n + 1$. La fonction $F(k) = \frac{w_k}{kd_k}$ a un développement asymptotique

$$F(k) = F_0 + \frac{F_1}{k} + \frac{F_2}{k^2} + \dots$$

avec les coefficients F_i rationnels. Le coefficient F_1 est appelé l'*invariant de Futaki* de l'action de \mathbb{C}^* sur (X, L) .

Dans le cas d'une variété lisse X , l'action infinitésimale de \mathbb{C}^* induit un champ de vecteurs holomorphe v , et l'invariant de Futaki F_1 se spécialise bien à l'invariant de Futaki \mathcal{F} introduit dans la section 1.3.4, à une constante près : $F_1 = -\frac{1}{V}\mathcal{F}(v)$, voir [20, proposition 2.2.2].

On définit la K-stabilité, introduite par Tian [60] dans son étude des variétés de Fano, puis adaptée par Donaldson [20] pour autoriser une fibre centrale non normale.

DÉFINITION 3.6. — *La variété polarisée (M, L) est K-stable (resp. K-semistable) si, pour toute configuration test non triviale $(\mathcal{M}, \mathcal{L})$, l'invariant de Futaki de l'action de \mathbb{C}^* induite sur $(\mathcal{M}_0, \mathcal{L}_0)$ est strictement négatif (resp. négatif ou nul). Elle est K-polystable si elle est semistable, et si l'invariant de Futaki ne s'annule que pour les configurations produits.*

À nouveau, il y a des variations dans la terminologie : ce que nous avons appelé polystabilité est appelé par Tian semistabilité propre, et par Donaldson stabilité ; notre terminologie, plus cohérente avec celle des fibrés vectoriels, suit [51].

Enfin, il n'est pas clair que la K-stabilité soit une bonne notion de stabilité au sens de la théorie géométrique des invariants, voir cependant [46].

3.2.3. La conjecture. — Dans la correspondance de Hitchin-Kobayashi (section 1.2.2), l'existence d'une métrique de Hermite-Einstein sur un fibré holomorphe est équivalente à la condition algébrique de polystabilité du fibré. On attend de même une correspondance de Hitchin-Kobayashi pour les variétés, dans laquelle le rôle de la condition algébrique serait joué par la K-stabilité (Tian [63], Donaldson [20]) :

CONJECTURE 3.7. — *La variété polarisée (M, L) admet une métrique kählérienne à courbure scalaire constante si et seulement si elle K-polystable.*

Dans la correspondance de Hitchin-Kobayashi, la direction Hermite-Einstein \Rightarrow stabilité est facile, et très difficile la réciproque. Sur les variétés, même la direction « facile » n'est pas connue : en effet, comme on a vu, l'existence d'une métrique kählérienne à courbure scalaire constante sur une variété polarisée (M, L) implique bien la stabilité asymptotique de Chow-Mumford, mais on a seulement l'implication

asymptotiquement stable \Rightarrow asymptotiquement semistable \Rightarrow K-semistable,

voir [52] pour une discussion précise des relations entre les différentes notions de stabilité, qui s'interprètent toutes en des conditions sur les $F(k)$, convenablement normalisés, associés aux configurations tests ; l'implication manquante sur la gauche, K-stabilité \Rightarrow stabilité asymptotique, est tentante, mais n'est pas démontrée pour le moment.

Ainsi donc, *une variété portant une métrique kählérienne à courbure scalaire constante est K-semistable*. Ce fait peut aussi être vu comme conséquence [13, 46] d’une borne inférieure sur la K-énergie :

THÉORÈME 3.8. — *Supposons que M admette une métrique kählérienne à courbure scalaire constante. Alors la K-énergie de Mabuchi y atteint un minimum.*

Si on dispose de géodésiques dans l’espace des métriques de Kähler \mathcal{K} , le théorème est une conséquence immédiate de la convexité de la K-énergie ; c’est la voie empruntée par Chen et Tian [13]. Dans le cas polarisé, et groupe d’automorphismes discret, la méthode de quantification est utilisée par Donaldson [22] pour une autre démonstration : la solution du problème des métriques équilibrées donne une borne inférieure sur la fonctionnelle Z de la section 2.1.2, qui, après passage au problème dual, fournit des bornes inférieures sur des fonctionnelles convergeant vers la K-énergie.

Une borne inférieure sur la K-énergie peut être pensée comme une condition analytique de semistabilité. Plus généralement, Tian [62, 63] conjecture que l’existence d’une métrique kählérienne à courbure scalaire constante est équivalente à une *stabilité analytique* définie comme la propriété (en un sens précis) de la K-énergie. Il a montré cette conjecture dans le cas Kähler-Einstein Fano [60]. Pour les liens entre propriété de la K-énergie et K-stabilité, voir [59, 61, 46].

Un autre développement concernant la stabilité est la définition par Ross et Thomas [51] d’une notion de pente, μ , pour les sous-schémas Z d’une variété projective polarisée M , analogue à la pente des sous-faisceaux d’un fibré vectoriel. Comme pour les fibrés, la semistabilité est définie en exigeant que $\mu(Z) \leq \mu(X)$ pour tout $Z \subset X$. Le lien avec la K-stabilité est obtenu en considérant la configuration provenant de la déformation au cône normal, de sorte que la semistabilité pour la pente est une conséquence de la K-semistabilité. Comme corollaire, Ross [50] construit sur P^2 éclaté en quatre points des classes de Kähler n’admettant pas de métriques kählériennes à courbure scalaire constante ; l’obstruction ne provient pas du groupe d’automorphismes, trivial. On consultera [51] pour d’autres exemples.

3.3. Le cas torique

L’image de la conjecture 3.7 devient un peu plus nette dans le cas torique, grâce au travail fondateur de Donaldson [20]. Nous donnons ici un bref aperçu des résultats.

Soit (M^{2n}, ω) une variété kählérienne torique. Le tore T^n agit avec application moment $\mu : M \rightarrow \mathbb{R}^n$, et l’image de μ est un polytope \bar{P} dans \mathbb{R}^n , le polytope de Delzant de M . L’action est libre sur l’image réciproque de l’intérieur, $M_0 = \mu^{-1}(P) \subset M$. On utilise alors les variables action-angle (x^i, θ_i) telles que

$$\omega = \sum dx^i \wedge d\theta_i,$$

avec l’application moment donnée par la projection sur les coordonnées x^i , et l’action de T^n par la translation dans les coordonnées θ_i .

Le problème se pose de manière élégante dans un formalisme proche de celui de la section 1.3.1, en fixant la forme symplectique ω et en faisant varier la structure complexe. Les métriques T^n -invariantes, kählériennes avec forme de Kähler ω , ont la description suivante, due à Guillemin : une telle métrique, g , est décrite par un *potentiel symplectique* u , fonction convexe sur P , de sorte que

$$g = \sum u_{,ij} dx^i dx^j + u^{ij} d\theta_i d\theta_j,$$

où (u^{ij}) est l'inverse de la métrique hessienne $(u_{,ij})$ de u .

On a alors la belle formule d'Abreu [1] pour la courbure scalaire :

$$s(u) = - \sum \frac{\partial^2 u^{ij}}{\partial x^i \partial x^j}.$$

Le problème des métriques kählériennes à courbure scalaire constante apparaît donc comme un problème d'équations aux dérivées partielles, non linéaire, d'ordre 4 sur le potentiel symplectique.

Guan [27] a montré que les géodésiques de l'espace des formes de Kähler dans la classe $[\omega]$ deviennent des droites dans l'espace des potentiels symplectiques. Il en déduit l'existence de géodésiques reliant deux métriques données, et donc l'unicité, à automorphisme holomorphe près, des métriques extrémales toriques.

Donaldson [20] étudie les surfaces toriques K-polystables. Il montre que :

1. la K-énergie de Mabuchi est bornée inférieurement ;
2. une suite minimisante de potentiels symplectiques a toujours une sous-suite convergente en un sens faible.

L'idée est d'utiliser la K-stabilité en l'appliquant à des configurations tests construites à partir de fonctions convexes, rationnelles, linéaires par morceau sur P . Un corollaire important est la construction de nouvelles surfaces toriques n'admettant pas de métrique kählérienne à courbure scalaire constante.

Si la limite faible obtenue était lisse et minimisait la K-énergie, elle fournirait la métrique kählérienne à courbure scalaire constante souhaitée : justifier cet espoir est manifestement un problème d'analyse très délicat ; voir cependant les estimations de [21].

RÉFÉRENCES

- [1] M. ABREU – « Kähler geometry of toric varieties and extremal metrics », *Internat. J. Math.* **9** (1998), no. 6, p. 641–651.
- [2] T. AUBIN – « Équations du type Monge-Ampère sur les variétés kählériennes compactes », *C. R. Acad. Sci. Paris Sér. A-B* **283** (1976), no. 3, p. Aiii, A119–A121.
- [3] S. BANDO & T. MABUCHI – « Uniqueness of Einstein Kähler metrics modulo connected group actions », *Algebraic geometry, Sendai, 1985, Adv. Stud. Pure Math.*, vol. 10, North-Holland, Amsterdam, 1987, p. 11–40.

- [4] A.L. BESSE – *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 10, Springer-Verlag, Berlin, 1987.
- [5] J.-P. BOURGUIGNON – « Métriques d'Einstein-Kähler sur les variétés de Fano: obstructions et existence (d'après Y. Matsushima, A. Futaki, S. T. Yau, A. Nadel et G. Tian) », *Astérisque* **245** (1997), p. 277–305, Séminaire Bourbaki, Vol. 1996/97, Exp. N° 830.
- [6] D. BURNS & P. DE BARTOLOMEIS – « Stability of vector bundles and extremal metrics », *Invent. Math.* **92** (1988), no. 2, p. 403–407.
- [7] E. CALABI – « Extremal Kähler metrics », Seminar on Differential Geometry, Ann. of Math. Stud., vol. 102, Princeton Univ. Press, Princeton, N.J., 1982, p. 259–290.
- [8] ———, « Extremal Kähler metrics. II », *Differential geometry and complex analysis*, Springer, Berlin, 1985, p. 95–114.
- [9] E. CALABI & X.X. CHEN – « The space of Kähler metrics. II », *J. Differential Geom.* **61** (2002), no. 2, p. 173–193.
- [10] D.M.J. CALDERBANK & M.A. SINGER – « Einstein metrics and complex singularities », *Invent. Math.* **156** (2004), no. 2, p. 405–443.
- [11] D. CATLIN – « The Bergman kernel and a theorem of Tian », *Analysis and geometry in several complex variables (Katata, 1997)*, Trends Math., Birkhäuser Boston, Boston, MA, 1999, p. 1–23.
- [12] X.X. CHEN – « The space of Kähler metrics », *J. Differential Geom.* **56** (2000), no. 2, p. 189–234.
- [13] X.X. CHEN & G. TIAN – « Geometry of Kähler metrics and holomorphic foliation by discs », [math.DG/0409433](https://arxiv.org/abs/math/0409433).
- [14] X. DAI, K. LIU & X. MA – « On the asymptotic expansion of Bergman kernel », [math.DG/0404494](https://arxiv.org/abs/math/0404494).
- [15] S.K. DONALDSON – « Infinite determinants, stable bundles and curvature », *Duke Math. J.* **54** (1987), no. 1, p. 231–247.
- [16] ———, « Remarks on gauge theory, complex geometry and 4-manifold topology », *Fields Medallists' lectures*, World Sci. Ser. 20th Century Math., vol. 5, World Sci. Publishing, River Edge, NJ, 1997, p. 384–403.
- [17] ———, « Symmetric spaces, Kähler geometry and Hamiltonian dynamics », *Northern California Symplectic Geometry Seminar*, Amer. Math. Soc. Transl. Ser. 2, vol. 196, Amer. Math. Soc., Providence, RI, 1999, p. 13–33.
- [18] ———, « Scalar curvature and projective embeddings. I », *J. Differential Geom.* **59** (2001), no. 3, p. 479–522.
- [19] ———, « Holomorphic discs and the complex Monge-Ampère equation », *J. Symplectic Geom.* **1** (2002), no. 2, p. 171–196.
- [20] ———, « Scalar curvature and stability of toric varieties », *J. Differential Geom.* **62** (2002), no. 2, p. 289–349.

- [21] ———, « Interior estimates for solutions of Abreu’s equation », *math.DG/0407486*.
- [22] ———, « Scalar curvature and projective embeddings, II », *math.DG/0407534*.
- [23] S.K. DONALDSON & P.B. KRONHEIMER – *The geometry of four-manifolds*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1990, Oxford Science Publications.
- [24] A. FUTAKI – « An obstruction to the existence of Einstein Kähler metrics », *Invent. Math.* **73** (1983), no. 3, p. 437–443.
- [25] P. GAUDUCHON – *Calabi’s extremal Kähler metrics: an elementary introduction*.
- [26] D. GIESEKER – « Global moduli for surfaces of general type », *Invent. Math.* **43** (1977), no. 3, p. 233–282.
- [27] D. GUAN – « On modified Mabuchi functional and Mabuchi moduli space of Kähler metrics on toric bundles », *Math. Res. Lett.* **6** (1999), no. 5-6, p. 547–555.
- [28] G. KEMPF & L. NESS – « The length of vectors in representation spaces », Algebraic geometry (Proc. Summer Meeting, Univ. Copenhagen, Copenhagen, 1978), Lecture Notes in Math., vol. 732, Springer, Berlin, 1979, p. 233–243.
- [29] J. KIM, C. LEBRUN & M. PONTECORVO – « Scalar-flat Kähler surfaces of all genera », *J. Reine Angew. Math.* **486** (1997), p. 69–95.
- [30] C. LEBRUN – « Scalar-flat Kähler metrics on blown-up ruled surfaces. », *J. Reine Angew. Math.* **420** (1991), p. 161–177.
- [31] ———, « Polarized 4-manifolds, extremal Kähler metrics, and Seiberg-Witten theory », *Math. Res. Lett.* **2** (1995), no. 5, p. 653–662.
- [32] C. LEBRUN & M. SINGER – « Existence and deformation theory for scalar-flat Kähler metrics on compact complex surfaces », *Invent. Math.* **112** (1993), no. 2, p. 273–313.
- [33] A. LICHNEROWICZ – « Sur les transformations analytiques des variétés kählériennes compactes », *C. R. Acad. Sci. Paris* **244** (1957), p. 3011–3013.
- [34] Z. LU – « On the lower order terms of the asymptotic expansion of Tian-Yau-Zelditch », *Amer. J. Math.* **122** (2000), no. 2, p. 235–273.
- [35] H. LUO – « Geometric criterion for Gieseker-Mumford stability of polarized manifolds », *J. Differential Geom.* **49** (1998), no. 3, p. 577–599.
- [36] T. MABUCHI – « K -energy maps integrating Futaki invariants », *Tohoku Math. J. (2)* **38** (1986), no. 4, p. 575–593.
- [37] ———, « Some symplectic geometry on compact Kähler manifolds. I », *Osaka J. Math.* **24** (1987), no. 2, p. 227–252.
- [38] ———, « Stability of extremal Kähler metrics », *Osaka J. Math.* **41** (2004), no. 3.
- [39] ———, « Uniqueness of extremal Kähler metrics for an integral Kähler class », *Internat. J. Math.* **15** (2004), no. 6, p. 531–546.
- [40] ———, « An energy-theoretic approach to the Hitchin-Kobayashi correspondence for manifolds, I », *Invent. Math.*, à paraître.

- [41] ———, « An energy-theoretic approach to the Hitchin-Kobayashi correspondence for manifolds, II », *J. Differential Geom.*, à paraître.
- [42] Y. MATSUSHIMA – « Sur la structure du groupe d'homéomorphismes analytiques d'une certaine variété kählérienne », *Nagoya Math. J.* **11** (1957), p. 145–150.
- [43] L. BOUTET DE MONVEL & J. SJÖSTRAND – « Sur la singularité des noyaux de Bergman et de Szegö », Journées: Équations aux Dérivées Partielles de Rennes (1975), Soc. Math. France, Paris, 1976, p. 123–164. Astérisque, No. 34–35.
- [44] D. MUMFORD, J. FOGARTY & F. KIRWAN – *Geometric invariant theory*, third éd., Ergebnisse der Mathematik und ihrer Grenzgebiete (2) [Results in Mathematics and Related Areas (2)], vol. 34, Springer-Verlag, Berlin, 1994.
- [45] D. MUMFORD – « Stability of projective varieties », *Enseignement Math. (2)* **23** (1977), no. 1-2, p. 39–110.
- [46] S.T. PAUL & G. TIAN – « Algebraic and analytic K-stability », [math.DG/0405530](https://arxiv.org/abs/math/0405530).
- [47] D.H. PHONG & J. STURM – « Stability, energy functionals, and Kähler-Einstein metrics », *Comm. Anal. Geom.* **11** (2003), no. 3, p. 565–597.
- [48] ———, « Scalar curvature, moment maps, and the Deligne pairing », *Amer. J. Math.* **126** (2004), no. 3, p. 693–712.
- [49] Y. ROLLIN & M. SINGER – « Non-minimal scalar-flat Kähler surfaces and parabolic stability », [math.DG/0404423](https://arxiv.org/abs/math/0404423), *Invent. Math.*, à paraître.
- [50] J. ROSS – « Instability of polarised algebraic varieties », PhD thesis, Imperial College, 2003.
- [51] J. ROSS & R. THOMAS – « An obstruction to the existence of constant scalar curvature Kähler metrics », [math.DG/0412518](https://arxiv.org/abs/math/0412518).
- [52] ———, « A study of the Hilbert-Mumford criterion for the stability of projective varieties », [math.AG/0412519](https://arxiv.org/abs/math/0412519).
- [53] W.-D. RUAN – « Canonical coordinates and Bergmann metrics », *Comm. Anal. Geom.* **6** (1998), no. 3, p. 589–631.
- [54] S. SEMMES – « Complex Monge-Ampère and symplectic manifolds », *Amer. J. Math.* **114** (1992), no. 3, p. 495–550.
- [55] C.T. SIMPSON – « Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization », *J. Amer. Math. Soc.* **1** (1988), no. 4, p. 867–918.
- [56] G. TIAN – « On Kähler-Einstein metrics on certain Kähler manifolds with $C_1(M) > 0$ », *Invent. Math.* **89** (1987), no. 2, p. 225–246.
- [57] ———, « On a set of polarized Kähler metrics on algebraic manifolds », *J. Differential Geom.* **32** (1990), no. 1, p. 99–130.
- [58] ———, « On Calabi's conjecture for complex surfaces with positive first Chern class », *Invent. Math.* **101** (1990), no. 1, p. 101–172.
- [59] ———, « The K -energy on hypersurfaces and stability », *Comm. Anal. Geom.* **2** (1994), no. 2, p. 239–265.

- [60] ———, « Kähler-Einstein metrics with positive scalar curvature », *Invent. Math.* **130** (1997), no. 1, p. 1–37.
- [61] ———, « Bott-Chern forms and geometric stability », *Discrete Contin. Dynam. Systems* **6** (2000), no. 1, p. 211–220.
- [62] ———, *Canonical metrics in Kähler geometry*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2000, Notes taken by Meike Akveld.
- [63] ———, « Extremal metrics and geometric stability », *Houston J. Math.* **28** (2002), no. 2, p. 411–432, Special issue for S.S. Chern.
- [64] K. UHLENBECK & S.-T. YAU – « On the existence of Hermitian-Yang-Mills connections in stable vector bundles », *Comm. Pure Appl. Math.* **39** (1986), no. S, suppl., p. S257–S293, *Frontiers of the mathematical sciences: 1985* (New York, 1985).
- [65] E. VIEHWEG – *Quasi-projective moduli for polarized manifolds*, *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)* [Results in Mathematics and Related Areas (3)], vol. 30, Springer-Verlag, Berlin, 1995.
- [66] S.T. YAU – « Calabi’s conjecture and some new results in algebraic geometry », *Proc. Nat. Acad. Sci. U.S.A.* **74** (1977), no. 5, p. 1798–1799.
- [67] ———, « On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I », *Comm. Pure Appl. Math.* **31** (1978), no. 3, p. 339–411.
- [68] ———, « Open problems in geometry », *Differential geometry: partial differential equations on manifolds* (Los Angeles, CA, 1990), *Proc. Sympos. Pure Math.*, vol. 54, Amer. Math. Soc., Providence, RI, 1993, p. 1–28.
- [69] S. ZELDITCH – « Szegő kernels and a theorem of Tian », *Internat. Math. Res. Notices* (1998), no. 6, p. 317–331.
- [70] S. ZHANG – « Heights and reductions of semi-stable varieties », *Compositio Math.* **104** (1996), no. 1, p. 77–105.

Olivier BIQUARD

IRMA, CNRS et Université Louis Pasteur,
7 rue René Descartes,
F-67084 STRASBOURG CEDEX
E-mail : biquard@math.u-strasbg.fr

LA MESURE D'ÉQUILIBRE D'UN ENDOMORPHISME DE $\mathbb{P}^k(\mathbb{C})$
[d'après Briend et Duval]

par **Xavier BUFF**

INTRODUCTION

Un endomorphisme de $\mathbb{P}^k(\mathbb{C})$ est une fonction holomorphe non constante de $\mathbb{P}^k(\mathbb{C})$ dans $\mathbb{P}^k(\mathbb{C})$. Nous noterons $z = [z_0 : z_1 : \dots : z_k]$ un point $z \in \mathbb{P}^k(\mathbb{C})$ défini par ses coordonnées homogènes et $\pi : \mathbb{C}^{k+1} \setminus \{0\} \rightarrow \mathbb{P}^k(\mathbb{C})$ la projection canonique. Une application rationnelle $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ de degré algébrique d est une application qui s'écrit $f = [F_0 : F_1 : \dots : F_k]$, les F_j étant des polynômes homogènes de degré d , sans facteurs communs. On associe à f l'application $F = (F_0, F_1, \dots, F_k) : \mathbb{C}^{k+1} \rightarrow \mathbb{C}^{k+1}$. On a alors $\pi \circ F = f \circ \pi$. Une application rationnelle $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ n'est pas nécessairement définie partout. Son ensemble d'indétermination est $\pi(F^{-1}\{0\})$.

Les endomorphismes de $\mathbb{P}^k(\mathbb{C})$ sont les applications rationnelles non constantes dont l'ensemble d'indétermination est vide. Le degré algébrique d'une application rationnelle est aussi le degré de la préimage d'un hyperplan générique.

Dans tout cet exposé, nous nous intéressons aux propriétés dynamiques d'un endomorphisme holomorphe $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ de degré algébrique $d \geq 2$. D'après le théorème de Bezout, son degré topologique, c'est-à-dire le nombre de préimages d'un point générique, est alors égal à d^k .

Remarque 0.1. — Nous ne parlerons pas des applications rationnelles dont l'ensemble d'indétermination n'est pas vide, par exemple les applications de Hénon. Le lecteur intéressé pourra consulter [25].

Le cas de la dimension $k = 1$ a une histoire assez ancienne, avec des contributions de Koenigs, Schröder, Böttcher à la fin du 19ème siècle et les fameux mémoires de Fatou et Julia au début du 20ème siècle. L'ensemble de Fatou Ω_f est le plus grand ouvert sur lequel la famille des itérés de f est normale (pour la convergence uniforme sur tout compact). L'ensemble de Julia J_f est le complémentaire de Ω_f .

Le fait même que les ensembles de Fatou et de Julia soient définis à partir de la notion de famille normale reflète les origines de la dynamique holomorphe. Un ingrédient majeur dans les travaux de Fatou et de Julia est le théorème de Montel qui dit qu'une famille de fonctions holomorphes définies sur un ouvert de $\mathbb{P}^1(\mathbb{C})$ et évitant 3 points de $\mathbb{P}^1(\mathbb{C})$ est une famille normale.

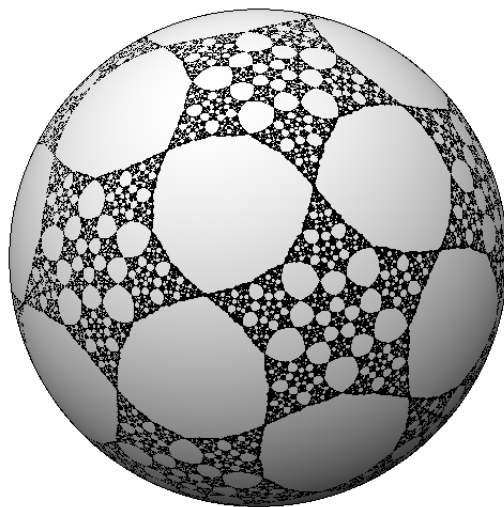


FIGURE 1. L'ensemble de Julia J_f de la fraction rationnelle $f : z \mapsto z \frac{z^{10} + 66z^5 - 11}{-11z^{10} - 66z^5 + 1}$.

L'ensemble J_f est compact, parfait et complètement invariant par f :

$$f(J_f) = J_f \quad \text{et} \quad f^{-1}(J_f) = J_f.$$

Il se peut qu'il y ait un ensemble fini complètement invariant par f . Un tel ensemble contient au plus 2 points. Le plus grand s'appelle l'ensemble exceptionnel E_f . On peut donner des définitions équivalentes de l'ensemble de Julia J_f d'une fraction rationnelle $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$:

- J_f est le plus petit fermé complètement invariant contenant au moins 3 points ;
- J_f est l'adhérence de l'ensemble des points périodiques répulsifs de f (i.e. les points $z \in \mathbb{P}^1(\mathbb{C})$ tels que $f^{ok}(z) = z$ et $|(f^{ok})'(z)| > 1$ pour un entier $k \geq 1$).

En 1965, Brolin [9] démontre que si f est un polynôme, il existe une mesure de probabilité invariante et mélangeante μ_f , supportée par l'ensemble de Julia J_f , qui reflète la distribution des préimages des points hors de l'ensemble exceptionnel E_f . L'outil central de la démonstration est la théorie du potentiel. La mesure μ_f peut être obtenue comme le Laplacien d'une fonction de Green dynamique. Tortrat [26] montre que la mesure μ_f reflète également la distribution des points périodiques.

En 1983, Lyubich [21] et Freire-Lopès-Mañé [16], [22] généralisent ce résultat pour toute fraction rationnelle f . De plus, ils montrent que μ_f est l'unique mesure d'entropie maximale de f .

Que reste-t-il de ces résultats en dimension $k \geq 2$?

Les définitions de l'ensemble de Fatou et de l'ensemble de Julia restent les mêmes. Mais en général, on ne peut plus caractériser l'ensemble de Julia comme le plus petit fermé invariant ou comme l'adhérence de l'ensemble des points périodiques répulsifs.

En 1990, Sibony a l'idée d'introduire la théorie du potentiel dans l'étude de la dynamique des difféomorphismes polynomiaux de \mathbb{C}^2 . Pour un tel difféomorphisme, il

définit une mesure de probabilité invariante. Avec Bedford [1] §3, il établit les premières propriétés de cette mesure. Les résultats sont complétés par Bedford, Lyubich et Smillie [2] [3].

En 1994, Hubbard et Papadopol [19] définissent, pour un endomorphisme holomorphe de $\mathbb{P}^k(\mathbb{C})$, une mesure de probabilité invariante naturelle, la *mesure d'équilibre* μ_f , en utilisant la théorie du pluripotential. Fornæss et Sibony [14] montrent que la mesure d'équilibre μ_f est mélangeante et reflète la distribution des préimages des points en dehors d'un ensemble pluripolaire. En dimension $k \geq 2$, le support de la mesure est en général plus petit que l'ensemble de Julia (nous verrons un exemple plus loin).

En 2001, Briend et Duval [7] montrent que les méthodes de Lyubich s'adaptent en toute dimension avec des résultats similaires. Nous allons présenter ces travaux de Briend et Duval. Dans les énoncés suivants, f est un endomorphisme holomorphe de $\mathbb{P}^k(\mathbb{C})$ ($k \geq 1$) de degré algébrique $d \geq 2$.

THÉORÈME ET DÉFINITION 0.2. — *Les ensembles algébriques complètement invariants par f sont en nombre fini. L'ensemble exceptionnel E_f de f est le plus grand ensemble algébrique propre complètement invariant par f .*

Il est naturel de tirer en arrière des fonctions continues sur $\mathbb{P}^k(\mathbb{C})$ et de pousser en avant des mesures sur $\mathbb{P}^k(\mathbb{C})$:

$$f^*\varphi = \varphi \circ f \quad \text{et} \quad \langle f_*\mu, \varphi \rangle = \langle \mu, f^*\varphi \rangle = \int_{\mathbb{P}^k(\mathbb{C})} \varphi \circ f \, d\mu.$$

Étant donné que $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est une application propre, elle a un degré topologique fini, à savoir d^k . La définition suivante a donc un sens :

$$\frac{1}{d^k} f_*\varphi = \frac{1}{d^k} \sum_{y \in f^{-1}(x)} \varphi(y),$$

les préimages de x étant comptées avec multiplicités. L'application $d^{-k} f_*\varphi$ obtenue en moyennant φ dans les fibres de f est alors une fonction continue (car on a pris en compte les multiplicités des préimages). On peut alors définir $d^{-k} f^*\mu$ par dualité :

$$\left\langle \frac{1}{d^k} f^*\mu, \varphi \right\rangle = \int_{\mathbb{P}^k(\mathbb{C})} \left(\frac{1}{d^k} \sum_{y \in f^{-1}(x)} \varphi(y) \right) d\mu(x),$$

les préimages de x étant comptées avec multiplicités. Par exemple, $\mu_{n,x} = d^{-kn} (f^{on})^* \delta_x$ est la mesure de comptage normalisée sur $f^{-n}(x)$:

$$\mu_{n,x} = \frac{1}{d^{kn}} \sum_{y \in f^{-n}(x)} \delta_y.$$

THÉORÈME 0.3. — *La mesure d'équilibre μ_f est l'unique mesure de probabilité vérifiant $d^{-k} f^*\mu_f = \mu_f$ et ne chargeant pas l'ensemble exceptionnel E_f . De plus, pour toute mesure de probabilité ν ne chargeant pas E_f , on a $d^{-kn} (f^{on})^*\nu \rightarrow \mu_f$. En particulier,*

$\mu_{n,x} \rightarrow \mu_f$ si et seulement si $x \in \mathbb{P}^k(\mathbb{C}) \setminus E_f$. La mesure μ_f est mélangeante, et donc ergodique.

Nous rappellerons les notions d'entropie topologique et d'entropie métrique plus loin. Depuis 1977, on sait que l'entropie topologique d'un endomorphisme de $\mathbb{P}^k(\mathbb{C})$ de degré algébrique $d \geq 2$ est $k \log d$. La minoration est due à Misiurewicz et Przytycki [23] et la majoration est due à Gromov [17]. Comme la mesure d'équilibre μ_f est de jacobien constant d^k , la formule de Rohlin-Parry [24] dit que l'entropie métrique de μ_f vaut $k \log d$. Enfin, le principe variationnel affirme que l'entropie topologique est le supremum des entropies métriques. Donc, μ_f est une mesure d'entropie maximale.

THÉORÈME 0.4. — *La mesure d'équilibre est l'unique mesure d'entropie maximale de f .*

L'ensemble des points périodiques de période n contient $(d^{n(k+1)} - 1)/(d - 1)$ points, en comptant les multiplicités. Le théorème suivant montre que la majorité d'entre eux sont répulsifs et que la mesure d'équilibre reflète leur distribution asymptotique.

THÉORÈME 0.5. — *La suite de mesures $\frac{1}{d^{nk}} \sum_{f^{on}(y)=y, y \text{ répulsif}} \delta_y$ converge faiblement vers la mesure d'équilibre μ_f .*

En dimension $k = 1$, tous les points périodiques, sauf au plus un nombre fini, sont répulsifs ; les points périodiques répulsifs sont dans l'ensemble de Julia qui est le support de μ_f . En dimension $k \geq 2$, Hubbard et Papadopol [19] (exemple 2, page 343) montrent qu'il se peut que les points périodiques répulsifs ne soient pas dans le support de la mesure d'équilibre. Fornæss et Sibony [15] (théorème 4.3 et remarque 4.4) montrent également qu'il se peut qu'il y ait une infinité de tels points.

Antérieurement aux résultats que nous présentons ici, Briend et Duval [6] en avaient obtenu un autre que nous ne démontrerons pas.

THÉORÈME 0.6. — *Les exposants de Liapounoff de f relativement à μ sont minorés par $(\log d)/2$.*

Signalons [10] où Dinh et Sibony ont généralisé toute cette théorie au cas des applications d'allure polynomiale (qui contient le cas des endomorphismes de $\mathbb{P}^k(\mathbb{C})$). Ils donnent une construction directe de la mesure d'équilibre, n'utilisant que les fonctions plurisousharmoniques. Ils montrent que la mesure d'équilibre est K -mélangeante (ce qui est la notion la plus forte de mélange) et qu'elle est exponentiellement mélangeante.

Signalons également les travaux de Guedj [18] et de Dinh et Sibony [11] qui généralisent de nombreux résultats dans le cadre des transformations méromorphes (l'ensemble d'indétermination n'est pas vide).

1. DÉFINITION POTENTIALISTE DE LA MESURE D'ÉQUILIBRE

1.1. La mesure de Brolin pour les polynômes

Considérons d'abord le cas d'un polynôme $P : \mathbb{C} \rightarrow \mathbb{C}$ de degré $d \geq 2$. Dans ce cas, on peut définir l'ensemble de Julia rempli

$$K_P = \{z \in \mathbb{C} ; \text{la suite } (P^{on}(z))_{n \geq 0} \text{ est bornée}\}.$$

L'ensemble de Julia J_P est le bord de l'ensemble de Julia rempli K_P .

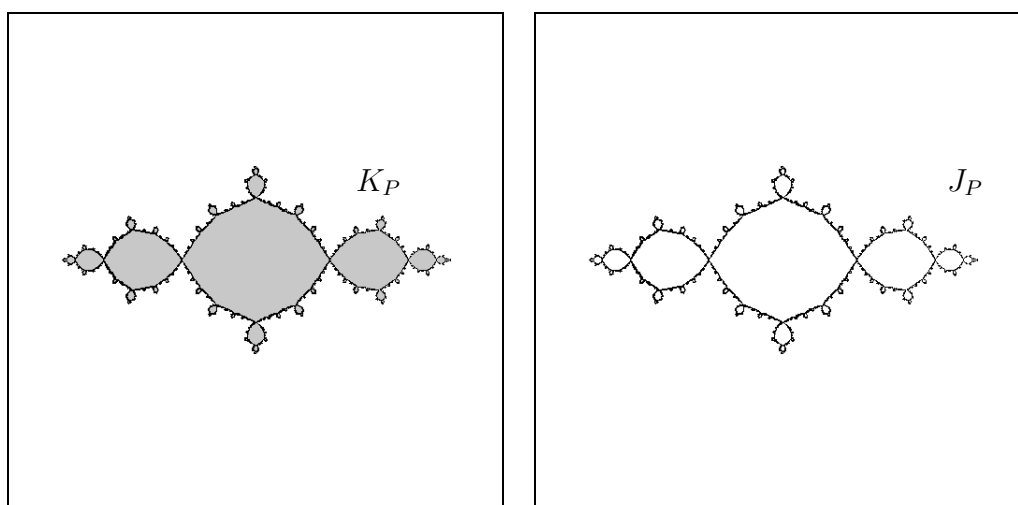


FIGURE 2. L'ensemble de Julia rempli et l'ensemble de Julia du polynôme $P : z \mapsto z^2 - 1$.

On peut alors définir une fonction $g_P : \mathbb{C} \rightarrow \mathbb{R}$ par

$$g_P(z) = \lim_{n \rightarrow +\infty} \frac{1}{d^n} \log^+ |P^{on}(z)|,$$

où $\log^+ = \max(0, \log)$. Il est facile de voir que la limite existe et est uniforme en remarquant que la fonction $u(z) = \log^+ |P(z)| - d \log^+ |z|$ est bornée sur \mathbb{C} et que

$$g_P(z) = \log^+ |z| + \sum_{m=1}^{+\infty} \frac{1}{d^m} u(P^{om}(z)).$$

La fonction g_P est continue et sousharmonique comme limite uniforme de fonctions continues et sousharmoniques. Elle s'annule sur K_P et est harmonique sur $\mathbb{C} \setminus K_P$. Comme g_P est sousharmonique, son Laplacien est une mesure. En fait, $\mu_P = \frac{1}{2\pi} \Delta g_P$ est une mesure de probabilité de support égal à J_P .

La mesure μ_P est appelée la mesure de Brolin du polynôme P . On voit facilement que $g_P(P(z)) = d \cdot g_P(z)$ et on en déduit que $d^{-1} P^* \mu_P = \mu_P$. Pour établir la distribution des images réciproques d'un point w , on est amené à montrer que la suite de fonctions sousharmoniques

$$\frac{1}{d^n} \log |P^{on}(z) - w|$$

converge vers g_P . De même, pour établir la distribution des points périodiques, on est amené à montrer que la suite de fonctions sousharmoniques

$$\frac{1}{d^n} \log |P^{\circ n}(z) - z|$$

converge vers g_P .

1.2. La mesure d'équilibre pour les endomorphismes holomorphes de $\mathbb{P}^k(\mathbb{C})$

1.2.1. Passage aux coordonnées homogènes. — Soient $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ un endomorphisme holomorphe de degré algébrique $d \geq 2$ et $F : \mathbb{C}^{k+1} \setminus \{0\} \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$ un relevé dont les coordonnées sont des polynômes homogènes de degré d :

$$\begin{array}{ccc} \mathbb{C}^{k+1} \setminus \{0\} & \xrightarrow{F} & \mathbb{C}^{k+1} \setminus \{0\} \\ \pi \downarrow & & \downarrow \pi \\ \mathbb{P}^k(\mathbb{C}) & \xrightarrow{f} & \mathbb{P}^k(\mathbb{C}). \end{array}$$

On écrit $F = (F_0, F_1, \dots, F_k)$ et $f = [F_0 : F_1 : \dots : F_k]$. On peut prolonger F à $\mathbb{P}^{k+1}(\mathbb{C})$ en posant

$$F([z_0 : z_1 : \dots : z_k : z_{k+1}]) = [F_0 : F_1 : \dots : F_k : z_{k+1}^d].$$

Dans ce numéro, nous allons préciser les relations qui existent entre la dynamique de $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ et la dynamique de $F : \mathbb{P}^{k+1}(\mathbb{C}) \rightarrow \mathbb{P}^{k+1}(\mathbb{C})$.

Le point fixe $O = [0 : \dots : 0 : 1]$ est complètement invariant par F . C'est un point super-attractif et toute orbite rencontrant un voisinage de O suffisamment petit converge vers O . Le bassin d'attraction Ω_0 de ce point fixe super-attractif est contenu dans l'ensemble de Fatou de F . Il est homéomorphe à une boule réelle de dimension $2(k + 1)$.

L'hyperplan à l'infini $H = \{z_{k+1} = 0\}$ est complètement invariant par F et $F : H \rightarrow H$ est conjuguée à $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$.

Pour $z = [z_0 : \dots : z_k] \in \mathbb{P}^k(\mathbb{C})$, on note $L_z \subset \mathbb{P}^{k+1}(\mathbb{C})$ la droite passant par O avec « pente » z :

$$L_z = \{[z_0 : \dots : z_k : 0]\} \cup \{[tz_0 : \dots : tz_k : 1] ; t \in \mathbb{C}\}.$$

L'application F envoie L_z sur $L_{f(z)}$ et $F : L_z \rightarrow L_{f(z)}$ est de la forme $t \mapsto \lambda t^d$. En particulier, un cercle de rayon r centré en O s'envoie sur un cercle de rayon $|\lambda|r^d$ centré en O .

L'ensemble de Julia J_F est la réunion du bord de Ω_0 , et du cône $C(J_f) \setminus \Omega_0$ où $C(J_f)$ est la réunion des droites L_z pour $z \in J_f$. Le bord de Ω_0 est homéomorphe à une sphère de dimension $2k + 1$.

Exemple 1.1. — Considérons l'exemple de l'endomorphisme f de $\mathbb{P}^2(\mathbb{C})$ défini par :

$$f([z_0 : z_1 : z_2]) = [z_0^d : z_1^d : z_2^d].$$

Il est de degré algébrique d et de degré topologique d^2 . Il admet trois points fixes super-attractifs complètement invariants : les points $[0 : 0 : 1]$, $[0 : 1 : 0]$ et $[1 : 0 : 0]$.

Les trois droites $\{z_0 = 0\}$, $\{z_1 = 0\}$ et $\{z_2 = 0\}$ sont complètement invariantes par f . La réunion des trois bassins super-attractifs constitue l'ensemble de Fatou. Son bord est l'ensemble de Julia.

Pour visualiser le bassin attractif de $[0 : 0 : 1]$, on peut se placer dans la carte $z_2 = 1$. On est dans \mathbb{C}^2 et $F(z_0, z_1) = (z_0^d, z_1^d)$. Le bassin d'attraction Ω_0 de $(0, 0)$ est le polydisque $\{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| < 1, |z_1| < 1\}$. Son bord est la réunion de deux tores pleins de dimension réelle 3 :

$$\partial\Omega_0 = \{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| < 1, |z_1| = 1\} \cup \{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| = 1, |z_1| < 1\}.$$

Pour obtenir l'ensemble de Julia J_F , il faut ajouter un troisième tore plein :

$$\{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| = |z_1| > 1\} \cup \{[z_0 : z_1 : 0] \in \mathbb{P}^2(\mathbb{C}) ; |z_0| = |z_1|\}.$$

L'intersection de ces trois tores est le tore réel $\{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| = |z_1| = 1\}$ qui se révélera être le support de la mesure d'équilibre.

Exemple 1.2. — On peut considérer un exemple un peu plus élaboré :

$$F([z_0 : z_1 : z_2]) = [z_0^2 - z_1^2 : z_1^2 : z_2^2].$$

Si on se place dans la carte $\{z_2 = 1\}$, c'est-à-dire dans \mathbb{C}^2 , l'expression de F est $F(z_0, z_1) = (z_0^2 - z_1^2, z_1^2)$ qui est l'expression en coordonnées homogènes du polynôme $f(z) = z^2 - 1$.

Le point $[0 : 0 : 1]$ est fixe et super-attractif et son bassin d'attraction Ω_0 est homéomorphe à une boule de dimension réelle 4. Son bord est homéomorphe à une 3-sphère.

L'ensemble de Julia J_F et le bord du bassin Ω_0 sont des ensembles cerclés : si $(z_0, \dots, z_k) \in \mathbb{C}^{k+1}$ appartient à J_F (respectivement à $\partial\Omega_0$) et si $|\lambda| = 1$, alors $(\lambda z_0, \dots, \lambda z_k)$ appartient à J_F (respectivement à $\partial\Omega_0$).

En particulier, on a une fibration de $\partial\Omega_0$ sur $\mathbb{P}^1(\mathbb{C})$, les fibres étant des cercles (c'est la fibration de Hopf). On peut se faire une idée de la géométrie de $\partial\Omega_0$ en quotientant par $(z_0, z_1) \sim (\lambda z_0, \lambda z_1)$ si $|\lambda| = 1$. À chaque point (z_0, z_1) on associe alors un point $z = [z_0 : z_1] \in \mathbb{P}^1(\mathbb{C})$ et un réel $r = \sqrt{|z_0|^2 + |z_1|^2} \in]0, +\infty[$. Le point $z \in \mathbb{P}^1(\mathbb{C})$ s'identifie avec un point $M_z \in S^2 \subset \mathbb{R}^3$ via la projection stéréographique (S^2 est la sphère unité dans \mathbb{R}^3) et on peut représenter le couple (z, r) par le point $r \cdot M_z \in \mathbb{R}^3$. Le bord de Ω_0 est alors représenté par une surface de \mathbb{R}^3 (voir figure 3). On peut représenter l'ensemble de Julia J_F en suivant le même principe (voir figure 4).

Exemple 1.3. — Considérons toujours une fraction rationnelle $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ et une application $F : \mathbb{P}^2(\mathbb{C}) \rightarrow \mathbb{P}^2(\mathbb{C})$ associée par passage aux coordonnées homogènes. Si l'ensemble de Julia de f est égal à $\mathbb{P}^1(\mathbb{C})$, alors, l'ensemble de Julia de F est égal au complémentaire du bassin attractif de $[0 : 0 : 1]$. Il ne sera donc pas d'intérieur vide, et il ne sera pas égal à $\mathbb{P}^2(\mathbb{C})$.

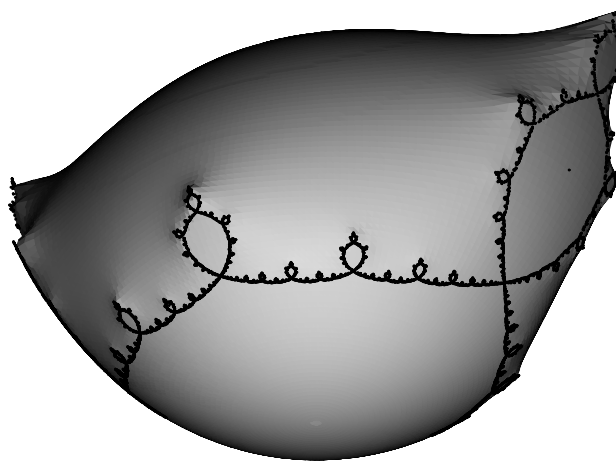


FIGURE 3. Une représentation du bord du bassin attractif Ω_0 de $(0,0)$ pour l'application $F : (z_0, z_1) \mapsto (z_0^2 - z_1^2, z_1^2)$. Cet ensemble est contenu dans l'ensemble de Julia J_F . En noir, on a représenté l'intersection $\partial\Omega_0 \cap C(J_f)$ entre le bord de Ω_0 et le cône $C(J_f)$. Cette intersection est le support de la mesure d'équilibre μ_F .

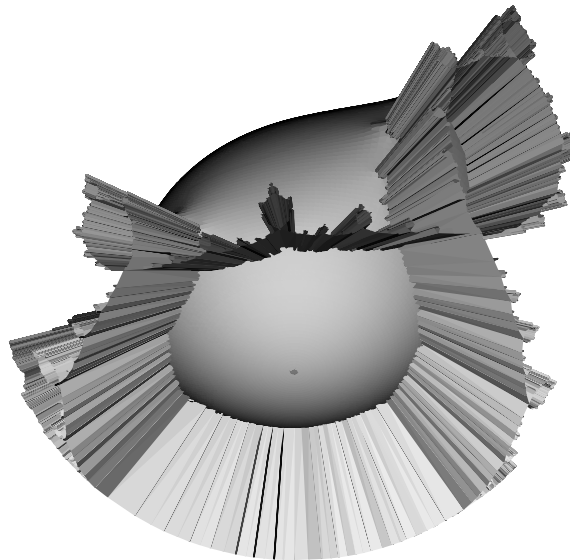


FIGURE 4. Une représentation de l'ensemble de Julia J_F pour l'application $F : (z_0, z_1) \mapsto (z_0^2 - z_1^2, z_1^2)$. On s'est en fait restreint à représenter l'intersection de J_F avec la boule de rayon 3 dans \mathbb{C}^2 .

1.2.2. *La fonction de Green d'un endomorphisme de $\mathbb{P}^k(\mathbb{C})$.* — Pour $x \in \mathbb{C}^{k+1} \setminus \{0\}$, on pose $u(x) = \log \|F(x)\| - d \log \|x\|$, où $\|\cdot\|$ est la norme euclidienne sur \mathbb{C}^{k+1} .

Pour $n \geq 0$, on définit

$$G_n(x) = \frac{1}{d^n} \log \|F^{\circ n}(x)\|.$$

Par homogénéité de F , on voit que la fonction u ne dépend que de $\pi(x) \in \mathbb{P}^k(\mathbb{C})$ et que u est donc bornée sur $\mathbb{C}^{k+1} \setminus \{0\}$. On en déduit immédiatement que la suite

$$G_n(x) - \log \|x\| = \sum_{m=1}^n G_m(x) - G_{m-1}(x) = \sum_{m=1}^n \frac{1}{d^m} u(F^{\circ m}(x)).$$

converge uniformément sur $\mathbb{C}^{k+1} \setminus \{0\}$ vers une fonction continue bornée : $G_F(x) - \log \|x\|$. On a donc la

PROPOSITION 1.4. — *La fonction*

$$G_F = \lim_{n \rightarrow +\infty} \frac{1}{d^n} \log \|F^{\circ n}\|$$

est continue et plurisousharmonique sur $\mathbb{C}^{k+1} \setminus \{0\}$. Pour tout $x \in \mathbb{C}^{k+1} \setminus \{0\}$, on a $G_F(F(x)) = d \cdot G_F(x)$ et pour tout $\lambda \in \mathbb{C}^*$, on a $G_F(\lambda x) = G_F(x) + \log |\lambda|$.

Exemple 1.5. — Soit $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ un polynôme de degré d . Considérons alors l'application $F : \mathbb{C}^2 \setminus \{0\} \rightarrow \mathbb{C}^2 \setminus \{0\}$ définie par

$$F(z_0, z_1) = (z_1^d f(z_0/z_1), z_1^d).$$

L'application F relève f en coordonnées homogènes. On a alors

$$F^{\circ n}(z_0, z_1) = (z_1^{d^n} f^{\circ n}(z_0/z_1), z_1^{d^n}).$$

Par conséquent,

$$\begin{aligned} G_F(z_0, z_1) &= \lim_{n \rightarrow +\infty} \frac{1}{d^n} \log \|F^{\circ n}(z_0, z_1)\| \\ &= \lim_{n \rightarrow +\infty} \frac{1}{2d^n} \log \left(|z_1^{d^n} f^{\circ n}(z_0/z_1)|^2 + |z_1^{d^n}|^2 \right) \\ &= \log |z_1| + g_f(z), \end{aligned}$$

où $z = [z_0 : z_1] \in \mathbb{P}^1(\mathbb{C})$ et g_f est la fonction de Green associée au polynôme f .

1.2.3. Courants positifs sur $\mathbb{P}^k(\mathbb{C})$. — Une forme lisse de bidegré (p, q) sur $\mathbb{P}^k(\mathbb{C})$ est une forme différentielle lisse qui, dans des cartes, s'écrit

$$\varphi = \sum_{|\alpha|=p, |\beta|=q} \varphi_{\alpha, \beta} dz_{\alpha} \wedge d\bar{z}_{\beta}$$

où $dz_{\alpha} = dz_{\alpha_1} \wedge \dots \wedge dz_{\alpha_p}$ et $d\bar{z}_{\beta} = d\bar{z}_{\beta_1} \wedge \dots \wedge d\bar{z}_{\beta_q}$. Un courant de bidegré (p, q) sur $\mathbb{P}^k(\mathbb{C})$ est un élément du dual des formes lisses de bidegré $(k-p, k-q)$ sur $\mathbb{P}^k(\mathbb{C})$. Dans des cartes, un courant S de bidegré (p, q) se représente comme une forme différentielle de bidegré (p, q) à coefficients distributions

$$S = \sum_{|\alpha|=p, |\beta|=q} S_{\alpha, \beta} dz_{\alpha} \wedge d\bar{z}_{\beta}.$$

L'opérateur d de Poincaré se décompose en $d = \partial + \bar{\partial}$ où

$$\partial\varphi = \sum \frac{\partial\varphi_{\alpha,\beta}}{\partial z_k} dz_k \wedge dz_\alpha \wedge d\bar{z}_\beta \quad \text{et} \quad \bar{\partial}\varphi = \sum \frac{\partial\varphi_{\alpha,\beta}}{\partial \bar{z}_k} d\bar{z}_k \wedge dz_\alpha \wedge d\bar{z}_\beta.$$

On définit l'opérateur

$$d^c = \frac{i}{2\pi}(\bar{\partial} - \partial).$$

On a

$$dd^c = \frac{i}{\pi}\partial\bar{\partial}.$$

Si S est un courant de bidegré (p, p) , on a

$$\langle dS, \varphi \rangle = -\langle S, d\varphi \rangle, \quad \langle d^c S, \varphi \rangle = -\langle S, d^c \varphi \rangle \quad \text{et} \quad \langle dd^c S, \varphi \rangle = \langle S, dd^c \varphi \rangle.$$

Remarque 1.6. — Si U est un ouvert de \mathbb{C} et $h : U \rightarrow \mathbb{R}$ est une fonction de classe \mathcal{C}^2 , alors $dd^c h = \left(\frac{1}{2\pi}\Delta h\right) dx \wedge dy$, où Δh est le Laplacien de h .

On dit qu'un courant S de bidegré (p, p) est positif si $\langle S, \varphi \rangle \geq 0$ pour toute forme test

$$\varphi = i\varphi_1 \wedge \bar{\varphi}_1 \wedge \dots \wedge i\varphi_{k-p} \wedge \bar{\varphi}_{k-p},$$

avec φ_j forme lisse de bidegré $(1, 0)$. Dans des cartes, les coefficients d'un courant positif S de bidegré (p, p) sont des mesures régulières : ce sont des distributions qui se prolongent en formes linéaires continues sur l'espace des fonctions continues.

On définit alors la masse d'un courant positif S de bidegré (p, p) par

$$\|S\| = \int_{\mathbb{P}^k(\mathbb{C})} S \wedge \frac{\omega^{k-p}}{(k-p)!}.$$

Si une suite de courants positifs sur $\mathbb{P}^k(\mathbb{C})$ est de masse bornée, on peut en extraire une sous-suite convergente. Si la suite est croissante et de masse bornée, alors elle est convergente.

Un courant positif fermé de bidegré $(1, 1)$ s'écrit localement $dd^c u$ avec u plurisousharmonique.

1.2.4. Le courant de Green d'un endomorphisme de $\mathbb{P}^k(\mathbb{C})$. — À partir de la fonction de Green G_F , on peut définir un courant positif fermé de bidegré $(1, 1)$ sur $\mathbb{P}^k(\mathbb{C})$ comme suit.

PROPOSITION 1.7. — *Il existe un courant positif fermé T_f de bidegré $(1, 1)$ sur $\mathbb{P}^k(\mathbb{C})$ tel que pour tout ouvert $U \subset \mathbb{P}^k(\mathbb{C})$ et toute section holomorphe $\sigma : U \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$, on ait*

$$T_f|_U = dd^c(G_F \circ \sigma).$$

PREUVE — Étant donné un ouvert $U \in \mathbb{P}^k(\mathbb{C})$ et une section holomorphe $\sigma : U \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$, la fonction $G_F \circ \sigma$ est une fonction plurisousharmonique, et $T_f|_U = dd^c(G_F \circ \sigma)$ est un courant positif fermé de bidegré $(1, 1)$. Remarquons que la définition de $T_f|_U$ ne dépend pas du choix de section. En effet, si $\sigma_0 : U \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$ et $\sigma_1 : U \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$ sont deux sections holomorphes, alors il existe une fonction holomorphe $\lambda : U \rightarrow \mathbb{C}^*$ telle que $\sigma_0 = \lambda \cdot \sigma_1$. Dans ce cas,

$$dd^c(G_F \circ \sigma_0) = dd^c \log |\lambda| + dd^c(G_F \circ \sigma_1) = dd^c(G_F \circ \sigma_1)$$

puisque $\log |\lambda|$ est une fonction pluriharmonique sur U . □

Remarque 1.8. — Hubbard et Papadopol [19] ont montré que le support de T_f est contenu dans l'ensemble de Julia J_f . Fornæss et Sibony [13] et indépendamment Ueda [27] ont montré que le support de T_f coïncide avec J_f .

Rappelons que la forme de Fubini-Study sur $\mathbb{P}^k(\mathbb{C})$ est la forme ω de bidegré $(1, 1)$ qui vérifie

$$\pi^* \omega = dd^c \log \|\cdot\|.$$

Étant donné que pour tout $x \in \mathbb{C}^{k+1} \setminus \{0\}$ et tout $\lambda \in \mathbb{C}^*$ on a

$$G_F(\lambda x) = G_F(x) + \log |\lambda| \quad \text{et} \quad \log \|\lambda x\| = \log \|x\| + \log |\lambda|,$$

la fonction $x \in \mathbb{C}^{k+1} \mapsto G_F(x) - \log \|x\|$ descend en une fonction continue φ sur $\mathbb{P}^k(\mathbb{C})$ et on a

$$T_f = \omega + dd^c \varphi.$$

Le courant T_f est donc cohomologue à la forme de Fubini-Study ω .

1.2.5. *Première définition de la mesure d'équilibre.* — Par construction, au voisinage de chaque point $x \in \mathbb{P}^k(\mathbb{C})$, $T_f = dd^c u$ avec u continue. Il est possible de définir le produit extérieur de courants positifs fermés de bidegrés $(1, 1)$ dont les potentiels sont continus (voir [4]). On peut donc parler de

$$T_f^{\wedge k} = \underbrace{T_f \wedge T_f \wedge \dots \wedge T_f}_{k \text{ fois}}.$$

Comme le courant T_f est cohomologue à la forme de Fubini-Study ω , $T_f^{\wedge k}$ est cohomologue à la forme volume $\Omega = \omega^{\wedge k}$ de la métrique de Fubini-Study. Ceci nous amène à la définition suivante.

DÉFINITION 1.9. — *La mesure d'équilibre μ_f d'un endomorphisme $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est la mesure de probabilité $\mu_f = T_f^{\wedge k}$.*

On a

$$G_F = \lim_{n \rightarrow +\infty} G_n \quad \text{avec} \quad G_n = \frac{1}{d^n} \log \|F^{on}\|.$$

Pour $n \geq 0$, il existe une forme lisse T_n de bidegré $(1, 1)$ sur $\mathbb{P}^k(\mathbb{C})$ telle que $\pi^* T_n = dd^c G_n$. En fait,

$$T_n = \frac{1}{d^n} (f^{on})^* \omega.$$

La fonction $G_n(x) - \log \|x\|$ descend en une fonction lisse φ_n sur $\mathbb{P}^k(\mathbb{C})$, $\varphi_n \rightarrow \varphi$ uniformément sur $\mathbb{P}^k(\mathbb{C})$ et

$$T_n = \omega + dd^c \varphi_n.$$

On a donc $T_n \rightarrow T_f$ au sens des courants, et comme la convergence de φ_n vers φ est uniforme, on a la convergence suivante au sens des courants (voir [4]) :

$$T_n \rightarrow T_f \quad \text{et} \quad T_n^{\wedge k} = \frac{1}{d^{kn}} (f^{\circ n})^* \Omega \rightarrow T_f^{\wedge k} = \mu_f.$$

On vient de démontrer la

PROPOSITION 1.10. — *La mesure d'équilibre μ_f est la limite faible de la suite de mesures de probabilité $\mu_n = \frac{1}{d^{kn}} (f^{\circ n})^* \Omega$, où $\Omega = \omega^{\wedge k}$ est la forme volume de la métrique de Fubini-Study.*

Exemple 1.11. — Considérons de nouveau l'endomorphisme f de $\mathbb{P}^2(\mathbb{C})$ défini par

$$f([z_0 : z_1 : z_2]) = [z_0^d : z_1^d : z_2^d].$$

L'ensemble de Julia J_f est la réunion de trois tores pleins qui s'intersectent le long du tore réel $\{(z_0, z_1) \in \mathbb{C}^2 ; |z_0| = 1 \text{ et } |z_1| = 1\}$. Ce tore réel est totalement invariant. Il contient tous les points périodiques répulsifs. Les préimages d'un point $(z_0, z_1) \in \mathbb{C}^2$ hors des droites $\{z_0 = 0\}$ et $\{z_1 = 0\}$ s'accroissent sur ce tore réel. Ce tore réel est le support de la mesure d'équilibre μ_f qui n'est autre que la mesure de Lebesgue sur le tore. Le support de la mesure d'équilibre μ_f n'est donc pas égal à J_f .

Exemple 1.12. — Considérons maintenant l'endomorphisme F de $\mathbb{P}^2(\mathbb{C})$ défini par

$$F([z_0 : z_1 : z_2]) = [z_0^2 - z_1^2 : z_1^2 : z_2^2].$$

Le bord du bassin attractif Ω_0 du point $[0 : 0 : 1]$ est représenté sur la figure 3. Le support de la mesure d'équilibre μ_F est contenu dans $\partial\Omega_0$. C'est l'intersection du bord de Ω_0 avec le cône $C(J_f)$. La projection $\pi_*(\mu_F)$ sur $\mathbb{P}^1(\mathbb{C})$ est la mesure d'équilibre μ_f de f ($\pi : \mathbb{C}^2 \setminus \{0\} \rightarrow \mathbb{P}^1(\mathbb{C})$ est la projection canonique).

Exemple 1.13. — La mesure d'équilibre d'un polynôme construite en passant aux coordonnées homogènes est égale à la mesure de Brolin obtenue à partir de la fonction de Green $g_f : \mathbb{C} \rightarrow [0, +\infty[$.

En effet, nous avons mentionné plus haut (exemple 1.5) que $G_F(z_0, z_1) = \log |z_1| + g_f(z)$, où $z = [z_0 : z_1] \in \mathbb{P}^1(\mathbb{C})$ et g_f est la fonction de Green associée au polynôme f . La fonction $(z_0, z_1) \mapsto G_F(z_0, z_1) - \log \|(z_0, z_1)\|$ définie sur $\mathbb{C}^2 \setminus \{0\}$ descend en une fonction $\varphi : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{R}$ définie par

$$\varphi(z) = g_f(z) - \frac{1}{2} \log(1 + |z|^2).$$

On a alors

$$\mu_f = \omega + \left(\frac{1}{2\pi} \Delta \varphi \right) dx \wedge dy = \left(\frac{1}{2\pi} \Delta g_f \right) dx \wedge dy,$$

où ω est la forme de Fubini-Study sur $\mathbb{P}^1(\mathbb{C})$.

1.2.6. *Seconde définition.* — Dans l’approche que nous venons de présenter, nous avons eu besoin d’utiliser la théorie de Bedford-Taylor pour définir μ_f comme puissance k -ième d’un courant positif fermé T_f de bidegré $(1, 1)$ et pour obtenir la convergence faible de la suite de mesures $\mu_n = T_n^{\wedge k}$ vers $\mu_f = T_f^{\wedge k}$.

Dinh et Sibony [10] puis Guedj [18] ont montré qu’on peut en fait directement définir la mesure d’équilibre μ_f comme la limite faible de la suite de mesures de probabilité $\mu_n = \frac{1}{d^{kn}}(f^{on})^*\Omega$. Nous allons maintenant présenter l’approche de Guedj.

On observe d’abord que $d^{-k}f^*\Omega$ est cohomologue à Ω , et on peut donc écrire

$$\frac{1}{d^k}f^*\Omega = \Omega + dd^c S$$

avec S une forme lisse de bidegré $(k - 1, k - 1)$. On a alors, par compacité de $\mathbb{P}^k(\mathbb{C})$, l’encadrement $-C\omega^{\wedge k-1} \leq S \leq C\omega^{\wedge k-1}$ pour une constante $C \geq 0$. Quitte à rajouter $C\omega^{\wedge k-1}$ à S , on peut supposer que $0 \leq S \leq C\omega^{\wedge k-1}$. On a alors

$$\frac{1}{d^{kn}}(f^{on})^*\Omega = \Omega + dd^c S_n \quad \text{avec} \quad S_n = S + \frac{1}{d^k}f^*S + \dots + \frac{1}{d^{k(n-1)}}(f^{o(n-1)})^*S.$$

Comme $S \geq 0$, la suite de courants positifs S_n est croissante. La masse des courants S_n est bornée :

$$0 \leq \|S_n\| = \int_{\mathbb{P}^k(\mathbb{C})} S_n \wedge \omega \leq C \sum_{j=0}^{n-1} \frac{1}{d^{kj}} \int_{\mathbb{P}^k(\mathbb{C})} (f^{oj})^*\omega^{k-1} \wedge \omega = C \sum_{j=0}^{n-1} \frac{1}{d^j} < +\infty.$$

La suite S_n converge donc faiblement vers un courant positif S_∞ de bidegré $(k - 1, k - 1)$, et

$$\frac{1}{d^{kn}}(f^{on})^*\Omega = \Omega + dd^c S_n \rightarrow \mu_f := \Omega + dd^c S_\infty.$$

L’opérateur $\mu \mapsto d^{-k}f^*\mu$ est un opérateur continu pour la convergence faible. En passant à la limite sur l’équation $\mu_{n+1} = d^{-k}f^*\mu_n$, on obtient

$$\mu_f = \frac{1}{d^k}f^*\mu_f.$$

Il en découle également que la mesure d’équilibre μ_f est f -invariante :

$$\mu_f(f^{-1}(A)) = \int \mathbf{1}_{f^{-1}(A)} d\mu_f = \int \mathbf{1}_{f^{-1}(A)} d\frac{f^*\mu_f}{d^k} = \int \mathbf{1}_A d\mu_f = \mu_f(A).$$

PROPOSITION 1.14. — *La mesure d’équilibre μ_f ne charge pas les ensembles pluripolaires. En particulier, elle ne charge pas les ensembles algébriques.*

PREUVE — Une fonction quasiplurisousharmonique sur $\mathbb{P}^k(\mathbb{C})$ est une fonction semi-continue supérieurement qui est localement la somme d’une fonction plurisousharmonique et d’une fonction lisse. Tout ensemble pluripolaire est contenu dans $\varphi^{-1}\{-\infty\}$ pour une fonction quasiplurisousharmonique φ . Nous allons voir que φ est intégrable par rapport à la mesure μ_f , ce qui implique $\varphi^{-1}\{-\infty\}$ est de mesure nulle.

Quitte à translater φ , on peut supposer que $\varphi \leq 0$ et quitte à multiplier φ par une constante, on peut supposer $dd^c\varphi \geq -\omega$. On a alors, en utilisant le théorème de Stokes,

$$\begin{aligned} 0 \leq \int_{\mathbb{P}^k(\mathbb{C})} (-\varphi) d\mu_f &= \int_{\mathbb{P}^k(\mathbb{C})} (-\varphi) d\Omega + \int_{\mathbb{P}^k(\mathbb{C})} (-dd^c\varphi) \wedge S_\infty \\ &\leq \int_{\mathbb{P}^k(\mathbb{C})} (-\varphi) d\Omega + \int_{\mathbb{P}^k(\mathbb{C})} S_\infty \wedge \omega, \end{aligned}$$

puisque S_∞ est un courant positif. Une fonction quasiplurisousharmonique est intégrable donc $\int_{\mathbb{P}^k(\mathbb{C})} (-\varphi) d\Omega < +\infty$, et $\int_{\mathbb{P}^k(\mathbb{C})} S_\infty \wedge \omega = \|S_\infty\| < +\infty$. \square

2. L'ENSEMBLE EXCEPTIONNEL

Dans le paragraphe suivant, nous allons donner une troisième construction de la mesure d'équilibre qui est due à Briend et Duval [7]. Cette approche est plus géométrique et s'inspire des travaux de Lyubich [21] dans le cas de la dimension $k = 1$. Une conséquence immédiate sera que la mesure d'équilibre μ_f reflète la distribution asymptotique des préimages des points hors d'un ensemble exceptionnel algébrique E_f .

Les premiers résultats concernant les ensembles exceptionnels algébriques pour les endomorphismes de $\mathbb{P}^k(\mathbb{C})$ avec $k \geq 2$ ont été obtenus par Fornæss et Sibony [12].

Nous allons maintenant montrer que les ensembles algébriques complètement invariants par f sont en nombre fini (résultat dû à Briend et Duval [7]). La réunion de tous ces ensembles hormis $\mathbb{P}^k(\mathbb{C})$ est donc un ensemble algébrique complètement invariant par f (éventuellement vide). C'est l'ensemble exceptionnel E_f .

Exemple 2.1. — Considérons d'abord une fraction rationnelle $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ de degré $d \geq 2$. Les ensembles algébriques propres contenus dans $\mathbb{P}^1(\mathbb{C})$ sont les ensembles finis. Si E est un ensemble fini et complètement invariant, son complémentaire $U = \mathbb{P}^1(\mathbb{C}) \setminus E$ est un ouvert complètement invariant de caractéristique d'Euler finie $\chi(U) = 2 - \#E$. La formule de Riemann-Hurwitz appliquée à $f : U \rightarrow U$ donne $\chi(U) = d\chi(U) - n$ où n est le nombre de points critiques de f contenus dans U , comptés avec multiplicités. On a alors $\chi(U) = n/(d-1) \geq 0$, ce qui montre que $\#E \leq 2$. L'ensemble exceptionnel E_f contient donc au plus 2 points.

Une étude au cas par cas montre alors que lorsque E_f contient un point fixe, f est conjuguée à un polynôme (le point fixe étant envoyé à l'infini). Lorsque E_f contient deux points, f est conjuguée soit à $z \mapsto z^d$, soit à $z \mapsto z^{-d}$.

Exemple 2.2. — Considérons maintenant une fraction rationnelle $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ de degré $d \geq 2$. Soit $F : \mathbb{C}^2 \setminus \{0\} \rightarrow \mathbb{C}^2 \setminus \{0\}$ un relèvement en coordonnées homogènes. L'application F se prolonge holomorphiquement en un endomorphisme de $\mathbb{P}^2(\mathbb{C})$ de degré algébrique d .

Le point $[0 : 0 : 1]$ est alors complètement invariant puisque c'est un point fixe de F qui n'a pas d'autre préimage que lui-même. Le degré topologique local de F en $[0 : 0 : 1]$ est d^2 . La droite à l'infini est un ensemble complètement invariant. La restriction de F à cette droite est conjuguée à $f : \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$. Le degré topologique local de F en un point générique de cette droite est d . Aux points qui correspondent aux points critiques de f , le degré topologique local de F sera $> d$.

Si l'ensemble exceptionnel de la fraction rationnelle f est vide, il n'y a pas d'autre ensemble algébrique propre complètement invariant par f . Sinon, il nous faut rajouter les points sur la droite à l'infini correspondant aux points de l'ensemble exceptionnel E_f et les droites de $\mathbb{P}^2(\mathbb{C})$ qui passent par l'origine et dont la pente correspond aux points de l'ensemble exceptionnel E_f .

Exemple 2.3. — Dans [12], Fornæss et Sibony présentent un exemple instructif qui montre qu'en dimension $k \geq 2$, contrairement à la dimension 1, les points complètement invariants ne sont pas nécessairement super-attractifs et qu'ils peuvent même se trouver dans l'ensemble de Julia. L'exemple est celui du point $[0 : 0 : 1]$ pour l'endomorphisme $f : \mathbb{P}^2(\mathbb{C}) \rightarrow \mathbb{P}^2(\mathbb{C})$ défini par

$$f([x_0 : x_1 : x_2]) = [\lambda x_0 x_2^{d-1} + x_1^d : x_0^d : x_2^d].$$

Dans la coordonnée $x_2 = 1$, l'expression de f est $f(x_0, x_1) = (\lambda x_0 + x_1^d, x_0^d)$. Si $|\lambda| > 1$, le point $[0 : 0 : 1]$ est dans l'ensemble de Julia.

2.1. Finitude des ensembles algébriques complètement invariants par f

Le degré topologique local $\deg_x f$ de f en x est le nombre de points de $f^{-1}(y)$ proches de x pour y générique proche de $f(x)$. C'est la multiplicité de x comme racine de $f(\cdot) - f(x)$. Le degré topologique local varie entre 1 et d^k et ses strates $\{x \in \mathbb{P}^k(\mathbb{C}) ; \deg_x f \geq s\}$ sont des ensembles algébriques. Si A est un ensemble algébrique irréductible, le degré local de f aux points génériques de A est $\deg_A f = \min_{x \in A} \deg_x f$.

LEMME 2.4. — *Soient A un ensemble algébrique irréductible, $s = \deg_A f$ et $p = \text{codim}(A)$. Alors $s \leq d^p$. Si de plus $f(A) = A$, alors $s = d^p$ si et seulement si A est complètement invariant par f .*

PREUVE — Soit x un point générique de A . Soit P le plan de codimension p tangent à $f(A)$ en $f(x)$. Soit Q un plan de dimension p transverse à A en x . La multiplicité de x comme point d'intersection de $f^{-1}(P)$ et de Q est s . Comme $f^{-1}(P)$ est de degré d^p , le théorème de Bezout implique que $f^{-1}(P) \cap Q$ contient exactement d^p points comptés avec multiplicités. On a donc $s \leq d^p$.

Supposons maintenant que $f(A) = A$; soit τ le degré de A . Alors, le degré de $f^{-1}(A)$, en comptant les multiplicités, est égal à $d^p \tau$ et comme $A \subset f^{-1}(A)$ intervient avec multiplicité s , on a $s\tau \leq d^p \tau$ avec égalité si et seulement si $A = f^{-1}(A)$. \square

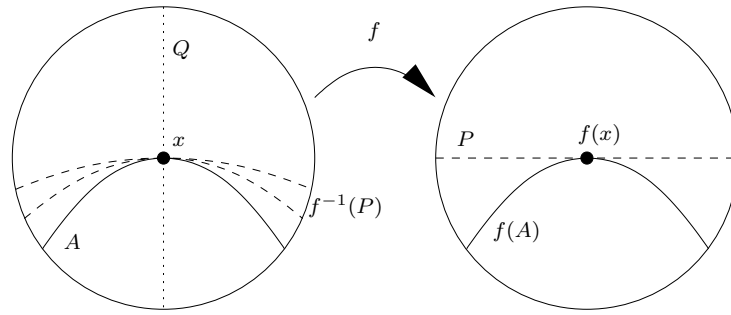


FIGURE 5. L'ensemble $f^{-1}(P) \cap Q$ contient au moins s points comptés avec multiplicités.

Les strates de degré topologique correspondant à une puissance de d vont donc jouer un rôle particulier.

DÉFINITION 2.5. — Pour $p \geq 1$, on pose

$$A_p = \{x \in \mathbb{P}^k(\mathbb{C}) ; \deg_x f \geq d^p\}.$$

LEMME 2.6. — Soit A un ensemble algébrique irréductible de codimension $p \geq 1$. Si A est complètement invariant par un itéré de f , alors l'orbite de A est contenue dans A_p .

PREUVE — On suppose que A est complètement invariant par f^{on} . D'une part, d'après le lemme 2.4, on a $\deg_A f^{on} = d^{pn}$. D'autre part,

$$\deg_A f^{on} = \prod_{j=0}^{n-1} \deg_{f^{oj}(A)} f$$

et d'après le lemme 2.4, on a $\deg_{f^{oj}(A)} f \leq d^p$. On a donc $\deg_A f = d^p$ et $A \subset A_p$ (de même que son orbite). \square

Soit E est un ensemble algébrique complètement invariant par f . Comme E est algébrique, ses composantes irréductibles de codimension p sont en nombre fini. Comme E est complètement invariant, ces composantes sont nécessairement permutées. Si A est une de ces composantes, elle est donc complètement invariante par un itéré de f . D'après le lemme 2.6, A est une des composantes de dimension maximale de A_p , qui sont en nombre fini puisque A_p est algébrique.

Les ensembles algébriques propres complètement invariants par f sont donc en nombre fini et leur réunion est un ensemble algébrique : l'ensemble exceptionnel E_f .

Exemple 2.7. — Dans le cas de la fonction $F : [x : y : t] \mapsto [x^2 - y^2 : y^2 : t^2]$ que nous avons évoqué plus haut, $E_2 = \{[0 : 0 : 1], [1 : 0 : 0]\}$. La strate E_1 est la réunion de la droite à l'infini et de la droite passant par l'origine avec pente $x/y = \infty \in \mathbb{P}^1(\mathbb{C})$.

La démonstration de Briend et Duval donne également le

LEMME 2.8. — L'ensemble exceptionnel E_f a les propriétés suivantes :

- 1) la strate E_p de codimension pure p de E est la réunion des cycles de composantes de codimension p entièrement contenus dans A_p ; en particulier, E est contenu dans le lieu critique C_f de f ;
- 2) la strate de codimension pure p de $A_p(f^{on})$ décroît avec n et se stabilise sur E_p pour n assez grand ;
- 3) l'ensemble exceptionnel d'un itéré de f coïncide avec celui de f .

Enfin, signalons le résultat de Briend-Cantat-Shishikura [5] qui donne une caractérisation plus précise de E_f .

THÉORÈME 2.9. — *Tout ensemble algébrique irréductible complètement invariant par f est un espace linéaire. En particulier, E_f est une union finie d'espaces linéaires.*

PREUVE — Soit A une composante irréductible de E_p . Quitte à remplacer f par un itéré, on peut supposer que $A = f(A) = f^{-1}(A)$. En reprenant la démonstration et les notations du lemme 2.4, on voit que si x est un point générique de A , P le plan de codimension p tangent à A en $f(x)$ et Q un plan de dimension p transverse à A en x , alors $f^{-1}(P) \cap Q$ contient d^p points, en comptant les multiplicités. Comme x intervient avec multiplicité d^p , on a $f^{-1}(P) \cap Q = \{x\}$. Comme Q est n'importe quel plan de dimension p transverse à A en x , $f^{-1}(P)$ est un plan de codimension p . C'est le plan tangent à A en x .

Le degré de $f^{-1}(P)$ est égal à 1 si on ne prend pas en compte les multiplicités, et il est égal à d^p si on tient compte des multiplicités. On voit donc que le degré topologique local de f en un point générique de $f^{-1}(P)$ est égal à d^p . On vient de montrer que, pour un point générique $x \in A$, le plan tangent à A en x est contenu dans l'ensemble algébrique A_p . On en déduit que A est un plan de codimension p . \square

3. PRÉIMAGES DE POINTS HORS DE L'ENSEMBLE EXCEPTIONNEL

On va maintenant présenter l'approche de la mesure d'équilibre due à Briend et Duval. Nous consacrerons cette partie à la démonstration du

THÉORÈME 3.1. — *La mesure d'équilibre μ_f est l'unique mesure de probabilité vérifiant $d^{-k}f^*\mu_f = \mu_f$ et ne chargeant pas E_f . Pour toute mesure de probabilité ν ne chargeant pas E_f , on a $d^{-kn}(f^{on})^*\nu \rightarrow \mu_f$.*

3.1. Existence

Les résultats de la partie 1 donnent l'existence d'une mesure de probabilité vérifiant $d^{-k}f^*\mu_f = \mu_f$ et ne chargeant pas E_f . Briend et Duval démontrent cette existence de la manière suivante.

Soit $\Omega = \omega^{\wedge k}$ la forme volume de la métrique de Fubini-Study et soit μ_n la suite de mesure de probabilités $\mu_n = d^{-kn}(f^{on})^*\Omega$. L'ensemble des mesures de probabilité

sur $\mathbb{P}^k(\mathbb{C})$ est compact. L'opérateur $d^{-k}f^*$ est continu. On en déduit que toute valeur d'adhérence de la suite $\nu_n = \frac{1}{n} \sum_{m=1}^n \mu_m$ est un point fixe de $d^{-k}f^*$.

LEMME 3.2. — Une telle valeur d'adhérence ne charge pas l'ensemble critique C_f , et donc ne charge pas E_f .

Remarque 3.3. — La preuve s'inspire largement de celle du théorème 4.3 qui est donnée plus bas.

PREUVE — Désignons par $\text{Jac}(f)$ le jacobien de f pour la forme volume Ω (on a $f^*\Omega = \text{Jac}(f)\Omega$). Soit M le maximum de $\text{Jac}(f)$ sur $\mathbb{P}^k(\mathbb{C})$. Fixons $\eta \in]0, 1[$. On peut alors choisir un voisinage U de C_f assez petit pour que $\varepsilon = \max_U \text{Jac}(f)$ vérifie $\varepsilon^\eta M^{1-\eta} < d^k$ (ceci est possible car $\text{Jac}(f) \equiv 0$ sur C_f). On va montrer que $\limsup_{n \rightarrow +\infty} \nu_n(U) \leq \eta$. On a

$$\nu_n(U) = \frac{1}{n} \sum_{m=1}^n \mu_m(U) = \frac{1}{n} \int_{\mathbb{P}^k(\mathbb{C})} \sum_{j=0}^{n-1} \mathbf{1}_U \circ f^{\circ j} \, d\mu_n \leq \mu_n(X_n) + \eta,$$

où $X_n = \{x \in \mathbb{P}^k(\mathbb{C}) ; \text{Card}(U \cap \{x, \dots, f^{n-1}(x)\}) \geq \eta n\}$ est l'ensemble des points dont la n -orbite visite souvent U . Or

$$\mu_n(X_n) = \int_{X_n} \frac{1}{d^{kn}} (f^{\circ n})^* \Omega = \int_{X_n} \frac{\text{Jac}(f^{\circ n})}{d^{kn}} \Omega = \int_{X_n} \frac{\prod_{j=0}^{n-1} (\text{Jac}(f) \circ f^{\circ j})}{d^{kn}} \Omega \leq \left(\frac{\varepsilon^\eta M^{1-\eta}}{d^k} \right)^n.$$

En passant à la limite quand $n \rightarrow +\infty$, on a $\mu_n(X_n) \rightarrow 0$ et donc

$$\limsup_{n \rightarrow +\infty} \nu_n(U) \leq \eta.$$

□

3.2. Unicité

La difficulté consiste à démontrer l'unicité d'une mesure de probabilité μ vérifiant $d^{-k}f^*\mu = \mu$ et ne chargeant pas E_f . Pour cela, Briend et Duval montrent que les préimages de deux points x et y hors de E_f s'équidistribuent de la même manière : $\mu_{n,x} - \mu_{n,y}$ converge vers 0 faiblement. Le théorème découle alors immédiatement en moyennant ce résultat. Si μ est une mesure vérifiant $d^{-k}f^*\mu = \mu$ et ne chargeant pas E_f et si ν ne charge pas E_f , on peut écrire

$$\mu = \int_{\mathbb{P}^k(\mathbb{C})} \delta_x \, d\mu(x) \quad \text{et} \quad \nu = \int_{\mathbb{P}^k(\mathbb{C})} \delta_x \, d\nu(x)$$

de sorte que

$$\mu - \frac{1}{d^{kn}} (f^{\circ n})^* \nu = \frac{1}{d^{kn}} (f^{\circ n})^* \mu - \frac{1}{d^{kn}} (f^{\circ n})^* \nu = \iint_{E_f^c \times E_f^c} (\mu_{n,x} - \mu_{n,y}) \, d\mu(x) \otimes d\nu(y) \rightarrow 0.$$

Il suffit donc de montrer que les préimages de deux points x et y hors de E_f s'équidistribuent de la même manière.

Notons C_f le lieu critique de f , $V_f = f(C_f)$ l'ensemble des valeurs critiques de f , $V_l = \bigcup_{j=1}^l f^{oj}(C_f)$ l'ensemble des valeurs critiques de f^{ol} , et $V_\infty = \bigcup_{l \geq 1} V_l$ l'ensemble post-critique de f .

Il nous faut donc montrer que pour toute fonction $\varphi : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{R}$ continue, et tous points x, y hors de E_f , la moyenne $\langle \mu_{n,x}, \varphi \rangle$ de φ aux préimages n -ièmes de x est proche de la moyenne $\langle \mu_{n,y}, \varphi \rangle$ de φ aux préimages n -ièmes de y , pour n assez grand.

On se donne une fonction $\varphi : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{R}$ continue et pour $x \in \mathbb{P}^k(\mathbb{C})$, on note $\varphi_n = d^{-kn}(f^{on})_*\varphi$ la moyenne de φ aux préimages n -ièmes. On va d'abord montrer le

LEMME 3.4. — *Pour tout $\varepsilon > 0$, il existe $l \geq 1$ tel que pour tout compact $K \subset \mathbb{P}^k(\mathbb{C}) \setminus V_l$, si n est suffisamment grand, la variation de φ_n sur K est inférieure à ε .*

On va ensuite montrer le

LEMME 3.5. — *Pour tout $l \geq 1$ et tout $x \notin E_f$, la proportion de préimages n -ièmes de x hors de V_l (en comptant les multiplicités) tend vers 1 quand n tend vers $+\infty$.*

On conclut alors de la manière suivante. Si x et y sont hors de E_f , on choisit m assez grand pour que la proportion de préimages m -ièmes de x et de y hors de V_l soit supérieure à $1 - \varepsilon$. La réunion de ces préimages est un compact $K \subset \mathbb{P}^k(\mathbb{C}) \setminus V_l$. Pour n assez grand, la variation de φ_n sur K est inférieure à ε . On obtient alors facilement

$$|\varphi_{m+n}(x) - \varphi_{m+n}(y)| \leq 2\varepsilon \sup_{\mathbb{P}^k(\mathbb{C})} |\varphi| + \varepsilon.$$

Venons-en à la démonstration des lemmes 3.4 et 3.5.

Un disque holomorphe est dit *plat générique* s'il est tracé sur une droite projective non contenue dans l'ensemble post-critique V_∞ . Le lemme 3.4 est un corollaire du

LEMME 3.6. — *Soit $\varepsilon > 0$. Il existe un entier $l \geq 0$ tel que sur tout disque plat générique Δ relativement compact dans $\mathbb{P}^k(\mathbb{C}) \setminus V_l$, on peut construire $(1 - \varepsilon)d^{kn}$ branches inverses de f^{on} pour n assez grand, d'images Δ_i^{-n} avec $\sup_i \text{Diam}(\Delta_i^{-n}) \xrightarrow{n \rightarrow +\infty} 0$.*

Remarque 3.7. — Briend et Duval montrent qu'on peut définir $(1 - \varepsilon)d^{kn}$ branches inverses d'images Δ_i^{-n} avec $\text{Diam}(\Delta_i^{-n}) = \mathcal{O}(d^{-n/2})$.

PREUVE — Soit τ le degré de V_f (l'ensemble des valeurs critiques de f). On fixe l de sorte que $2\tau d^{-l}(1 - 1/d)^{-1} < \varepsilon$. On considère un disque holomorphe Δ relativement compact dans $\mathbb{P}^k(\mathbb{C}) \setminus V_l$ et tracé sur une droite $L \not\subset V_\infty$. Soit $\tilde{\Delta} \supset \Delta$ un disque holomorphe tracé sur L et évitant V_l .

On commence par contrôler le nombre de branches inverses de f^{on} sur $\tilde{\Delta}$. Puisque Δ évite V_l , on dispose de d^{kl} branches inverses f_i^{-l} de f^{ol} sur $\tilde{\Delta}$, d'images $\tilde{\Delta}_i^{-l} = f_i^{-l}(\tilde{\Delta})$. Une telle branche inverse f_i^{-l} en engendre d^k pour f^{ol+1} si $\tilde{\Delta}_i^{-l}$ n'intersecte pas V_f . Or, les disques $\tilde{\Delta}_i^{-l}$ sont disjoints et tracés sur la courbe $f^{-l}(L)$ qui est de degré $d^{(k-1)l}$. Par le théorème de Bezout, cette droite intersecte V_f en au plus $\tau d^{(k-1)l}$ points. Ainsi, $d^{kl}(1 - \tau d^{-l})$ disques $\tilde{\Delta}_i^{-l}$ évitent V_f et contribuent à $d^{k(l+1)}(1 - \tau d^{-l})$ branches inverses

de $f^{o(l+1)}$ sur $\tilde{\Delta}$. En procédant par récurrence, on montre que pour $n \geq l$, il y a au moins $d^{kn}(1 - \tau d^{-l}(1 + d^{-1} + \dots + d^{-n+l+1})) \geq d^{kn}(1 - \varepsilon/2)$ branches inverses f_i^{-n} de f^{on} sur $\tilde{\Delta}$, d'images $\tilde{\Delta}_i^{-n} = f_i^{-n}(\tilde{\Delta})$.

On va maintenant contrôler l'aire de la majorité des disques $\Delta_i^{-n} = f_i^{-n}(\Delta)$. La courbe $f^{-n}(L)$ étant de degré $d^{(k-1)n}$, l'aire totale de $f^{-n}(L)$ pour la forme de Fubini-Study ω est $d^{(k-1)n}$. Comme les disques $\tilde{\Delta}_i^{-n}$ sont disjoints sur $f^{-n}(L)$, au plus $(\varepsilon/2)d^{nk}$ d'entre eux auront une aire supérieure à $(2/\varepsilon)d^{-n}$. Ainsi, pour $(1 - \varepsilon)d^{nk}$ d'entre les disques $\tilde{\Delta}_i^{-n}$, l'aire sera majorée par $(2/\varepsilon)d^{-n}$.

Ceci est suffisant pour contrôler le diamètre des disques $\Delta_i^{-n} \subset \tilde{\Delta}_i^{-n}$. En effet, la famille des branches inverses $f_i^{-n} : \tilde{\Delta} \rightarrow \tilde{\Delta}_i^{-n}$ est normale (voir la remarque 3.8 ci-dessous) et puisque l'aire des disques $\tilde{\Delta}_i^{-n}$ tend vers 0, les seules valeurs d'adhérence sont des constantes. \square

Remarque 3.8. — On peut voir que la famille des branches inverses $f_i^{-n} : \tilde{\Delta} \rightarrow \tilde{\Delta}_i^{-n}$ est normale, en passant aux coordonnées homogènes. En effet, si $F : \mathbb{C}^{k+1} \setminus \{0\} \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$ est un relevé de $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ et si $\sigma : \tilde{\Delta} \rightarrow \mathbb{C}^{k+1} \setminus \{0\}$ est une section holomorphe, les branches inverses $f_i^{-n} : \tilde{\Delta} \rightarrow \tilde{\Delta}_i^{-n}$ se relèvent en branches inverses $F_i^{-n} : \sigma(\tilde{\Delta}) \rightarrow F^{-n}(\sigma(\tilde{\Delta}))$. Ces branches inverses évitent un voisinage de 0 et un voisinage de ∞ qui sont super-attractifs pour F . Elles forment donc une famille normale.

Le lemme 3.5 est un corollaire du

LEMME 3.9. — *Si x n'appartient pas à E_f , alors $\mu_{n,x}(C_f) \rightarrow 0$ quand $n \rightarrow +\infty$.*

PREUVE — Quitte à remplacer f par un itéré, on peut supposer que pour tout $p \geq 1$, la strate E_p de codimension pure p de E_f coïncide avec A_p . On a alors $A_k = E_k$ (c'est la strate de dimension 0, une réunion finie de points avec degré local d^k). Et pour $p \leq k - 1$, les composantes irréductibles de $A_p \setminus E_f$ sont alors de codimension $\geq p + 1$.

On peut alors montrer par récurrence descendante sur p , que $\mu_{n,x}(A_p)$ décroît exponentiellement vite. Pour $p = k$, on a $A_k = E_k$ et comme $x \notin E_f \supset E_k$, on a $\mu_{n,x}(A_k) = 0$ pour tout $n \geq 0$.

Les composantes de A_{p-1} de codimension $p - 1$ sont les composantes de E_{p-1} . Comme $f^{-n}(x) \cap E_{p-1} = \emptyset$, on voit que $f^{-n}(x)$ ne rencontre A_{p-1} qu'en des composantes de codimension $\geq p$. En utilisant le théorème de Bezout, on voit alors que quand $n \rightarrow +\infty$,

$$\text{Card}(f^{-n}(x) \cap A_{p-1}) = \mathcal{O}(d^{n(k-p)}).$$

Pour majorer $\mu_{n,x}(A_{p-1})$, il suffit de majorer les multiplicités de f^{on} en ces points.

Fixons $\rho < 1$. Si les $\lfloor \rho n \rfloor$ premiers itérés d'un point $y \in f^{-n}(x)$ ne rencontrent pas A_p , la multiplicité de f^{on} en y est majorée par κ_p^n avec $\kappa_p = (d^p - 1)^\rho d^{k(1-\rho)}$. La contribution de ces points à $\mu_{n,x}(A_{p-1})$ est donc majorée par

$$\frac{\kappa_p^n \cdot \text{Card}(f^{-n}(x) \cap A_{p-1})}{d^{kn}}$$

qui décroît exponentiellement vite si on choisit $\rho < 1$ suffisamment proche de 1 pour que $\kappa_p < d^p$.

Pour les autres points, on utilise l’hypothèse de récurrence. On sait que $\mu_{n,x}(A_p)$ décroît exponentiellement vite : $\mu_{n,x}(A_p) = \mathcal{O}(\lambda_p^n)$. On en déduit qu’en comptant les multiplicités, le nombre de points $y \in f^{-n}(x)$ pour lesquels il existe $j \leq \rho n$ avec $f^{oj}(y) \in A_p$ est un $\mathcal{O}(\lambda_p^{(1-\rho)n} \cdot d^{kn})$. Leur contribution à $\mu_{n,x}(A_{p-1})$ est donc $\mathcal{O}(\lambda_p^{(1-\rho)n})$ qui décroît exponentiellement vite. \square

3.3. Mélange

PROPOSITION 3.10. — *La mesure μ_f est mélangeante, et donc ergodique.*

PREUVE — La convergence de $\mu_{n,x}$ vers μ_f se traduit dualement de la manière suivante : si φ est une fonction continue sur $\mathbb{P}^k(\mathbb{C})$, alors sur E^c ,

$$d^{-kn}(f^{on})_*\varphi \rightarrow \int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\mu_f.$$

Si ψ est une autre fonction continue, on aura donc

$$\begin{aligned} \int_{\mathbb{P}^k(\mathbb{C})} \varphi \cdot (\psi \circ f^{on}) \, d\mu_f &= \int_{\mathbb{P}^k(\mathbb{C})} \varphi \cdot (\psi \circ f^{on}) \, d\frac{(f^{on})^*\mu_f}{d^{kn}} \\ &= \int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\frac{(f^{on})^*(\psi\mu_f)}{d^{kn}} \\ &= \int_{\mathbb{P}^k(\mathbb{C})} \frac{(f^{on})_*\varphi}{d^{kn}} \psi \, d\mu_f \\ &\rightarrow \left(\int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\mu_f \right) \left(\int_{\mathbb{P}^k(\mathbb{C})} \psi \, d\mu_f \right). \end{aligned}$$

C’est le mélange. \square

Dans [10], Dinh et Sibony montrent que μ_f est K -mélangeante, c’est-à-dire que pour toute fonction $\varphi \in L^2(\mu_f)$, on a

$$\sup_{\|\psi\|_{L^2(\mu_f)}} \left| \int_{\mathbb{P}^k(\mathbb{C})} \varphi \cdot (\psi \circ f^{on}) \, d\mu_f - \left(\int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\mu_f \right) \left(\int_{\mathbb{P}^k(\mathbb{C})} \psi \, d\mu_f \right) \right| \rightarrow 0.$$

Ils montrent également que la mesure μ_f est exponentiellement mélangeante et que la vitesse de mélange est d’ordre d^{-n} : il existe une constante $A > 0$ telle que pour toute fonction $\psi \in L^\infty(\mu_f)$, et toute fonction φ de classe \mathcal{C}^2 , on a

$$\left| \int_{\mathbb{P}^k(\mathbb{C})} \varphi \cdot (\psi \circ f^{on}) \, d\mu_f - \left(\int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\mu_f \right) \left(\int_{\mathbb{P}^k(\mathbb{C})} \psi \, d\mu_f \right) \right| \leq A \|\psi\|_\infty \|\varphi\|_{\mathcal{C}^2} d^{-n}.$$

Antérieurement, Fornæss et Sibony [14] avaient montré que la vitesse était d’ordre $(d - \varepsilon)^{-n}$. Dinh et Sibony ont depuis montré que dans le cas où φ est une fonction

lipschitzienne, la mesure est encore exponentiellement mélangeante et la vitesse de mélange est d'ordre $d^{-n/2}$:

$$\left| \int_{\mathbb{P}^k(\mathbb{C})} \varphi \cdot (\psi \circ f^{cn}) \, d\mu_f - \left(\int_{\mathbb{P}^k(\mathbb{C})} \varphi \, d\mu_f \right) \left(\int_{\mathbb{P}^k(\mathbb{C})} \psi \, d\mu_f \right) \right| \leq A \|\psi\|_\infty \|\varphi\|_{\text{Lip}} d^{-n/2}.$$

4. PROPRIÉTÉS ENTROPIQUES DE LA MESURE D'ÉQUILIBRE

Nous allons maintenant montrer que la mesure d'équilibre μ_f est l'unique mesure d'entropie maximale de f .

4.1. Entropie topologique et entropie métrique

Si X est un espace compact et métrisable, et si $f : X \rightarrow X$ est une application continue, l'entropie de f est censée mesurer la rapidité à laquelle on acquiert de l'information sur la position d'un point $x \in X$ quand on observe (avec une certaine imprécision) la trajectoire pendant longtemps.

4.1.1. Entropie métrique. — Si X est muni d'une mesure de probabilité μ , savoir que x appartient à un ensemble $A \subset X$ est censé fournir une quantité d'information égale à $\log \frac{1}{\mu(A)}$. Un questionnaire qui définit une partition mesurable \mathcal{U} de X en sous-ensembles A_i correspondant aux réponses possibles procure une quantité d'information moyenne égale à

$$h_\mu(\mathcal{U}) = \sum_i \mu(A_i) \log \frac{1}{\mu(A_i)}.$$

Supposons que μ soit f -invariante et que l'on observe la trajectoire d'un point x sous itération de f pendant un temps n . On répond alors au questionnaire pour $x_j = f^{oj}(x)$, $j = 0, \dots, n-1$. Ceci définit une partition $\bigvee^n \mathcal{U}$ de X . L'entropie métrique de f définie à partir de \mathcal{U} et de μ est alors

$$h_\mu(\mathcal{U}, f) = \lim_{n \rightarrow +\infty} \frac{1}{n} h_\mu \left(\bigvee^n \mathcal{U} \right).$$

Cette limite existe et est un infimum car la suite $h_\mu(\bigvee^n \mathcal{U})$ est sous-additive (une suite (u_n) de nombres réels et positifs est sous-additive si $u_{n+n'} \leq u_n + u_{n'}$ pour tous n, n' ; pour une telle suite, la limite $\lim \frac{1}{n} u_n$ existe et est un infimum).

L'entropie de f par rapport à la mesure f -invariante μ est alors définie par

$$h_\mu(f) = \sup \{ h_\mu(\mathcal{U}, f) ; \mathcal{U} \text{ une partition mesurable de } X \text{ avec } h_\mu(\mathcal{U}) < +\infty \}.$$

Si μ et ν sont deux mesures f -invariantes, alors pour $t \in [0, 1]$, on a

$$h_{t\mu+(1-t)\nu}(f) = th_\mu(f) + (1-t)h_\nu(f).$$

L'entropie métrique de la réunion de deux ensembles invariants est donc la somme des entropies de chaque ensemble pondéré par sa mesure.

4.1.2. *Entropie topologique.* — Quand aucune mesure particulière n'intervient, on peut alors définir l'entropie topologique en comptant le nombre de réponses possibles. Si \mathcal{U} est un recouvrement de X par des ouverts, on définit

$$\#^*\mathcal{U} = \min\{\text{Card}(\mathcal{V}) ; \text{où } \mathcal{V} \text{ est un recouvrement de } X \text{ contenu dans } \mathcal{U}\}.$$

On pose $f^*\mathcal{U} = \{f^{-1}(U)\}_{U \in \mathcal{U}}$. Si \mathcal{U} et \mathcal{V} sont deux recouvrements de X par des ouverts, on pose $\mathcal{U} \vee \mathcal{V} = \{U \cap V\}_{U \in \mathcal{U}, V \in \mathcal{V}}$. On pose

$$\bigvee^n \mathcal{U} = \mathcal{U} \vee f^*\mathcal{U} \vee \dots \vee (f^{(n-1)*}\mathcal{U}).$$

Un élément non vide $W = U_0 \cap f^{-1}(U_1) \cap \dots \cap f^{-(n-1)}(U_{n-1})$ de $\bigvee^n \mathcal{U}$ correspond à un n -itinéraire dans \mathcal{U} , c'est-à-dire une suite (U_0, \dots, U_{n-1}) telle qu'il existe $x \in X$ avec $f^{oi}(x) \in U_i$ pour $i = 0, \dots, n - 1$.

On définit l'entropie topologique de f par rapport à \mathcal{U} de la manière suivante :

$$h_{\text{top}}(\mathcal{U}, f) = \lim \frac{1}{n} \log \#^* \bigvee^n \mathcal{U}.$$

Cette limite existe et est un infimum car

$$\#^* \bigvee^{n+n'} \mathcal{U} \leq \left(\#^* \bigvee^n \mathcal{U} \right) \left(\#^* \bigvee^{n'} \mathcal{U} \right)$$

et donc, le log est une fonction sous-additive de n . On définit l'entropie topologique de f par

$$h_{\text{top}}(f) = \sup_{\mathcal{U}} h_{\text{top}}(\mathcal{U}, f).$$

Cette définition ne dépend clairement que de la topologie sur X .

4.1.3. *Espaces métrisables.* — Si X est métrisable, on peut donner des caractérisations de l'entropie métrique et de l'entropie topologique faisant intervenir la distance d définissant la topologie sur X . Pour $n \geq 1$, on définit la distance dynamique d_n par

$$d_n(x, y) = \max_{0 \leq j \leq n-1} \{d(f^{oj}(x), f^{oj}(y))\}.$$

La distance d_n mesure la distance entre les n -orbites de x et de y dans l'espace produit X^n . On note $B_n(x, r)$ les boules dynamiques associées.

Dans le cas où μ est une mesure ergodique, un théorème de Brin et Katok [8] permet de définir l'entropie métrique en terme de la décroissance des masses des boules dynamiques $B_n(x, r)$.

THÉORÈME 4.1 (Brin-Katok). — Pour μ presque tout x , on a

$$h_{\mu}(f) = \sup_{\varepsilon > 0} \liminf_{n \rightarrow +\infty} -\frac{1}{n} \log \mu(B_n(x, \varepsilon)).$$

Une manière de définir l'entropie topologique consiste à considérer des ensembles F qui sont (n, ε) -séparés, c'est-à-dire tels que pour toute paire de points $(x, y) \in F^2$, on ait $d_n(x, y) \geq \varepsilon$. On pose alors

$$h_d(f, \varepsilon) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N_d(f, \varepsilon, n),$$

où $N_d(f, \varepsilon, n)$ est le nombre maximal de points de X dans un ensemble (n, ε) -séparé. On a alors la propriété suivante :

$$h_{\text{top}}(f) = \lim_{\varepsilon \rightarrow 0} h_d(f, \varepsilon) = \sup_{\varepsilon > 0} h_d(f, \varepsilon).$$

4.1.4. Le principe variationnel. — Historiquement, l'entropie topologique a été introduite après l'entropie métrique. Bien qu'il y ait des analogies entre les définitions, l'absence d'une mesure canonique de la taille des ensembles invariants amène à des différences entre les deux notions. Par exemple, l'entropie métrique de f sur la réunion de deux ensembles invariants est un barycentre des entropies sur chaque ensemble, pondéré par les mesures de chaque ensemble. L'entropie topologique de f sur la réunion de deux ensembles invariants est le maximum de l'entropie topologique sur chaque ensemble. Par conséquent, l'entropie topologique mesure la complexité dynamique maximale, alors que l'entropie métrique mesure une complexité moyenne. Il est donc naturel d'espérer que l'entropie métrique soit inférieure à l'entropie topologique. Le principe variationnel dit que l'entropie topologique est le supremum des entropies métriques :

$$h_{\text{top}}(f) = \sup\{h_{\mu}(f) ; \mu \text{ mesure de probabilité } f\text{-invariante}\}.$$

Le principe variationnel pose la question, centrale en théorie ergodique, de l'existence et de l'unicité d'une mesure μ d'entropie maximale, c'est-à-dire telle que $h_{\mu}(f) = h_{\text{top}}(f)$. Nous allons voir plus loin que dans le cas d'un endomorphisme holomorphe $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$, la mesure d'équilibre μ_f est l'unique mesure d'entropie maximale.

Nous aurons besoin d'une version relative du principe variationnel. L'entropie topologique se localise naturellement. Si on considère un ensemble $Y \subset X$ non nécessairement invariant, on définit l'entropie de f relative à Y par :

$$h_{\text{top}}(f|Y) = \sup_{\varepsilon > 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log (\max\{\text{Card}(F) ; F \text{ } (n, \varepsilon)\text{-séparé, } F \subset Y\}).$$

On regarde donc la rapidité de croissance du nombre de n -orbites ε discernables en partant de Y . Mais on ne demande pas à ce que l'orbite reste dans Y . On a alors le

LEMME 4.2 (Principe variationnel relatif). — *Si μ est une mesure ergodique et si Y est un ensemble mesurable tel que $\mu(Y) > 0$, alors*

$$h_{\mu}(f) \leq h_{\text{top}}(f|Y).$$

4.2. La mesure d'équilibre d'un endomorphisme de $\mathbb{P}^k(\mathbb{C})$ est une mesure d'entropie maximale

Commençons par rappeler le théorème suivant dû à Misiurewicz et Przytycki [23].

THÉORÈME 4.3. — *Si X est une variété compacte orientable et $f : X \rightarrow X$ est une application de classe \mathcal{C}^1 , alors $h_{\text{top}}(f) \geq \log |\text{deg } f|$.*

PREUVE — Soient d une distance définissant la topologie sur X , ω une forme volume, $\alpha \in]0, 1[$, $L = \sup_X |\text{Jac}f|$; posons $\varepsilon = L^{-\alpha/(\alpha-1)}$. Considérons l'ensemble $B = \{x \in X ; |\text{Jac}f(x)| \geq \varepsilon\}$.

(1) Il existe $\delta > 0$ tel que si $x \neq y$ sont dans B et $d(x, y) < \delta$, alors $f(x) \neq f(y)$.

(2) Supposons maintenant que x est une valeur régulière de f^{on} . On construit alors un ensemble (n, δ) -séparé dans $f^{-n}(x)$ de la manière suivante. D'abord, comme x est une valeur régulière de f , $f^{-1}(x)$ contient au moins $N = |\text{deg}f|$ préimages. Si N d'entre elles sont dans B , on dit qu'on a une bonne transition et on pose $Q_1 = \{\text{ces } N \text{ préimages}\}$. Sinon, on dit qu'on a une mauvaise transition et on pose $Q_1 = \{y\}$ avec $y \in B^c$ et $f(y) = x$. Chaque élément de Q_1 est une valeur régulière de f^{on-1} et on peut recommencer la même procédure pour tous les $y \in Q_1$. En collectant ainsi tous les ensembles obtenus, on définit $Q_2 \in f^{-2}(x)$. En itérant le procédé on obtient un ensemble $Q_n \in f^{-n}(x)$ qui est (n, δ) -séparé.

(3) Si $x \in X \setminus f^{on}(A)$ avec

$$A = \{y \in X ; \text{Card}(B \cap \{y, f(y), \dots, f^{n-1}(y)\}) \leq \alpha n\},$$

on a beaucoup de bonnes transitions. En effet, si $x \notin f^{on}(A)$, alors $Q_n \cap A = \emptyset$, et pour passer de x à n'importe quel point $y \in Q_n$, on a au moins $\lfloor \alpha n \rfloor + 1$ bonnes transitions. Par conséquent $\text{Card}(Q_n) \geq N^{\alpha n}$.

(4) Il ne reste plus qu'à démontrer que, pour tout n , on peut trouver une valeur régulière $x \in X \setminus f^{on}(A)$. Le cardinal maximal d'un ensemble (n, δ) -séparé sera alors minoré par $N^{\alpha n}$ et on aura $h_{\text{top}}(f) \geq \alpha \log N$ pour tout $\alpha \in]0, 1[$.

Si $y \in A$, alors

$$|\text{Jac}f^{on}(y)| = \prod_{j=0}^{n-1} |\text{Jac}f(f^{oj}(x))| < L^{\alpha n} \varepsilon^{n-\alpha n} = 1$$

(par choix de ε). Donc le volume de $f^{on}(A)$ est strictement inférieur au volume de X et d'après le théorème de Sard, il existe une valeur régulière $x \in X \setminus f^{on}(A)$. \square

COROLLAIRE 4.4. — Si $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est un endomorphisme holomorphe de degré algébrique d , son entropie topologique est minorée par $k \log d$.

La majoration de l'entropie d'un endomorphisme holomorphe de $\mathbb{P}^k(\mathbb{C})$ est due à Gromov [17]. Il montre que l'entropie topologique de f est majorée par $\text{lov}(f)$, où $\text{lov}(f)$ est le taux de croissance du volume du graphe itéré de f .

DÉFINITION 4.5. — On pose

$$\Gamma_n = \{(x, f(x), \dots, f^{on-1}(x)), x \in \mathbb{P}^k(\mathbb{C})\}$$

et on définit

$$\text{lov}(f) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(\text{Vol}(\Gamma_n)) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \left(\int_{\Gamma_n} \omega_n^{\wedge k} \right),$$

où ω_n est la forme de Kähler sur $\mathbb{P}^k(\mathbb{C})^n$ induite par la forme de Fubini-Study sur chaque facteur.

LEMME 4.6. — Si $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est un endomorphisme holomorphe de degré algébrique d , on a $\text{lov}(f) = k \log d$.

PREUVE — C'est un calcul cohomologique. On a

$$\begin{aligned} \text{Vol}(\Gamma_n) &= \int_{\Gamma_n} \omega_n^{\wedge k} = \int_{\mathbb{P}^k(\mathbb{C})} (\omega + f^*\omega + \dots + (f^{\circ n-1})^*\omega)^{\wedge k} \\ &= \sum_{\underline{i} \in \{0, \dots, n-1\}^k} \int_{\mathbb{P}^k(\mathbb{C})} (f^{\circ i_1})^*\omega \wedge \dots \wedge (f^{\circ i_{k-1}})^*\omega \\ &= \sum_{\underline{i} \in \{0, \dots, n-1\}^k} d^{i_1 + \dots + i_{k-1}} \\ &= (1 + d + \dots + d^{n-1})^k = \left(\frac{d^n - 1}{d - 1} \right)^k. \end{aligned}$$

Par conséquent,

$$\text{lov} f = \limsup_{n \rightarrow +\infty} \frac{k}{n} \log \frac{d^n - 1}{d - 1} = k \log d.$$

□

THÉORÈME 4.7. — Si $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est un endomorphisme holomorphe, on a $h_{\text{top}}(f) \leq \text{lov}(f)$.

PREUVE — La démonstration repose sur le théorème de Lelong [20]. Un ensemble (n, ε) -séparé F donne, via ses n -orbites, un ensemble ε -séparé G dans le graphe itéré Γ_n pour la distance produit, qui n'est autre que la distance dynamique d_n . On a donc

$$\text{Vol}(\Gamma_n) \geq \sum_{y \in G} \text{Vol}(B_n(y, \varepsilon/2) \cap \Gamma_n),$$

puisque les boules $B_n(y, \varepsilon/2)$ sont disjointes. Le théorème de Lelong fournit une minoration indépendante de n et de y du volume de Γ_n dans ces boules :

$$\text{Vol}(B_n(y, \varepsilon/2) \cap \Gamma_n) \geq c.$$

On en déduit que

$$\frac{1}{n} \log(\text{Vol}(\Gamma_n)) \geq \frac{\log c}{n} + \frac{1}{n} \log(\max\{\text{Card}(F), F \text{ } (n, \varepsilon)\text{-séparé}\}).$$

Ceci donne la majoration souhaitée.

□

En combinant les résultats de cette partie, on obtient le théorème suivant.

THÉORÈME 4.8. — Si $f : \mathbb{P}^k(\mathbb{C}) \rightarrow \mathbb{P}^k(\mathbb{C})$ est un endomorphisme holomorphe de degré algébrique d , son entropie topologique est égale à $k \log d$.

Nous allons maintenant voir que puisque la mesure d'équilibre μ_f est de jacobien constant égal à d^k (pour tout borélien B sur lequel f est injective, on a $\mu_f(B) = d^{-k}\mu_f(f(B))$), elle est d'entropie maximale.

D'après le théorème 4.1 de Brin et Katok, il suffit d'expliciter, pour $\alpha > 0$, un borélien X_α de mesure non nulle, avec $\mu_f(B_n(x, \varepsilon)) \leq d^{-kn(1-\alpha)}$ pour x dans X_α et n assez grand. On choisit un voisinage U de l'ensemble des valeurs critiques de f , assez petit pour que $\mu_f(U) \leq \alpha/2$. On pose alors

$$X_n(\alpha) = \{x \in \mathbb{P}^k(\mathbb{C}) ; \text{Card}\{j ; f^{\circ j}(x) \in U\} \leq n\alpha\}$$

et

$$X_\alpha = \liminf_{n \rightarrow +\infty} X_n(\alpha).$$

D'après le théorème de Birkhoff, pour μ_f presque tout x , on a

$$\lim_{n \rightarrow +\infty} \frac{\text{Card}\{j \in [0, n-1] ; f^{\circ j}(x) \in U\}}{n} = \lim_{n \rightarrow +\infty} \sum_{j=0}^{n-1} \mathbf{1}_U \circ f^{\circ j}(x) = \mu_f(U) \leq \alpha/2.$$

Par conséquent, $\mu_f(X_\alpha) = 1$. La masse des boules dynamiques centrées sur X_α s'estiment par récurrence grâce à la propriété de jacobien constant de μ_f . Soient $x \in X_\alpha$ et $\varepsilon < d(V, \partial U)$. Si $f^{\circ j+1}(x) \notin U$, ce qui arrive au moins $n(1-\alpha)$ fois, f réalise une injection de $B_{n-j}(f^{\circ j}(x), \varepsilon)$ dans $B_{n-j-1}(f^{\circ j+1}(x), \varepsilon)$. On a alors

$$\mu_f(B_{n-j}(f^{\circ j}(x), \varepsilon)) \leq d^{-k}\mu_f(B_{n-j-1}(f^{\circ j+1}(x), \varepsilon)).$$

Sinon, on a toujours, par invariance de μ_f ,

$$\mu_f(B_{n-j}(f^{\circ j}(x), \varepsilon)) \leq \mu_f(B_{n-j-1}(f^{\circ j+1}(x), \varepsilon)).$$

On a donc

$$\mu_f(B_n(x, \varepsilon)) \leq d^{-kn(1-\alpha)}.$$

□

4.3. Unicité de la mesure d'entropie maximale

Dans cette partie, nous allons montrer que la mesure d'équilibre est l'unique mesure d'entropie maximale. Nous supposons qu'il existe une mesure ergodique $\nu \neq \mu_f$ d'entropie maximale $k \log d$ et nous chercherons une contradiction.

LEMME 4.9. — *La mesure ν ne charge pas l'ensemble des valeurs critiques V_f .*

PREUVE — L'argument de Gromov s'adapte pour démontrer que

$$h_{\text{top}}(f|V_f) \leq \text{lov}(f|V_f) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(\text{Vol}(\Gamma_n|V_f)).$$

Le même calcul cohomologique que dans la preuve du lemme 4.6 montre que

$$\text{Vol}(\Gamma_n|V_f) = \tau \left(\frac{d^n - 1}{d - 1} \right)^{k-1}$$

où τ est le degré de V_f . On voit donc que $\text{lov}(f|V) \leq (k - 1) \log d$. D’après le principe variationnel relatif, si ν chargeait V_f , on aurait

$$h_\nu(f) = k \log d \leq h_{\text{top}}(f|V_f) \leq (k - 1) \log d,$$

d’où une contradiction. □

On voit donc que ν ne charge pas l’ensemble exceptionnel E_f . La mesure ν n’est donc pas un point fixe de l’opérateur $d^{-k} f^*$. Elle n’est pas de jacobien constant d^k .

On peut alors paver $\mathbb{P}^k(\mathbb{C}) \setminus V_f$ par des simplexes S à bords de mesure nulle pour ν . Soit $U = \bigcup_S \overset{\circ}{S}$. La préimage de chaque simplexe S est une union disjointe de d^k composantes S_1, \dots, S_{d^k} que l’on indexe de sorte que $\nu(S_1) \geq \dots \geq \nu(S_{d^k})$. Pour $j = 1, \dots, d^k$, on pose

$$U_j = \bigcup_S S_j.$$

On a ainsi $f^{-1}(U) = U_1 \sqcup \dots \sqcup U_{d^k}$ et $f : U_j \rightarrow U$ est bijective. Comme ν est invariante mais pas de jacobien constant d^k , on peut supposer, quitte à prendre un pavage de $\mathbb{P}^k(\mathbb{C}) \setminus V_f$ par des simplexes suffisamment petits, que $\nu(U_1) > d^{-k}$.

On peut alors trouver $\sigma > d^{-k}$ et un ouvert O relativement compact dans U_1 tels que $\nu(O) > \sigma$. On choisit $\varepsilon > 0$ suffisamment petit pour que le ε -voisinage de O soit contenu dans U_1 . Soit X l’ensemble des points visitant assez souvent O :

$$X = \{x \in \mathbb{P}^k(\mathbb{C}) ; r_n(x) \geq \sigma n \text{ pour } n \geq m\},$$

avec

$$r_n(x) = \text{Card}(\{j \in [0, n - 1] ; f^{\circ j}(x) \in O\}) = \sum_{j=0}^{n-1} \mathbf{1}_O \circ f^{\circ j}(x).$$

D’après le théorème de Birkhoff, on sait que pour ν presque tout x , on a

$$\lim_{n \rightarrow +\infty} \frac{r_n(x)}{n} = \int_{\mathbb{P}^k(\mathbb{C})} \mathbf{1}_O \, d\nu = \nu(O) > \sigma.$$

On peut donc prendre m assez grand pour que $\nu(X) > 0$.

D’après le principe variationnel relatif, on a donc

$$k \log d = h_\nu(f) \leq h_{\text{top}}(f|X).$$

L’argument de Gromov s’adapte pour démontrer que

$$h_{\text{top}}(f|X) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log(\text{Vol}(\Gamma_n|X)_\varepsilon),$$

où $(\Gamma_n|X)_\varepsilon$ est le ε -voisinage de la restriction de Γ_n à X dans Γ_n . On va montrer qu’il existe $\tau < 1$ tel que pour n assez grand, on a

$$\text{Vol}((\Gamma_n|X)_\varepsilon) = \mathcal{O}(d^{\tau kn}).$$

On en déduira alors que

$$k \log d \leq h_{\text{top}}(f|X) \leq \tau k \log d,$$

ce qui donnera la contradiction recherchée.

Pour estimer le volume de $(\Gamma_n|X)_\varepsilon$, on introduit le codage suivant. À un ensemble de volume nul près, $\{U_1, \dots, U_{d^k}\}$ est une partition de $\mathbb{P}^k(\mathbb{C})$. Pour $\alpha \in \{1, \dots, d^k\}^n$, on note

$$U_\alpha^{-n} = \{x \in \mathbb{P}^k(\mathbb{C}) ; f^{\circ j}(x) \in U_{\alpha_j} \text{ pour } j \in [0, n-1]\}$$

et

$$\Gamma_n(\alpha) = \Gamma_n \cap (U_{\alpha_0} \times \dots \times U_{\alpha_{n-1}}).$$

À des ensembles de volume nul près, $\{U_\alpha^{-n}\}_{\alpha \in \{1, \dots, d^k\}^n}$ et $\{\Gamma_n(\alpha)\}_{\alpha \in \{1, \dots, d^k\}^n}$ sont respectivement des partitions de $\mathbb{P}^k(\mathbb{C})$ et de Γ_n . Puisque la restriction de f à chaque U_j est injective, la restriction de $f^{\circ j}$ à chaque U_α^{-n} est injective pour $j \in [0, n-1]$. Le point x appartient à U_α^{-n} si et seulement si sa n -orbite $(x, \dots, f^{n-1}(x))$ appartient à $\Gamma_n(\alpha)$. On a donc

$$\int_{\Gamma_n(\alpha)} \omega_n^{\wedge k} = \int_{U_\alpha^{-n}} (\omega + \dots + (f^{\circ n-1})^* \omega)^{\wedge k}.$$

De plus, si une n -orbite $(y, \dots, f^{n-1}(y)) \in \Gamma_n$ est dans le ε -voisinage de $(\Gamma_n|X)_\varepsilon$, il existe un $x \in X$ tel que $d(f^{\circ j}(x), f^{\circ j}(y)) \leq \varepsilon$ pour tout $j \in [0, n-1]$. Si $f^{\circ j}(x) \in O$ alors $f^{\circ j}(y) \in O$ par choix de ε . Par conséquent, on a l'inclusion

$$(\Gamma_n|X)_\varepsilon \subset \bigcup_{\alpha \in \Sigma_n} \Gamma_n(\alpha),$$

avec

$$\Sigma_n = \{\alpha \in \{1, \dots, d^k\}^n ; \text{Card}(\{j ; \alpha_j = 1\}) \geq \sigma n\}.$$

LEMME 4.10. — *Il existe $\rho < 1$ tel que*

$$\text{Card}(\Sigma_n) \leq (d^{k\rho})^n.$$

PREUVE — On a

$$\text{Card}(\Sigma_n) = \sum_{\sigma n \leq j \leq n} \frac{n!}{j!(n-j)!} (d^k - 1)^{n-j}.$$

Il suffit alors de majorer en utilisant la formule de Stirling. □

Maintenant, on a

$$\begin{aligned} \text{Vol}((\Gamma_n|X)_\varepsilon) &\leq \sum_{\alpha \in \Sigma_n} \int_{\Gamma_n(\alpha)} \omega_n^{\wedge k} \\ &\leq \sum_{\underline{i} \in \{1, \dots, n-1\}^k} \sum_{\alpha \in \Sigma_n} \int_{U_\alpha^{-n}} (f^{\circ i_1})^* \omega \wedge \dots \wedge (f^{\circ i_k})^* \omega. \end{aligned}$$

On choisit λ tel que $\rho < \lambda < 1$. On scinde la somme sur $\underline{i} \in \{1, \dots, n-1\}^k$ en deux parties. L'une sur $\{[\lambda n], \dots, n-1\}^k$ et l'autre sur le complémentaire.

Pour $\underline{i} \in \{[\lambda n], \dots, n-1\}^k$ on a, en posant $q = [\lambda n]$,

$$\begin{aligned} \int_{U_\alpha^{-n}} (f^{\circ i_1})^* \omega \wedge \dots \wedge (f^{\circ i_k})^* \omega &= \int_{U_\alpha^{-n}} (f^{\circ q})^* ((f^{\circ i_1 - q})^* \omega \wedge \dots \wedge (f^{\circ i_k - q})^* \omega) \\ &\leq \int_{\mathbb{P}^k(\mathbb{C})} (f^{\circ i_1 - q})^* \omega \wedge \dots \wedge (f^{\circ i_k - q})^* \omega \\ &= d^{i_1 + \dots + i_k - kq} \leq d^{kn(1-\lambda)}. \end{aligned}$$

La première somme est donc majorée par

$$n^k \text{Card}(\Sigma_n) d^{k(1-\lambda)n} \leq n^k (d^{k(1+\rho-\lambda)})^n.$$

Pour $\underline{i} \notin \{[\lambda n], \dots, n-1\}^k$ on majore globalement la deuxième somme en utilisant le fait que les U_α^{-n} sont deux à deux disjoints. On a

$$\begin{aligned} \sum_{\alpha \in \Sigma_n} \int_{U_\alpha^{-n}} (f^{\circ i_1})^* \omega \wedge \dots \wedge (f^{\circ i_k})^* \omega &\leq \int_{\mathbb{P}^k(\mathbb{C})} (f^{\circ i_1})^* \omega \wedge \dots \wedge (f^{\circ i_k})^* \omega \\ &= d^{i_1 + \dots + i_k} \leq d^{(k-1)(n-1) + \lambda n + 1}. \end{aligned}$$

La deuxième somme est donc majorée par $n^k (d^{k-1+\lambda})^n$. On a donc

$$\text{Vol}((\Gamma_n|X)_\varepsilon) \leq n^k (d^{k(1+\rho-\lambda)})^n + n^k (d^{k-1+\lambda})^n = \mathcal{O}(d^{\tau n})$$

avec

$$\tau = \max\{1 + \rho - \lambda, k - 1 + \lambda\} < 1.$$

□

5. POINTS PÉRIODIQUES RÉPULSIFS

Nous allons maintenant montrer que la mesure d'équilibre reflète la distribution des points périodiques répulsifs. Avant les travaux de Briend et Duval, on ne savait pas montrer que tout endomorphisme de $\mathbb{P}^k(\mathbb{C})$ admet un cycle répulsif. Briend et Duval montrent non seulement qu'il y en a toujours un, mais qu'il y en a une infinité et qu'ils sont denses dans le support de la mesure d'équilibre.

Le lemme clé est le suivant.

LEMME 5.1. — *Soit $\varepsilon > 0$. Il existe l tel que pour tout $x \in \mathbb{P}^k(\mathbb{C}) \setminus V_l$, il existe une boule $B(x, r)$ sur laquelle on peut définir $(1 - \varepsilon)d^{kn}$ branches inverses f_i^{-n} de f^n , d'images B_i^{-n} avec $\sup_i \text{Diam}(B_i^{-n}) \xrightarrow{n \rightarrow +\infty} 0$.*

Remarque 5.2. — Il est possible de montrer qu'on peut définir $(1 - \varepsilon)d^{kn}$ branches inverses d'images B_i^{-n} avec $\text{Diam}(B_i^{-n}) = \mathcal{O}(d^{-n/2})$ (voir par exemple [10], Prop. 3.4.7).

Nous allons d’abord expliquer comment utiliser ce lemme. On pose

$$\nu_n = \frac{1}{d^{nk}} \sum_{f^{\circ n}(y)=y, y \text{ répulsif}} \delta_y.$$

Le nombre total de points périodiques de période n comptés avec multiplicités est $\frac{d^{n(k+1)}-1}{d-1}$. On voit donc que la masse totale de ν_n est bornée. A priori, cette masse pourrait être nulle (s’il n’y avait pas de point périodique répulsif de période n). Considérons une valeur d’adhérence ν de la suite de mesures ν_n . La masse totale de la mesure ν est alors majorée par 1.

Fixons $\varepsilon > 0$. Considérons l comme dans le lemme 5.1 et pour x hors de V_l , soit $B = B(x, r)$ une boule sur laquelle on peut définir $(1 - \varepsilon)d^{kn}$ branches inverses f_i^{-n} de $f^{\circ n}$, d’images B_i^{-n} avec $\sup_i \text{Diam}(B_i^{-n}) \rightarrow 0$ quand $n \rightarrow +\infty$. Nous allons montrer que

$$\nu(B) \geq (1 - 3\varepsilon)^2 \mu_f(B).$$

La mesure d’équilibre μ_f ne charge pas les ensembles algébriques (proposition 1.14). Elle ne charge donc pas l’ensemble V_l . On en déduit que $\nu \geq (1 - 3\varepsilon)^2 \mu_f$. Comme ceci est vrai pour tout $\varepsilon > 0$, on a $\nu \geq \mu_f$. Comme μ_f est une mesure de probabilité et comme la masse de ν est majorée par 1, on en déduit que $\nu = \mu_f$. On a donc

$$\frac{1}{d^{nk}} \sum_{f^{\circ n}(y)=y, y \text{ répulsif}} \delta_y \xrightarrow{n \rightarrow +\infty} \mu_f.$$

Montrons maintenant que $\nu(B) \geq (1 - 3\varepsilon)^2 \mu_f(B)$. On doit montrer que pour n assez grand, le nombre de points périodiques répulsifs de période n contenus dans B est supérieur à $d^{nk}(1 - 3\varepsilon)^2 \mu_f(B)$. Pour cela, il suffit de montrer que le nombre de boules B_i^{-n} relativement compactes dans B est supérieur à $d^{nk}(1 - 3\varepsilon)^2 \mu_f(B)$. En effet, si B_i^{-n} est relativement compacte dans B , alors $f_i^{-n} : B \rightarrow B_i^{-n}$ est contractante et a un point fixe attractif dans B_i^{-n} . Ce point est un point périodique répulsif de période n pour f . Comme les B_i^{-n} sont deux à deux disjointes, les points périodiques répulsifs ainsi obtenus sont tous distincts.

Donnons-nous une boule B' relativement compacte dans B telle que $\mu_f(B') \geq (1 - \varepsilon)\mu_f(B)$. Si n est assez grand, le diamètre des boules B_i^{-n} est petit, et toute boule B_i^{-n} qui intersecte B' est relativement compacte dans B . Il suffirait donc de montrer que la boule B' contient au moins $d^{nk}(1 - 3\varepsilon)^2 \mu_f(B)$ points $x_i^{-n} = f_i^{-n}(x)$. Comme la mesure d’équilibre reflète la distribution des préimages de x , la boule B' contient au moins $d^{nk}(1 - 2\varepsilon)\mu_f(B)$ préimages n -ièmes de x pour n assez grand. Le problème majeur est que nous avons «jeté» εd^{nk} de ces préimages. Étant donné que la taille de la boule B est déterminée par ε , rien n’empêche $\mu_f(B)$ d’être de l’ordre de grandeur de ε , et donc, il se pourrait qu’aucun des points x_i^{-n} ne se trouve dans la boule B' .

On peut cependant s’en sortir en procédant comme suit. On pose $\varepsilon' = \varepsilon \mu_f(B)$. Pour cet ε' , il existe l' tel que pour tout $y \in \mathbb{P}^k(\mathbb{C}) \setminus V_{l'}$, il existe une boule $B(y, r'_y)$ sur laquelle on peut définir $(1 - \varepsilon')d^{km}$ branches inverses $f^{\circ m}$ dont les images ont un petit diamètre. La mesure μ_f ne charge pas les ensembles algébriques. On peut donc trouver

un voisinage ouvert U' de V_l' dont la masse est inférieure à ε (le ε de départ). Comme $\mathbb{P}^k(\mathbb{C}) \setminus U'$ est compact, on peut le recouvrir par un nombre fini de boules $B(y, r'_y/2)$. On choisit alors n assez grand pour que

- les boules B_i^{-n} aient un diamètre suffisamment petit de sorte que

$$x_i^{-n} \in B(y, r'_y/2) \implies B_i^{-n} \subset B(y, r'_y)$$

(ceci est possible car on a un nombre fini de boules $B(y, r'_y/2)$ et que le diamètre des boules B_i^{-n} tend vers 0 quand n tend vers $+\infty$) et

- le nombre de préimages n -ièmes de x contenues dans le voisinage U' de V_l' soit inférieur à $2\varepsilon d^{nk}$ (ceci est possible car $\mu_{n,x}(U') \rightarrow \mu_f(U') \leq \varepsilon$).

On a alors $(1-3\varepsilon)d^{nk}$ branches inverses f_i^{-n} de $f^{\circ n}$ sur B , telles que pour tout i , la boule $B_i^{-n} = f_i^{-n}(B)$ est contenue dans une des boules $B(y, r'_y)$. Pour chaque i et pour m assez grand, on peut alors construire $(1-\varepsilon')d^{mk}$ branches inverses de $f^{\circ m}$ sur B_i^{-n} d'images $B_j^{-(n+m)}$ ayant un diamètre strictement inférieur à la distance de B' au bord de B . Utilisons le fait que $\mu_{m,x_i^{-n}} \rightarrow \mu_f$ quand $m \rightarrow +\infty$. Quand $m \rightarrow +\infty$, la proportion de préimages m -ièmes de x_i^{-n} contenues dans B' tend vers $\mu_f(B') \geq (1-\varepsilon)\mu_f(B)$. Pour m assez grand, le nombre de préimages m -ièmes de x_i^{-n} contenues dans B' est donc supérieur à $(1-2\varepsilon)\mu_f(B)d^{mk}$. On en a «jeté» $\varepsilon'd^{mk} = \varepsilon\mu_f(B)d^{mk}$. Il y a donc au moins $(1-3\varepsilon)\mu_f(B)d^{mk}$ branches inverses de $f^{\circ m}$ sur B_i^{-n} qui ont des images relativement compactes dans B .

En faisant le total sur les $(1-3\varepsilon)d^{nk}$ points x_i^{-n} , on obtient $(1-3\varepsilon)d^{nk} \cdot (1-3\varepsilon)\mu_f(B)d^{mk}$ branches inverses $f_j^{-(n+m)}$ de $f^{\circ(n+m)}$ sur B d'images $B_j^{-(n+m)}$ relativement compactes dans B . Ces branches inverses produisent, comme expliqué ci-dessus, $(1-3\varepsilon)^2\mu_f(B)d^{(n+m)k}$ points répulsifs de période $n+m$ dans B et on a donc, pour n et m assez grands,

$$\nu_{n+m}(B) \geq (1-3\varepsilon)^2\mu_f(B).$$

Venons-en maintenant à la démonstration du lemme 5.1. On procède essentiellement comme dans le lemme 3.6, avec cependant quelques modifications. La démonstration que nous présentons nous a été expliquée par Julien Duval.

PREUVE DU LEMME 5.1 — Soient $\varepsilon > 0$ et $l \geq 1$ grand. Soient $x \in \mathbb{P}^k(\mathbb{C}) \setminus V_l$ et r_0 assez petit pour que $B_0 = B(x, r_0)$ évite V_l . On peut alors définir d^{kl} branches inverses f_i^{-l} de $f^{\circ l}$ sur B_0 .

On ne va pas montrer qu'une grosse proportion des boules $B_i^{-l} = f_i^{-l}(B_0)$ n'intersecte pas l'ensemble V_f des valeurs critiques. Mais on va montrer que c'est vrai si on s'autorise à diminuer le rayon de la boule de départ. Considérons la suite décroissante r_n définie par

$$r_{n+1} = r_n - \frac{r_0}{2(n+1)^2}.$$

Cette suite converge vers $r = r_0(1 - \pi^2/12) > 0$. Posons $B_1 = B(x, r_1)$ et supposons que $f_i^{-l}(B_1)$ intersecte V_f . Alors, d’après le théorème de Lelong, on a

$$\text{Vol}(f^{ol}(f_i^{-l}(B_1) \cap V_f)) \geq c(r_0 - r_1)^{2(k-1)},$$

où c est une constante universelle. Or, ce volume est donné par l’intégrale

$$\int_{f_i^{-l}(B_1) \cap V_f} (f^{ol})^* \omega^{\wedge k-1}.$$

La somme pour toutes les branches inverses f_i^{-l} est donc majorée par

$$\int_{V_f} (f^{ol})^* \omega^{\wedge k-1} = \tau d^{l(k-1)},$$

où τ est le degré de V_f . Parmi les d^{lk} ensembles $f_i^{-l}(B_1)$, il y en a au plus

$$\frac{\tau d^{l(k-1)}}{c(r_0 - r_1)^{2(k-1)}}$$

qui intersectent V_f . Les autres contribuent chacun à d^{lk} branches inverses de f^{ol+1} sur B_1 . On a donc

$$d^{lk} \left(1 - \frac{\tau}{cd^l(r_0 - r_1)^{2(k-1)}} \right)$$

branches inverses de f^{ol+1} sur B_{l+1} . En procédant par récurrence, on montre que pour $n \geq l$, il y a au moins

$$d^{kn} \left(1 - \frac{\tau 2^{2(k-1)}}{cd^l r_0^{2(k-1)}} \left(1 + \frac{2^{4(k-1)}}{d} + \dots + \frac{(n-l)^{4(k-1)}}{d^{n-l-1}} \right) \right)$$

branches inverses de f^{on} sur $B(x, r_{n-l})$. Si on prend l assez grand, on a donc $(1 - \varepsilon/2)d^{kn}$ branches inverses f_i^{-n} de f^{on} sur $B(x, r)$.

Nous allons maintenant contrôler le diamètre des boules $B_i^{-n} = f_i^{-n}(B(x, r'))$ pour $r' < r$. Donnons-nous une famille dénombrable de droites L_j passant par x telles que la réunion des disques $\Delta_j = L_j \cap B(x, r)$ soit dense dans $B(x, r)$. Pour chaque n et pour chaque $j \leq n$, on oublie les branches inverses pour lesquelles

$$\text{Aire}(f_i^{-n}(\Delta_j)) \geq \frac{2n}{\varepsilon d^n}.$$

On en oublie au plus n fois $\varepsilon d^{nk}/(2n)$ (voir la preuve du lemme 3.6). Donc, il reste $(1 - \varepsilon)d^{kn}$ branches inverses f_i^{-n} . Ces branches inverses sont telles que pour $j \leq n$, on a $\text{Aire}(f_i^{-n}(\Delta_j)) = \mathcal{O}(nd^{-n})$.

Pour $n \geq 1$, on a donc gardé $(1 - \varepsilon)d^{nk}$ branches inverses f_i^{-n} telles que

$$\text{Aire}(f_i^{-n}(\Delta_j)) \xrightarrow{n \rightarrow +\infty} 0.$$

Les branches inverses $f_i^{-n} : B(x, r) \rightarrow \mathbb{P}^k(\mathbb{C})$ forment une famille normale (l’argument est le même que celui de la remarque 3.8). Les valeurs d’adhérence sont constantes le long de chaque disque Δ_j . Elles sont donc constantes sur $B(x, r)$ puisque les disques Δ_j sont denses dans $B(x, r)$. □

REMERCIEMENTS

Je tiens à remercier tous ceux qui m'ont aidé à préparer cet exposé, notamment François Berteloot, Jean-Yves Briend, Arnaud Chéritat, Adrien Douady, Julien Duval, John H. Hubbard et Vincent Guedj. Je tiens également à remercier Nessim Sibony qui a suggéré de nombreuses améliorations pour la version finale.

RÉFÉRENCES

- [1] E. BEDFORD & J. SMILLIE – *Polynomial diffeomorphisms of \mathbf{C}^2 : Currents, equilibrium measure and hyperbolicity*, Invent. Math. **87** (1990) 69–99.
- [2] E. BEDFORD & J. SMILLIE – *Polynomial diffeomorphisms of \mathbf{C}^2 (III)*, Math. Ann **294** (1992) 395–420.
- [3] E. BEDFORD, M. LYUBICH & J. SMILLIE – *Polynomial diffeomorphisms of \mathbf{C}^2 (V), The measure of maximal entropy and laminar currents*, Invent. Math. **112** (1993) 77–125.
- [4] E. BEDFORD & B. TAYLOR – *A new capacity for plurisubharmonic functions*, Acta Math. **149** (1982), 1–39.
- [5] J.-Y. BRIEND, S. CANTAT & M. SHISHIKURA – *Linearity of the exceptional set for maps of $\mathbb{P}^k(\mathbb{C})$* , Math. Ann. **330** (2004), 39–43.
- [6] J.-Y. BRIEND & J. DUVAL – *Exposants de Liapounoff et distribution des points périodiques d'un endomorphisme de $\mathbb{C}\mathbb{P}^k$* , Acta Math. **182** (1999), 143–157.
- [7] J.-Y. BRIEND & J. DUVAL – *Deux caractérisations de la mesure d'équilibre d'un endomorphisme de $\mathbb{P}^k(\mathbb{C})$* , Publ. Math. IHÉS **93** (2001) 145–159.
- [8] M. BRIN & A. KATOK – *On local entropy*, dans Geometric dynamics, Lect. Notes in Math. **1007**, Springer-Verlag (1983), 30–38.
- [9] H. BROLIN – *Invariant sets under iteration of rational functions*, Arkiv Math. **6** (1965) 103–144.
- [10] T.-C. DINH & N. SIBONY – *Dynamique des applications d'allure polynomiale*, J. Math. Pures Appl. **82** (2003) 367–423.
- [11] T.-C. DINH & N. SIBONY – *Distribution des valeurs de transformations méromorphes et applications*, Preprint arXiv: math.DS/0306095.
- [12] J.E. FORNÆSS & N. SIBONY – *Complex dynamics in higher dimension I*, Astérisque **222** (1994) 201–231.
- [13] J.E. FORNÆSS & N. SIBONY – *Complex dynamics in higher dimension, II*, Modern Methods in Complex Analysis, Princeton N.J 1992, Ann. Math. Studies **137** (1995) 135–187.
- [14] J.E. FORNÆSS & N. SIBONY – *Complex dynamics in higher dimension*, Notes partially written by Estela A. Gavosto. NATO Adv. Inst. Ser. C Math. Phys. Sci., 439, Complex potential theory (Montreal, PQ, 1993), Kluwer Acad. Press, Dordrecht (1994) 131–186.

- [15] J.E. FORNÆSS & N. SIBONY – *Dynamics of \mathbf{P}^2 (Examples)*, Contemporary Mathematics **269** (2001) 47–85.
- [16] A. FREIRE, A. LOPES & R. MAÑE – *An invariant measure for rational maps*, Bol. Soc. Brasil. Mat. **14** (1983), 45–62.
- [17] M. GROMOV – *On the entropy of holomorphic maps*. Manuscrit (1977) publié dans l'Enseignement Mathématique **49** (2003), 217–235.
- [18] V. GUEDJ – *Ergodic properties of rational mappings with large topological degree*, à paraître dans Annals of Math. (2005).
- [19] J.H. HUBBARD & P. PAPADOPOULOS – *Superattractive fixed points in \mathbb{C}^n* , Indiana Univ. Math. J. **43** (1994) 321–365.
- [20] P. LELONG – *Propriétés métriques des variétés analytiques complexes définies par une équation*, Ann. Sci. Éc. Norm. Sup. **67** (1950), 393–419.
- [21] M. LYUBICH – *Entropy properties of rational endomorphisms of the Riemann sphere*, Erg. Th. and Dyn. Sys. **3** (1983), 351–385.
- [22] R. MAÑE – *On the uniqueness of the maximizing measure for rational maps*, Bol. Soc. Brasil. Mat. **14** (1983), 27–43.
- [23] M. MISIUREWICZ & F. PRZYTYCKI – *Topological entropy and degree of smooth mappings*, Bull. Acad. Polon. Sci., Ser. Math. Astron. et Phys. **25** (1977), 573–574.
- [24] W. PARRY – *Entropy and generators in ergodic theory*, Benjamin Press (1969).
- [25] N. SIBONY – *Dynamique des applications rationnelles de \mathbf{P}^k* , Dynamique et géométrie complexes (Lyon, 1997), Panor. Synthèses, **8**, Soc. Math. France, Paris (1999), 97–185.
- [26] P. TORTRAT – *Aspects potentialistes de l'itération des polynômes*, Lecture Notes in Math. **1255** Springer (1987), 195–209.
- [27] T. UEDA – *Fatou sets in complex dynamics on projective spaces*, J. Math. Soc. Japan **46** (1994), no. 3, 545–555.

Xavier BUFF

Université Paul Sabatier
 Laboratoire Émile Picard
 UMR 5580 du CNRS
 UFR MIG

118, route de Narbonne,
 F-31062 TOULOUSE Cédex

E-mail : buff@picard.ups-tlse.fr

**LEMME FONDAMENTAL ET ENDOSCOPIE,
UNE APPROCHE GÉOMÉTRIQUE**
[d'après Gérard Laumon et Ngô Bao Châu]

par **Jean-François DAT**

Malgré son nom, le «lemme fondamenta» de Langlands et Shelstad est un énoncé conjectural, ou plutôt une famille d'énoncés conjecturaux. Il est de nature locale au sens arithmétique, *i.e.* il concerne des objets relatifs à un corps local, mais est apparu dans les deux problèmes majeurs, de nature globale, du fameux «programme de Langlands» : le principe de fonctorialité et l'expression des fonctions L de variétés de Shimura en termes automorphes. En effet, les méthodes initiées par Langlands pour chacun de ces problèmes reposent sur la formule des traces d'Arthur-Selberg : on cherche à comparer deux telles formules pour deux groupes différents dans le cas de la fonctorialité, ou bien une seule, pour un groupe donné, avec la formule des traces de Grothendieck-Lefschetz pour sa variété de Shimura associée. Mais la formule des traces d'Arthur-Selberg n'est pas en général le bon outil pour appliquer cette méthode, elle est «instable» (*cf.* 1.3). Pour la stabiliser, Langlands [25], à la suite de travaux avec Labesse [24] et Shelstad, a créé la théorie d'endoscopie qui repose sur deux conjectures «locales», relevant de l'analyse harmonique sur les groupes p -adiques : le *transfert* et le *lemme fondamental*. Depuis vingt-cinq ans, beaucoup de mathématiciens ont été écrites, soit pour résoudre ces conjectures, soit pour aller au-delà, en les admettant.

Expliquons brièvement la nature de ces conjectures en renvoyant le lecteur à 1.6 pour un énoncé détaillé. Langlands et Shelstad ont défini la notion de groupe endoscopique d'un groupe p -adique G et pour un tel groupe, disons H , deux familles de distributions dépendant d'une classe de conjugaison «stable» γ semi-simple et « G -régulière» dans H : l'intégrale orbitale «stable» SO_γ sur H et l'intégrale orbitale «endoscopique» $O_\gamma^{G|H}$ sur G . La conjecture de *transfert* prévoit l'existence, pour chaque fonction lisse à support compact f_G sur G , d'une fonction semblable f_H sur H telle que $SO_\gamma(f_G) = O_\gamma^{G|H}(f_H)$ pour tout γ . Lorsque les groupes G et H sont quasi déployés et non-ramifiés, et donc munis de sous-groupes «hyperspéciaux» (*cf.* 1.1), le *lemme fondamental* prévoit, lui, que lorsque f_G est la fonction caractéristique d'un sous-groupe «hyperspécial» de G , alors on peut prendre pour f_H la fonction caractéristique d'un sous-groupe «hyperspécial» de H .

Waldspurger a introduit une variante du lemme fondamental où les intégrales orbitales sont prises sur les algèbres de Lie et non sur les groupes. Ses travaux [35] et [37] ont la conséquence remarquable suivante (*cf.* 1.7) : pour résoudre les conjectures de

transfert de Langlands-Shelstad relatives à un groupe réductif G sur un corps p -adique, il suffit de savoir résoudre le lemme fondamental dans sa version «algèbres de Lie» pour toute forme quasi déployée non-ramifiée du système de racines absolu du centralisateur connexe d'un élément semi-simple de G , sur un corps $\mathbb{F}_{p^d}((T))$ avec p' «grand».

Ce dernier problème a l'avantage de pouvoir être traduit en termes géométriques. En effet les distributions SO_γ et $O_\gamma^{G|H}$ sont des combinaisons linéaires d'intégrales orbitales ; dans le cas d'un groupe classique –associé à un espace vectoriel muni d'une forme symplectique, symétrique ou hermitienne– évaluer l'intégrale orbitale au point x de l'algèbre de Lie en la fonction caractéristique d'un compact hyperspécial revient à compter certains réseaux stables sous x et qui sont «autoduaux» pour la forme considérée. Lorsqu'on est sur un corps de séries formelles $F = k((T))$, les réseaux sont «classifiés» par la grassmannienne affine (une ind-variété sur k) et ceux qui sont stables sous x par la *fibres de Springer affine* telle que définie par Kazhdan et Lusztig dans [16]. Ainsi, Goresky, Kottwitz et Macpherson ont interprété dans [11] l'évaluation de SO_γ , resp. de $O_\gamma^{G|H}$, en la fonction caractéristique d'un compact hyperspécial de $\text{Lie}(H)$, resp. de $\text{Lie}(G)$, comme la trace alternée de Frobenius sur certains morceaux des groupes de cohomologie d'un quotient de la fibre de Springer affine de H en γ , resp. des groupes de cohomologie d'un système local «endoscopique» sur un quotient de la fibre de Springer affine de G en un élément $\gamma' \in \text{Lie}(G)$ associé à γ . Le problème est alors de comparer ces morceaux de groupes de cohomologie ! Les trois auteurs précédents ont résolu, modulo une conjecture de pureté cohomologique des fibres de Springer affines, le cas très particulier où le tore centralisateur de γ est non ramifié, et ont introduit un outil qui sera fondamental pour tous les travaux ultérieurs : la cohomologie équivariante sous un certain tore adapté à la situation. Laumon [28] a ensuite résolu le cas où G est un groupe unitaire, sans restriction sur γ mais toujours sous la conjecture de pureté, en introduisant de manière *ad hoc* un argument de déformation.

Ces approches géométriques sont de nature purement locales, et butent sur un obstacle, la conjecture de pureté, qui ne sera pas forcément facile à lever. Récemment, Ngô [33] a introduit une nouvelle approche géométrique, de nature globale, basée sur la fibration de Hitchin [15] introduite par ce dernier dans un tout autre contexte. Elle permet d'interpréter très naturellement le procédé de *pré-stabilisation* de Langlands-Kottwitz, cf. 1.4, qui est historiquement la motivation *globale* à l'origine du lemme fondamental. Elle fournit à la fois un moyen de contourner le problème de la pureté et de déformer de manière naturelle une situation locale donnée. Cette approche vaut *a priori* pour un groupe général, mais n'a pour l'instant été poussée jusqu'au bout, c'est-à-dire jusqu'à la preuve du lemme fondamental, que dans le cas des groupes unitaires dans l'article [30] de Laumon et Ngô. Signalons enfin qu'elle utilise de manière cruciale le langage des champs algébriques.

Le but principal de l'exposé est de présenter l'approche de Laumon et Ngô. Cependant, comme l'énoncé du lemme fondamental n'a rien d'immédiatement «naturel», il a paru souhaitable de commencer par une présentation succincte de la théorie

d'endoscopie et de ses motivations. On évoquera d'ailleurs autant la théorie globale que locale, puisque l'approche de Ngô est de nature globale.

L'auteur remercie P.-H. Chaudouard, J.-P. Labesse, G. Laumon et Ngô B.C. pour leur aide dans la préparation de cet exposé.

1. ENDOSCOPIE : UNE BRÈVE INTRODUCTION

1.1. Notations

Nous aurons à considérer deux contextes différents, l'un global et l'autre local, pour lesquels certaines notations seront similaires.

Contexte local : F est un corps local non-archimédien, donc soit $\mathbb{F}_q((T))$, soit une extension finie de \mathbb{Q}_p . Un groupe réductif connexe \mathbf{G} sur F est quasi-déployé et non-ramifié (*i.e.* déployé sur une extension non-ramifiée) si et seulement si \mathbf{G} se prolonge en un groupe lisse à fibres réductives connexes sur l'anneau des entiers \mathcal{O}_F . Un sous-groupe *hypersécial* de $\mathbf{G}(F)$ est alors par définition un groupe de la forme $\mathbf{G}(\mathcal{O}_F)$ pour un tel prolongement. C'est un sous-groupe compact maximal du groupe localement compact $\mathbf{G}(F)$. Sous ces hypothèses, on normalisera toujours les mesures de Haar sur ce dernier de sorte que la mesure d'un sous-groupe hypersécial soit 1.

Contexte global : F est un corps global, donc une extension finie de \mathbb{Q} ou de $\mathbb{F}_q(T)$, dont on note \mathcal{O}_F l'anneau des entiers et $\mathcal{P}(F)$ l'ensemble des places. Pour toute place $v \in \mathcal{P}(F)$, on note F_v le complété en v et \mathcal{O}_v son anneau d'entiers. On note aussi $\mathbb{A}_F := \prod'_v F_v$ l'anneau topologique localement compact des adèles.

Soit \mathbf{G} un groupe réductif connexe sur F . On peut le prolonger en un groupe lisse à fibres réductives connexes sur $\mathcal{O}_F[\frac{1}{N}]$ pour un élément non-nul N de \mathcal{O}_F . On peut alors parler de $\mathbf{G}(\mathcal{O}_v)$ pour $v \in \mathcal{P}(F)$ ne divisant pas N . On définit $\mathbf{G}(\mathbb{A}_F)$ comme le produit restreint $\prod'_v \mathbf{G}(F_v)$ relativement aux $\mathbf{G}(\mathcal{O}_v)$. C'est un groupe localement compact. On le munira de la mesure «canonique» de Tamagawa dont la définition précise n'a pas d'importance ici, et qui se décompose en un produit $dx = \prod'_v dx_v$ de mesures de Haar locales qui, aux places v ne divisant pas N , sont normalisées comme au paragraphe précédent (sauf éventuellement en une place si $N = 1$ et F est un corps de fonctions).

Rappelons que le plongement diagonal $\mathbf{G}(F) \hookrightarrow \mathbf{G}(\mathbb{A}_F)$ fait de $\mathbf{G}(F)$ un sous-groupe discret de $\mathbf{G}(\mathbb{A}_F)$ qui est de covolume fini si et seulement si le centre \mathbf{Z} de \mathbf{G} est anisotrope, et qui est cocompact si et seulement si \mathbf{G} lui-même est anisotrope.

Enfin nous fixerons une clôture séparable \overline{F} de F et noterons Γ le groupe de Galois correspondant et Γ_v le groupe de décomposition en $v \in \mathcal{P}(F)$. Nous supposerons toujours que la caractéristique de F est «assez grande pour \mathbf{G} » (première à l'ordre de son groupe de Weyl en particulier) pour éviter les soucis d'inséparabilité.

1.2. Quelques mots sur la formule des traces

Le contexte ici est *global*. Supposons, pour simplifier, le centre \mathbf{Z} de \mathbf{G} anisotrope sur F . Soit $\mathcal{C}_c^\infty(\mathbf{G}(\mathbb{A}_F))$ l'espace des fonctions complexes lisses à support compact sur $\mathbf{G}(\mathbb{A}_F)$. De telles fonctions sont toujours de la forme $f_S \otimes (\otimes_{v \in \mathcal{P}(F) \setminus S} \mathbf{1}_{\mathbf{G}(\mathcal{O}_v)})$ pour un ensemble fini $S \subset \mathcal{P}(F)$ et une fonction $f_S \in \mathcal{C}_c^\infty(\prod_{v \in S} \mathbf{G}(F_v))$. Si $f \in \mathcal{C}_c^\infty(\mathbf{G}(\mathbb{A}_F))$, la représentation ρ de $\mathbf{G}(\mathbb{A}_F)$ sur l'espace $L^2(\mathbf{G}(F) \backslash \mathbf{G}(\mathbb{A}_F))$ et la mesure dx induisent un opérateur $\rho(f)$, intégral de noyau

$$K_f(x, y) = \sum_{\gamma \in \mathbf{G}(F)} f(x^{-1}\gamma y).$$

Lorsque \mathbf{G} est F -anisotrope, et donc $\mathbf{G}(F) \backslash \mathbf{G}(\mathbb{A}_F)$ est compact, l'opérateur $\rho(f)$ est traçable et sa trace est l'intégrale de son noyau sur la diagonale. Un simple calcul fournit alors la «formule des traces» pour $\mathbf{G}(\mathbb{A}_F)$:

$$\begin{aligned} \text{Tr}(\rho(f)) &= \sum_{\gamma \in \mathbf{G}(F)/\text{conj}} \int_{\mathbf{G}_\gamma(F) \backslash \mathbf{G}(\mathbb{A}_F)} f(x^{-1}\gamma x) dx \\ &= \sum_{\gamma \in \mathbf{G}(F)/\text{conj}} \text{vol}_{dy}(\mathbf{G}_\gamma(F) \backslash \mathbf{G}_\gamma(\mathbb{A}_F)) \int_{\mathbf{G}_\gamma(\mathbb{A}_F) \backslash \mathbf{G}(\mathbb{A}_F)} f(x^{-1}\gamma x) dx/dy. \end{aligned}$$

Nous avons noté \mathbf{G}_γ le (groupe algébrique) centralisateur de γ dans \mathbf{G} . Notons que l'extraction du volume à la seconde ligne permet d'utiliser des calculs locaux, puisque lorsque $f = \otimes_v f_v$, on a le produit (dont presque tous les facteurs sont égaux à 1)

$$\int_{\mathbf{G}_\gamma(\mathbb{A}_F) \backslash \mathbf{G}(\mathbb{A}_F)} f(x^{-1}\gamma x) dx/dy = \prod_v \int_{\mathbf{G}_\gamma(F_v) \backslash \mathbf{G}(F_v)} f_v(x_v^{-1}\gamma x_v) dx_v/dy_v.$$

Le terme de gauche est l'intégrale orbitale globale de f sur la classe de conjugaison de γ , et sera encore noté $O_\gamma(f)$. De même, les intégrales orbitales locales du terme de droite seront notées $O_\gamma(f_v)$, $v \in \mathcal{P}(F)$. Enfin on posera $\tau(\mathbf{G}_\gamma) := \text{vol}_{dy}(\mathbf{G}_\gamma(F) \backslash \mathbf{G}_\gamma(\mathbb{A}_F))$; de par nos conventions, c'est le nombre de Tamagawa de \mathbf{G}_γ , si celui-ci est connexe. Par la suite on supposera souvent \mathbf{G}_{der} simplement connexe, ce qui assurera la connexité des centralisateurs.

Lorsque \mathbf{G} n'est plus supposé F -anisotrope, la formule des traces est beaucoup plus compliquée et n'est d'ailleurs pas une expression de la trace d'un opérateur. Il s'agit tout de même d'une égalité entre deux distributions sur $\mathbf{G}(\mathbb{A}_F)$, l'une renfermant des informations spectrales, comme la trace sur le spectre automorphe discret, l'autre des informations «géométriques» comme les intégrales orbitales le long de classes de conjugaison semi-simples de $\mathbf{G}(F)$. Un peu plus précisément, le côté géométrique s'écrit :

$$\sum_{\gamma \in \mathbf{G}(F)^{\text{ell}}/\text{conj}} \tau(\mathbf{G}_\gamma) O_\gamma(f) + \text{autres termes.}$$

Nous ne nous occuperons pas dans cet exposé des «autres termes», qui sont importants pour le but ultime mais qui pour certaines applications peuvent être annulés en choisissant bien la fonction-test f (cf. [19] par exemple). Puisqu'il faut définir l'ensemble

$\mathbf{G}(F)^{ell}$ qui apparaît dans la sommation ci-dessus, on en profite pour introduire un peu de vocabulaire. Un élément $\gamma \in \mathbf{G}(F)$ est dit

- *régulier* si son centralisateur est de dimension minimale (égale au rang de \mathbf{G}).
- *fortement régulier* si ce centralisateur est un tore.
- *elliptique* s’il est contenu dans un F -tore F -anisotrope « modulo le centre »⁽¹⁾. Un tel tore sera lui-même dit « elliptique » par la suite.

Par exemple, lorsque $\mathbf{G} = GL(N)$, un élément de $\mathbf{G}(F)$ est semi-simple régulier et elliptique si et seulement si son polynôme caractéristique est irréductible. Les tores elliptiques de $GL(N)$ sont de la forme $\text{Res}_{F'|F} \mathbf{G}_m$ pour une extension séparable $F'|F$ de degré N . Si $\mathbf{G} = U_{E|F}(N)$ est le groupe unitaire (quasi déployé) associé à l’extension quadratique E de F , alors les tores elliptiques sont de la forme $\prod_{i=1}^r \text{Res}_{F_i|F} U_{E_i|F_i}(1)$ où les F_i sont des extensions séparables de F disjointes de E telles que $\sum_{i=1}^r \deg(F_i|F) = N$.

1.3. Pourquoi la stabilisation

Tentons d’expliquer brièvement l’origine des problèmes que nous allons présenter par la suite. La discussion qui suit est de nature informelle et contient des simplifications nécessairement abusives. Pour plus de détails et plus de rigueur, le lecteur pourra consulter [8] sur les formes et représentations automorphes et [7] sur les L -groupes et la functorialité.

Une représentation *automorphe* de $\mathbf{G}(\mathbb{A}_F)$ est par définition un sous-quotient irréductible de la représentation régulière de ce groupe dans un certain espace de fonctions sur $\mathbf{G}(F) \backslash \mathbf{G}(\mathbb{A}_F)$ qui sont dites aussi *automorphes* et qui généralisent la notion classique de forme modulaire que l’on a lorsque $\mathbf{G} = GL(2)$. De même que ces dernières, les formes automorphes recèlent des propriétés arithmétiques cachées et très intéressantes. Notons qu’une représentation automorphe se décompose de manière essentiellement unique en un produit tensoriel « restreint » $\Pi = \bigotimes'_v \pi_v$ de représentations irréductibles des groupes locaux $\mathbf{G}(F_v)$.

Le principe de functorialité que Langlands a imaginé en 1967 permet de transférer ces représentations d’un groupe à un autre. À \mathbf{G} réductif connexe sur F on associe un groupe algébrique complexe ${}^L\mathbf{G} := \widehat{\mathbf{G}} \rtimes \Gamma$ où $\widehat{\mathbf{G}}$ est « le » groupe réductif connexe sur \mathbb{C} de système de racines basé *dual* de celui de \mathbf{G} , et Γ agit par automorphismes extérieurs, via l’action sur le système de racines basé que l’on déduit de la F -structure rationnelle sur \mathbf{G} . Ainsi, ${}^L\mathbf{G}$ ne détermine pas \mathbf{G} mais seulement sa forme intérieure quasi déployée. Le principe de functorialité prévoit que tout morphisme de L -groupes ${}^L\mathbf{H} \rightarrow {}^L\mathbf{G}$ doit induire un transfert des représentations automorphes de \mathbf{H} vers celles de \mathbf{G} , qui est décrit explicitement sur *presque tous* les facteurs locaux π_v (cf. [7, 16.2]). Parmi les conséquences mirifiques qu’aurait (un cas particulier de) cet énoncé, on peut citer la conjecture d’Artin sur les fonctions L de représentations galoisiennes.

⁽¹⁾Avec cette définition, un élément elliptique est en particulier semi-simple.

Langlands a aussi initié une stratégie pour attaquer la fonctorialité dans certains contextes favorables⁽²⁾ qui repose sur la formule des traces d’Arthur-Selberg évoquée au paragraphe précédent. Dans certaines situations de transfert, on peut faire correspondre les classes de conjugaison semi-simples des deux groupes $\mathbf{H}(\overline{F})$ et $\mathbf{G}(\overline{F})$, et ce de manière Γ -équivariante. C’est par exemple le cas lorsque le morphisme de L -groupes induit un morphisme Γ -équivariant $\widehat{\mathbf{T}}_{\mathbf{H}}/W_{\widehat{\mathbf{H}}} \rightarrow \widehat{\mathbf{T}}_{\mathbf{G}}/W_{\widehat{\mathbf{G}}}$ où $\mathbf{T}_{\mathbf{H}}$ et $\mathbf{T}_{\mathbf{G}}$ sont des F -tores maximaux de \mathbf{H} et \mathbf{G} et $W_{\widehat{\mathbf{H}}}$, $W_{\widehat{\mathbf{G}}}$ les groupes de Weyl correspondants.

Conjugaison stable : supposons pour simplifier que \mathbf{G}_{der} est simplement connexe et disons que deux éléments semi-simples de $\mathbf{G}(F)$ sont *stablement conjugués* s’ils sont conjugués dans $\mathbf{G}(\overline{F})$. Par exemple, si $\mathbf{G} = GL(N)$ ou $U_{E|F}(N)$, deux éléments semi-simples sont *stablement* conjugués si et seulement si ils ont même polynôme caractéristique. Dans le cas de $GL(N)$, ils sont alors conjugués, mais dans le cas $U_{E|F}(N)$, ils ne le sont pas nécessairement.

Par la discussion précédente, un morphisme des L -groupes induit parfois une correspondance entre classes de conjugaison *stables* d’éléments semi-simples de $\mathbf{G}(F)$ et $\mathbf{H}(F)$. L’idée est alors de déduire le transfert spectral que l’on cherche d’un transfert géométrique des intégrales orbitales, grâce à la formule des traces d’Arthur-Selberg. Malheureusement cette dernière, dans sa forme originale, n’est pas *stablement* invariante. On est donc amené à la modifier (la « stabiliser ») pour l’adapter aux besoins de la stratégie de Langlands. Ce travail a été entrepris par Labesse, Langlands [25], prolongé par Kottwitz [18][17], Kottwitz-Shelstad [21], Labesse [23][22], et bien sûr, Arthur [3], [2] et [4].

1.4. La pré-stabilisation de Langlands-Kottwitz

À titre d’exemple et de motivation pour le lemme fondamental, mais aussi pour comprendre la pertinence de l’approche géométrique de Ngô, décrivons la première partie de la stratégie de Langlands pour stabiliser la contribution

$$T_G^{regell}(f) := \sum_{\gamma \in \mathbf{G}(F)_{conj}^{regell}} \tau(\mathbf{G}_{\gamma}) O_{\gamma}(f), \quad f = \otimes'_v f_v \in \mathcal{C}_c^{\infty}(\mathbf{G}(\mathbb{A}_F))$$

des termes elliptiques réguliers au côté géométrique de la formule des traces. On suppose \mathbf{G}_{der} simplement connexe, pour simplifier les énoncés. Commençons par la remarque suivante : la classe de conjugaison de γ (elliptique régulier) s’identifie à $\mathbf{G}(F)/\mathbf{G}_{\gamma}(F)$ tandis que la classe *stable* s’identifie à $\mathbf{G}/\mathbf{G}_{\gamma}(F)$ (points rationnels de la F -variété quotient $\mathbf{G}/\mathbf{G}_{\gamma}$). Il est donc naturel de définir l’intégrale orbitale stable :

$$SO_{\gamma}(f) := \int_{\mathbf{G}/\mathbf{G}_{\gamma}(\mathbb{A}_F)} f(x\gamma x^{-1}) dx = \prod_v \int_{\mathbf{G}/\mathbf{G}_{\gamma}(F_v)} f_v(x_v \gamma x_v^{-1}) dx_v.$$

⁽²⁾On peut citer deux autres stratégies : l’une basée sur les théorèmes inverses «à la Weil» de Piatetski-Shapiro, l’autre basée sur la correspondance Theta de Howe.

La fonction $x \mapsto f(x\gamma x^{-1})$ est définie sur $\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)$ via la bijection $(\mathbf{G}(\mathbb{A}_F)/\mathbf{G}_\gamma(\mathbb{A}_F))^F \simeq \mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)$ sur laquelle nous laissons le lecteur méditer. Les mesures $\mathbf{G}(F_v)$ -invariantes sur chaque $\mathbf{G}/\mathbf{G}_\gamma(F_v)$ sont déterminées par celles qu'on a choisies sur $\mathbf{G}(F_v)$ et $\mathbf{G}_\gamma(F_v)$ puisqu'on a une partition $\mathbf{G}/\mathbf{G}_\gamma(F_v) = \sqcup_{\gamma' \sim_{st} \gamma} \mathbf{G}(F_v)/\mathbf{G}_{\gamma'}(F_v)$ de l'orbite stable en orbites ordinaires et des isomorphismes $\mathbf{G}_{\gamma'} \simeq \mathbf{G}_\gamma$ pour $\gamma' \in \mathbf{G}(F_v)$ stablement conjugué à γ . Les facteurs locaux sont presque tous égaux à 1 en vertu d'un résultat de Kottwitz [18, 7.1]. Enfin, l'expression $SO_\gamma(f)$ ne dépend *que de la classe stable de γ* .

On voudrait maintenant comparer les expressions :

$$\sum_{\gamma' \in \mathbf{G}(F)/\text{conj}, \gamma' \sim_{st} \gamma} \tau(\mathbf{G}_{\gamma'}) O_{\gamma'}(f) \quad \text{et} \quad SO_\gamma(f)$$

et pour cela on doit exprimer le terme de gauche comme une certaine intégrale sur $\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)$.

Il est instructif de regarder d'abord le problème local associé, qui consiste à écrire l'intégrale orbitale $O_\gamma(f_v)$ comme une intégrale sur $\mathbf{G}/\mathbf{G}_\gamma(F)$. Pour cela, il faut caractériser l'image de l'injection $\mathbf{G}(F)/\mathbf{G}_\gamma(F) \hookrightarrow \mathbf{G}/\mathbf{G}_\gamma(F)$ de la classe de conjugaison de γ dans sa classe stable. En fait, on dispose pour tout F -corps K d'une application cobord

$$\mathbf{G}/\mathbf{G}_\gamma(K) \xrightarrow{\text{inv}_K(\gamma, -)} \ker (H^1(K, \mathbf{G}_\gamma) \longrightarrow H^1(K, \mathbf{G})) =: \partial^1(K, \mathbf{G}_\gamma, \mathbf{G})$$

qui induit une bijection entre l'ensemble $\mathbf{G}(K) \backslash (\mathbf{G}/\mathbf{G}_\gamma)(K)$ des classes de conjugaison dans la classe stable et $\partial^1(K, \mathbf{G}_\gamma, \mathbf{G})$, et qui envoie la classe de conjugaison de γ sur l'élément « neutre » de $\partial^1(K, \mathbf{G}_\gamma, \mathbf{G})$. En d'autres termes,

$$O_\gamma(f_v) = \int_{\mathbf{G}/\mathbf{G}_\gamma(F_v)} \delta_{e, \text{inv}_{F_v}(\gamma, x)} f_v(x\gamma x^{-1}).$$

Dans notre situation et lorsque v est non-archimédienne, $\partial^1(F_v, \mathbf{G}_\gamma, \mathbf{G})$ est un groupe abélien fini. On peut alors faire une transformation de Fourier en introduisant son dual $\mathfrak{K}(\gamma)_v$ pour obtenir

$$\begin{aligned} O_\gamma(f_v) &= |\mathfrak{K}(\gamma)_v|^{-1} \sum_{\kappa \in \mathfrak{K}(\gamma)_v} \int_{\mathbf{G}/\mathbf{G}_\gamma(F_v)} \langle \kappa, \text{inv}_{F_v}(\gamma, x) \rangle f_v(x\gamma x^{-1}) dx \\ &=: |\mathfrak{K}(\gamma)_v|^{-1} \sum_{\kappa \in \mathfrak{K}(\gamma)_v} O_\gamma^\kappa(f_v). \end{aligned}$$

Les termes $O_\gamma^\kappa(f_v)$ sont appelés κ -intégrales orbitales.

Revenons au problème global, et introduisons l'application « canonique »

$$\phi_\gamma : \mathbf{G}(F) \backslash (\mathbf{G}/\mathbf{G}_\gamma)(F) \longrightarrow \mathbf{G}(\mathbb{A}_F) \backslash (\mathbf{G}/\mathbf{G}_\gamma)(\mathbb{A}_F).$$

Un résultat de Borel-Serre montre la finitude des fibres de cette application, et puisque $\tau(\mathbf{G}'_\gamma) = \tau(\mathbf{G}_\gamma)$ pour γ' stablement conjugué à γ , on peut alors réécrire

$$(1) \quad \sum_{\gamma' \in \mathbf{G}(F)/\text{conj}, \gamma' \sim_{st} \gamma} \tau(\mathbf{G}_{\gamma'}) O_{\gamma'}(f) = \tau(\mathbf{G}_\gamma) \int_{\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)} |\phi_\gamma^{-1}(x)| f(x\gamma x^{-1}) dx.$$

Traduisons alors ϕ_γ en termes de cohomologie galoisienne : d’après ce qui a été dit dans la discussion du cas local ci-dessus, elle s’identifie à la flèche naturelle

$$(2) \quad \partial^1(F, \mathbf{G}_\gamma, \mathbf{G}) \longrightarrow \bigoplus_v \partial^1(F_v, \mathbf{G}_\gamma, \mathbf{G}).$$

Supposons, pour simplifier encore un peu, que \mathbf{G} est semi-simple (et toujours simplement connexe). Dans le cas non-archimédien, le théorème de Kneser-Bruhat-Tits selon lequel $H^1(F_v, \mathbf{G}) = 0$ et la dualité de Tate-Nakayama appliquée au tore \mathbf{G}_γ nous donnent alors une injection $\partial^1(F_v, \mathbf{G}_\gamma, \mathbf{G}) \hookrightarrow X_*(\mathbf{G}_\gamma)_{\Gamma_v}$. Un morphisme analogue existe pour v archimédienne, et par composition et somme on en déduit une flèche

$$(3) \quad \bigoplus_v \partial^1(F_v, \mathbf{G}_\gamma, \mathbf{G}) \longrightarrow X_*(\mathbf{G}_\gamma)_\Gamma,$$

ce dernier groupe étant fini, puisque \mathbf{G}_γ est F -anisotrope. Nous noterons $\mathfrak{K}(\gamma)$ le groupe dual de $X_*(\mathbf{G}_\gamma)_\Gamma$.

THÉORÈME 1.1 (Langlands, Kottwitz). — *L’image de l’application (2) coïncide avec le noyau de la flèche (3). Ses fibres ont pour cardinal $\tau(G)\tau(G_\gamma)^{-1}|\mathfrak{K}(\gamma)|$, les τ désignant les nombres de Tamagawa.*

Ce théorème est valable sans hypothèse de semi-simplicité, mais la définition générale de $\mathfrak{K}(\gamma)$ est plus compliquée. Notons maintenant $\text{inv}(\gamma, -)$ la composée

$$\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F) = \prod_v \mathbf{G}/\mathbf{G}_\gamma(F_v) \xrightarrow{\oplus \text{inv}_v} \bigoplus_v \partial^1(F_v, \mathbf{G}_\gamma, \mathbf{G}) \longrightarrow \mathfrak{K}(\gamma)^*.$$

Par le théorème précédent, on peut reformuler (1) en

$$\sum_{\gamma' \in \mathbf{G}(F)/\text{conj}, \gamma' \sim_{st} \gamma} \tau(\mathbf{G}_{\gamma'}) O_{\gamma'}(f) = \tau(\mathbf{G}) |\mathfrak{K}(\gamma)| \int_{\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)} \delta_{\text{inv}(\gamma, x), 0} f(x\gamma x^{-1}) dx,$$

ce qui par transformation de Fourier sous le groupe $\mathfrak{K}(\gamma)$, nous donne :

$$(4) \quad \sum_{\gamma' \in \mathbf{G}(F)/\text{conj}, \gamma' \sim_{st} \gamma} \tau(\mathbf{G}_{\gamma'}) O_{\gamma'}(f) = \tau(\mathbf{G}) \sum_{\kappa \in \mathfrak{K}(\gamma)} O_\gamma^\kappa(f)$$

où

$$O_\gamma^\kappa(f) = \int_{\mathbf{G}/\mathbf{G}_\gamma(\mathbb{A}_F)} \langle \kappa, \text{inv}(\gamma, x) \rangle f(x\gamma x^{-1}) dx = \prod_v O_{\gamma_v}^\kappa(f_v).$$

L’expression ci-dessus est la κ -intégrale orbitale globale. Il est bon de souligner que celle-ci ne dépend que de la classe stable de γ dans $G(F)$, tandis que les κ -intégrales orbitales locales *dépendent d’un point-base dans la classe stable* : si γ_v est stablement conjugué à γ dans $G(F_v)$, alors $O_{\gamma_v}^\kappa(f_v) = \langle \kappa, \text{inv}_{F_v}(\gamma, \gamma_v) \rangle O_\gamma^\kappa(f_v)$.

1.5. Endoscopie et transfert global

On a obtenu l'expression

$$T_G^{regell}(f) = \tau(\mathbf{G}) \sum_{\gamma \in \mathbf{G}(F)_{/st}^{regell}} \sum_{\kappa \in \mathcal{R}(\gamma)} \mathcal{O}_\gamma^\kappa(f).$$

Le but de l'endoscopie est d'exprimer les κ -intégrales orbitales comme des intégrales stables sur des groupes quasi déployés de dimension inférieure ou égale à celle de \mathbf{G} . Pour cela, introduisons le système de racines absolu $\Phi = (\mathbb{X}^*, \Sigma, \mathbb{X}_*, \Sigma^\vee)$ et le groupe de Weyl W_G de \mathbf{G} , et rappelons qu'on a une bijection canonique $\mathbf{G}(\overline{F})_{/conj}^{ss} \xrightarrow{\sim} (\mathbb{X}_* \otimes \overline{F}^\times)/W_G$. Ainsi, si s est un caractère de \mathbb{X}_* , «le» groupe réductif connexe \mathbf{G}_s sur \overline{F} de système de racines $\Phi_s := (\mathbb{X}^*, \Sigma_s, \mathbb{X}_*, \Sigma_s^\vee)$ où $\Sigma_s^\vee := \ker(s) \cap \Sigma^\vee$ est muni d'une application

$$(5) \quad \mathbf{G}_s(\overline{F})_{/conj}^{ss} \xrightarrow{\phi} \mathbf{G}(\overline{F})_{/conj}^{ss}$$

de «transfert» des classes de conjugaison semi-simples. Si une classe γ dans $\mathbf{G}(\overline{F})$ est fortement régulière, la conjugaison dans \mathbf{G} induit des isomorphismes canoniques entre les centralisateurs dans \mathbf{G} des divers éléments de cette classe, et on peut donc noter sans ambiguïté \mathbf{G}_γ le centralisateur «commun». Appelons « G -régulière» toute classe semi-simple δ de $\mathbf{G}_s(\overline{F})$ dont l'image $\phi(\delta)$ est fortement régulière dans \mathbf{G} . Pour une telle classe, on a un isomorphisme canonique

$$(6) \quad (\mathbf{G}_s)_\delta \xrightarrow{\sim} \mathbf{G}_{\phi(\delta)}.$$

Enfin, la conjugaison des tores maximaux dans \mathbf{G}_s permet de transporter *sans ambiguïté* le caractère s de \mathbb{X}_* à $X_*((\mathbf{G}_s)_\delta)$, ce qui par l'isomorphisme précédent fournit encore un caractère

$$(7) \quad s : X_*((\mathbf{G}_s)_\delta) \longrightarrow \mathbb{C}^\times.$$

La théorie de l'endoscopie s'intéresse aux structures F -rationnelles sur \mathbf{G}_s telles que les applications (5), (6) et (7) soient Γ -invariantes. Voici une manière concrète de les construire : la structure F -rationnelle de \mathbf{G} induit toute une famille d'actions de Γ sur Φ . Pour se fixer les idées, on peut choisir l'unique telle action qui stabilise une base Δ préalablement choisie de Σ ; toutes les autres s'en déduisent par composition de $W_G \rtimes \Gamma \longrightarrow \text{Aut}(\Phi)$ avec une section $\Gamma \longrightarrow W_G \rtimes \Gamma$.

DÉFINITION 1.2. — Une donnée endoscopique⁽³⁾ est un couple (s, ρ) où s est un caractère de \mathbb{X}_* et $\rho : \Gamma \longrightarrow W_G \rtimes \Gamma$ est une section d'image contenue dans le fixateur $(W_G \rtimes \Gamma)_s$ de s . Elle est dite elliptique si $\mathbb{X}_*^{W_{G_s} \rtimes \rho \Gamma} = \{0\}$.

⁽³⁾Cette définition ne coïncide avec la référence [17, par. 7] que sous l'hypothèse simplificatrice de semi-simplicité et simple connexité de . Il est probable que l'on puisse plus généralement se contenter de cette version simplifiée dès que le groupe satisfait le principe de Hasse, et à condition d'adapter la notion d'isomorphisme de données endoscopiques. Par ailleurs, nous avons préféré éviter d'utiliser le L -groupe, comme dans [22] et [14].

Il y a une notion d'isomorphisme de données endoscopiques que nous passons sous silence. La donnée de ρ induit une action $\Gamma \rightarrow \text{Aut}(\Phi_s)$ qui détermine sur \mathbf{G}_s une F -structure quasi déployée que nous noterons $\mathbf{G}_{s,\rho}$, et pour laquelle les applications (5), (6) et (7) sont Γ -invariantes. On peut remarquer que lorsque s est le caractère trivial et ρ la section triviale, le groupe $\mathbf{G}_{s,\rho}$ est la forme intérieure quasi déployée de \mathbf{G} .

Partons maintenant d'une classe *stable* G -régulière δ de $\mathbf{G}_{s,\rho}(F)$; par Γ -équivariance de (5), la classe $\phi(\delta) \in \mathbf{G}(\overline{F})_{\text{conj}}^{ss}$ contient au plus une classe stable γ de $\mathbf{G}(F)$, et exactement une lorsque \mathbf{G} est quasi-déployé avec \mathbf{G}_{der} simplement connexe, en vertu d'un résultat de Steinberg-Kottwitz. On a donc une application *partielle*

$$\mathbf{G}_{s,\rho}(F)_{/\text{st}}^{G\text{-reg}} \dashrightarrow \mathbf{G}(F)_{/\text{st}}^{\text{reg}}$$

dont les fibres sont finies. Par Γ -invariance de (7), le caractère s induit un caractère $X_*(\mathbf{G}_\gamma)_\Gamma \rightarrow \mathbb{C}^*$ que nous noterons κ . On notera symboliquement $(s, \rho, \delta) \mapsto (\gamma, \kappa)$ l'application partielle à fibres finies ainsi obtenue.

CONJECTURE 1.3 (Transfert). — *Soient (s, ρ) une donnée endoscopique et $f \in \mathcal{C}_c^\infty(\mathbf{G}(\mathbb{A}_F))$. Il existe une fonction $f_{s,\rho} \in \mathcal{C}_c^\infty(\mathbf{G}_{s,\rho}(\mathbb{A}_F))$ telle que pour toute classe stable elliptique G -régulière δ de $\mathbf{G}_{s,\rho}(F)$, on ait*

$$SO_\delta(f_{s,\rho}) = O_\gamma^\kappa(f),$$

si $(s, \rho, \delta) \mapsto (\gamma, \kappa)$.

En admettant cette conjecture et en étudiant les fibres de l'application $(s, \rho, \delta) \mapsto (\gamma, \kappa)$, Langlands a montré la formule (pour F un corps de nombres)

$$(8) \quad T_G^{\text{regell}}(f) = \sum_{(s,\rho)/\sim} \iota(G, s, \rho) ST_{G_{s,\rho}}^{G\text{-regell}}(f_{s,\rho})$$

où la somme porte sur les classes d'équivalence de données endoscopiques *elliptiques* et

$$ST_{G_{s,\rho}}^{G\text{-regell}}(g) := \tau(\mathbf{G}_{s,\rho}) \sum_{\gamma \in \mathbf{G}_{s,\rho}(F)_{/\text{conj. st.}}^{G\text{-regell}}} SO_\gamma(g), \quad \text{pour } g \in \mathcal{C}_c^\infty(\mathbf{G}_{s,\rho}(\mathbb{A}_F)).$$

La formule (8) est la «stabilisation de la partie elliptique régulière de la formule des traces de \mathbf{G} sur F » [25]. Elle sert de modèle pour la stabilisation de toute la partie elliptique [18], de toute la formule des traces [3][2][4], et des analogues tordus [21] et [22]. *Tous ces travaux* supposent la conjecture de transfert (éventuellement une variante tordue) vérifiée.

En guise de (brèves) explications sur la preuve de la formule (8), nous allons vérifier que l'application partielle $(s, \rho, \delta) \mapsto (\gamma, \kappa)$ est *surjective*. Fixons donc une classe stable γ de $\mathbf{G}(F)$ et un caractère κ de $X_*(\mathbf{G}_\gamma)_\Gamma$. Puisque \mathbf{G}_γ est un tore maximal de \mathbf{G} , on a une W_G -classe de conjugaison canonique d'isomorphismes $X_*(\mathbf{G}_\gamma) \xrightarrow{\sim} \mathbb{X}_*$ de groupes abéliens. *Choisissons* un tel isomorphisme : il permet de transporter κ en un caractère s_κ de \mathbb{X}_* et l'action de Γ sur $X_*(\mathbf{G}_\gamma)$ en une action sur \mathbb{X}_* ; cette dernière est donnée par une section $\rho_\gamma : \Gamma \rightarrow W_G \rtimes \Gamma$ d'image contenue dans le

fixateur $(W_G \rtimes \Gamma)_{s_\kappa}$, puisque κ était lui-même Γ -invariant. On obtient donc une donnée endoscopique (s_κ, ρ_γ) , qui est elliptique si et seulement si γ l'est. De plus, la $\mathbf{G}_{s_\kappa, \rho_\gamma}(\overline{F})$ -classe de conjugaison de \overline{F} -plongements $\mathbf{G}_\gamma \hookrightarrow \mathbf{G}_{s_\kappa, \rho_\gamma}$ déterminée par l'isomorphisme $X_*(\mathbf{G}_\gamma) \xrightarrow{\sim} \mathbb{X}_*$ est, par construction, Γ -stable. Puisque $\mathbf{G}_{s_\kappa, \rho_\gamma}$ est quasi-déployé, le résultat de Steinberg-Kottwitz mentionné plus haut assure alors l'existence dans cette classe d'un F -plongement de \mathbf{G}_γ dans $\mathbf{G}_{s_\kappa, \rho_\gamma}$, qui est donc bien défini à conjugaison *stable* près. Posons alors $\gamma_{s_\kappa, \rho_\gamma}$ la classe stable image de γ par ce plongement. On a bien $(s_\kappa, \rho_\gamma, \gamma_{s_\kappa, \rho_\gamma}) \mapsto (\gamma, \kappa)$. Notons que cette construction dépend du choix initial de l'isomorphisme $X_*(\mathbf{G}_\gamma) \xrightarrow{\sim} \mathbb{X}_*$. Néanmoins, on peut vérifier que les triplets (s, ρ, δ) au-dessus de (γ, κ) sont tous «isomorphes» à des triplets de la forme $(s_\kappa, \rho_\gamma, \gamma')$.

1.6. Transfert local et lemme fondamental

Puisque les intégrales orbitales stables et les κ -intégrales orbitales sont produits de facteurs locaux analogues, une manière d'établir (et d'expliciter un peu) le transfert de la conjecture 1.3 est de le déduire de transferts locaux à toutes les places. La forme naïve que prendrait un tel transfert serait (avec les notations de 1.3) : «étant donnée $f_v \in \mathcal{C}_c^\infty(\mathbf{G}(F_v))$, il existe $f'_v \in \mathcal{C}_c^\infty(\mathbf{G}_{s, \rho}(F_v))$ telle que pour toute classe G_v -régulière δ dans $\mathbf{G}_{s, \rho}(F_v)$, on ait $SO_\delta(f'_v) = O_\gamma^\kappa(f_v)$ si $(s, \rho, \delta) \mapsto (\kappa, \delta)$ ». Mais ceci n'est pas possible pour au moins deux raisons : tout d'abord les κ -intégrales orbitales dépendent d'un point-base dans la classe stable γ , donc un seul choix de point-base est permis et il faut trouver un moyen de le déterminer. D'autre part, les intégrales orbitales $O_\gamma(f)$, en tant que fonction de γ élément semi-simple régulier, «explosent» au voisinage des éléments singuliers. Ainsi quand δ approche un élément G_v -singulier mais régulier, $SO_\delta(f'_v)$ reste bornée tandis que $O_\gamma(f_v)$ «explose», ce qui empêche évidemment l'égalité des deux expressions.

Une possibilité pour résoudre ces problèmes est d'introduire des facteurs $\Delta_v(\delta, \gamma)$ qui dépendent de la classe stable de δ et de la classe *ordinaire* de γ . En prenant en compte toutes les contraintes de la situation, Langlands et Shelstad [26] ont trouvé une définition de ces facteurs, et les ont appelés *facteurs de transfert*. La conjecture locale prend alors la forme suivante, où l'on se place dans un contexte *local* (voir introduction), avec une définition des données endoscopiques et des applications $(s, \rho, \delta) \mapsto (\gamma, \kappa)$ analogues au cas global :

CONJECTURE 1.4 (Transfert local). — *Soient (s, ρ) une donnée endoscopique de \mathbf{G} sur F et $f \in \mathcal{C}_c^\infty(\mathbf{G}(F))$. Il existe une fonction $f_{s, \rho} \in \mathcal{C}_c^\infty(\mathbf{G}_{s, \rho}(F))$ telle que pour toute classe stable fortement G -régulière δ de $\mathbf{G}_{s, \rho}(F)$ et toute classe ordinaire γ de $\mathbf{G}(F)$, on ait*

$$SO_\delta(f_{s, \rho}) = \Delta(\delta, \gamma)O_\gamma^\kappa(f),$$

si $(s, \rho, \delta) \mapsto (\gamma, \kappa)$. ⁽⁴⁾

⁽⁴⁾Pour renouer avec les notations de l'introduction, il faut poser $O_\delta^{G|H} := \Delta(\delta, \gamma)O_\gamma^\kappa$ si $H = G_{s, \rho}$.

Pour que cette conjecture locale serve à la conjecture globale, il faut deux choses : d'une part, dans un contexte global, le produit des facteurs de transfert locaux doit être égal à 1 : ceci fait partie des contraintes que s'étaient fixées Langlands et Shelstad. D'autre part, aux places non-ramifiées, les fonctions caractéristiques des compacts hyperspéciaux intervenant dans la définition des groupes adéliques doivent se correspondre. On est donc amené au supplément suivant :

CONJECTURE 1.5 (Lemme fondamental). — *Soient \mathbf{G} un groupe non ramifié sur F et (s, ρ) une donnée endoscopique non-ramifiée. Alors pour toute classe stable fortement G -régulière δ de $\mathbf{G}_{s, \rho}(F)$ et toute classe ordinaire γ de $\mathbf{G}(F)$, on a*

$$SO_{\delta}(\mathbf{1}_{\mathbf{G}_{s, \rho}(\mathcal{O})}) = \Delta(\delta, \gamma) O_{\gamma}^{\kappa}(\mathbf{1}_{\mathbf{G}(\mathcal{O})}),$$

si $(s, \rho, \delta) \mapsto (\gamma, \kappa)$.

En fait les facteurs de transfert ne sont définis qu'à un multiple près, ce qui n'est pas gênant pour la conjecture de transfert mais l'est pour le lemme fondamental, une fois les choix de mesures naturels effectués. Dans le cas \mathbf{G} quasi-déployé, la définition de Langlands-Shelstad associe à tout choix d'épinglage de \mathbf{G} sur F une normalisation des facteurs de transfert. D'autre part, l'énoncé du lemme fondamental dépend de prolongements entiers réductifs connexes de \mathbf{G} et $\mathbf{G}_{s, \rho}$; il faut choisir ces prolongements entiers de façon compatible à l'épinglage dont dépendent les facteurs de transfert, cf. [12, 7.1].

Le facteur de transfert $\Delta(\delta, \gamma)$ est le produit d'une puissance de p , un quotient de discriminants faciles à définir, et d'une racine de l'unité (qui est même un signe pour les groupes classiques) qui, elle, est difficile à définir. Nous ne donnerons pas sa définition générale, mais signalons qu'il a été calculé par Waldspurger [36] pour tous les groupes classiques. De plus, dans le cas non-ramifié, une définition générale plus simple a été proposée par Hales [12] et dans le cas des algèbres de Lie, une caractérisation agréable a été donnée par Kottwitz dans [20].

Pour un peu plus de détails et d'exemples sur le lemme fondamental, nous recommandons [14], qui toutefois recule aussi devant les facteurs de transfert. On y trouvera une (courte) liste des cas particuliers du lemme fondamental (non tordu) connus jusqu'à présent, la seule famille de groupes, $SL(n)$, $n \in \mathbb{N}$, étant due à Waldspurger [38].

1.7. Les réductions de Waldspurger (après Langlands-Shelstad et Hales)

La variante «algèbres de Lie» du lemme fondamental introduite par Waldspurger est formellement analogue à l'énoncé 1.5 sauf que les fonctions-test sont $\mathbf{1}_{\mathbf{g}_{s, \rho}(\mathcal{O})}$ et $\mathbf{1}_{\mathbf{g}(\mathcal{O})}$ et les classes de conjugaison γ et δ sont respectivement dans les algèbres de Lie rationnelles $\mathbf{g}(F)$ et $\mathbf{g}_{s, \rho}(F)$. Les centralisateurs de tels éléments sont toujours connexes, et il n'y a donc pas lieu de parler de «fortement» régulier.

Lorsque le corps F est de caractéristique nulle (une extension finie de \mathbb{Q}_p), Waldspurger a montré [35], à la suite des travaux de Langlands-Shelstad [27] et de Hales [13], que pour résoudre la conjecture de transfert pour une paire endoscopique $(\mathbf{G}, (s, \rho))$

sur F , il suffit de connaître la version «algèbres de Lie» du lemme fondamental pour certaines familles de paires endoscopiques non-ramifiées $(\mathbf{G}', (s', \rho'))$ de même système de racines absolu qu'un centralisateur d'élément semi-simple dans \mathbf{G} , et définies sur d'autres corps locaux F' de caractéristique nulle, dont on peut de plus supposer les caractéristiques résiduelles supérieures à un entier donné.

Par ailleurs, il a montré récemment [37] qu'en supposant la caractéristique résiduelle assez grande (avec minorant explicite en fonction du système de racines de \mathbf{G}), l'énoncé du lemme fondamental pour les algèbres de Lie ne dépend *que du corps résiduel*. Plus précisément : notons k le corps résiduel de F et fixons un prolongement de \mathbf{G} en un groupe réductif connexe sur \mathcal{O}_F . Puisque \mathcal{O}_F et $k[[T]]$ ont des groupes de Galois isomorphes (à $\widehat{\mathbb{Z}}$), \mathbf{G} détermine un «unique» groupe réductif \mathbf{G}' sur $k[[T]]$ et chaque donnée endoscopique non-ramifiée de \mathbf{G} se transfère à \mathbf{G}' . Waldspurger définit alors une correspondance $\delta \leftrightarrow \delta'$ entre classes de conjugaison stable dans $\mathfrak{g}_{s,\rho}(F)^{G-reg}$ et $\mathfrak{g}_{s',\rho'}(k((T)))^{G-reg}$, de sorte qu'on a les égalités $SO_\delta(\mathbf{1}_{\mathbf{H}(\mathcal{O}_F)}) = SO_{\delta'}(\mathbf{1}_{\mathbf{H}'(k[[T]])})$ et $\Delta(\delta, \gamma)O_\delta^s(\mathbf{1}_{\mathbf{G}(\mathcal{O}_F)}) = \Delta(\delta', \gamma')O_{\delta'}^{s'}(\mathbf{1}_{\mathbf{G}'(k[[T]])})$ (pour n'importe quels γ, γ' dans la classe stable associée à δ , resp. δ').

Soulignons une conséquence de tout ceci, reposant sur le fait que le centralisateur d'un élément semi-simple dans $GL(N)$ est un produit de $GL(m)$. Pour résoudre les conjectures de transfert globales ou locales de Langlands-Shelstad relatives aux groupes unitaires sur un corps de nombres ou p -adique, il suffit de savoir résoudre le lemme fondamental dans sa version «algèbres de Lie» pour tout groupe unitaire non-ramifié sur $\mathbb{F}_{(p')^d}((T))$ avec p' «grand».

1.8. Le lemme fondamental pour les groupes unitaires

Pour conclure cette partie, nous explicitons l'énoncé du lemme fondamental pour l'algèbre de Lie d'un groupe unitaire non-ramifié d'égales caractéristiques, sous la forme qui est démontrée par Laumon et Ngô. Fixons une puissance q d'un nombre premier $p > 2$ et posons $F := \mathbb{F}_q((T))$.

1.8.1. Groupes unitaires. — Pour $n \in \mathbb{N}$, notons Φ_n la matrice ayant des 1 sur l'anti-diagonale et des 0 ailleurs. À toute extension quadratique F'/F , on associe le schéma en groupes réductifs unitaire quasi-déployé en n variables sur F défini par

$$\mathbf{U}_{F'|F}(n)(R) = \{g \in GL_n(R \otimes_F F'), \Phi_n \cdot \tau({}^t g^{-1}) \cdot \Phi_n^{-1} = g\}$$

pour toute F -algèbre R , et où τ désigne l'élément non trivial de $Gal(F'/F)$. On note $U_{F'|F}(n) := \mathbf{U}_{F'|F}(n)(F)$ les points rationnels et $\mathbf{u}_{F'|F}(n)$ l'algèbre de Lie. Ces groupes ont un centre compact et un groupe dérivé simplement connexe. Lorsque F' est «1»'extension quadratique non-ramifiée de F , on note simplement $\mathbf{U}(n)$ à la place de $\mathbf{U}_{F'|F}(n)$. Dans ce cas, on le prolonge en un schéma en groupes réductifs connexe sur $\mathcal{O} = \mathcal{O}_F$, en remplaçant simplement F par \mathcal{O} et F' par $\mathcal{O}_{F'}$ dans la définition ci-dessus.

1.8.2. *Tores elliptiques.* — On a déjà signalé que les tores maximaux elliptiques de $\mathbf{U}(n)$ sont de la forme $\mathbf{T}_{I,(F_i)} := \prod_{i \in I} \text{Res}_{F_i|F} \mathbf{U}_{F'_i|F_i}(1)$ où les F_i sont des extensions séparables de F disjointes de F' (*i.e.* de degré résiduel impair) telles que $\sum_{i \in I} \deg(F_i|F) = n$ et $F'_i = F_i F'$ est l'extension quadratique non ramifiée de F_i . Nous allons décrire les classes de conjugaison de F -plongements $\mathbf{T}_{I,(F_i)} \hookrightarrow \mathbf{U}(n)$ dans une classe stable fixée. On a pour tout F, F', n des isomorphismes

$$H^1(F, \mathbf{U}_{F'|F}(n)) \xrightarrow{\sim} F^\times / \text{Nr}_{F'|F}(F'^\times) \simeq \mathbb{Z}/2\mathbb{Z}$$

et ce groupe de cohomologie ne fait rien d'autre que classifier les classes d'équivalence de formes hermitiennes sur F'^n par leur discriminant. Ainsi

$$\partial^1(F, \mathbf{T}_{I,(F_i)}, \mathbf{U}(n)) \simeq \ker \left((\mathbb{Z}/2\mathbb{Z})^I \longrightarrow \mathbb{Z}/2\mathbb{Z} \right)$$

où le morphisme de droite est le morphisme somme. Concrètement, soient $(\lambda_i)_{i \in I} \in \partial^1(F, \mathbf{T}_{I,(F_i)}, \mathbf{U}(n))$; choisissons des éléments $c_i \in F_i$ tels que la forme hermitienne $\Psi_{i,c_i} : (x, y) \mapsto \text{Tr}_{F'_i|F'}(c_i x^\tau y)$ sur le F' -espace vectoriel F'_i soit de discriminant λ_i . Alors les F -espaces hermitiens (F'^n, Φ_n) et $(\oplus_i F'_i, \oplus_i \Psi_{i,c_i})$ sont isomorphes, et tout isomorphisme induit un F -plongement $\mathbf{T}_{I,(F_i)} \hookrightarrow \mathbf{U}(n)$. Deux tels isomorphismes sont conjugués et, en faisant varier $(\lambda_i)_{i \in I}$, on obtient ainsi les classes de conjugaison dans la classe stable associée à un isomorphisme F' -linéaire $\oplus_i F'_i \xrightarrow{\sim} (F')^n$.

1.8.3. *Éléments elliptiques réguliers.* — Un élément de l'algèbre de Lie rationnelle de $\mathbf{T}_{I,(F_i)}$ est un I -uplet $\gamma_I = (\gamma_i)_{i \in I}$ avec $\gamma_i \in F'_i$ vérifiant $\gamma_i + \gamma_i^\tau = 0$. Un tel élément est régulier (pour tout F -plongement $\mathbf{T}_{I,(F_i)} \hookrightarrow \mathbf{U}(n)$) si γ_i engendre F'_i sur F' et les polynômes minimaux $P_i(X)$ des γ_i sont premiers entre eux, deux à deux. Pour les besoins du lemme fondamental, on peut supposer que $\gamma_i \in \mathcal{O}_{F'_i}$ pour tout i .

1.8.4. *Intégrales orbitales.* — Fixons $\lambda_I := (\lambda_i)_{i \in I} \in \partial^1(F, \mathbf{T}_{I,(F_i)}, \mathbf{U}(n))$ et un F -plongement associé $\iota : \mathbf{T}_{I,(F_i)} \hookrightarrow \mathbf{U}(n)$. On notera $O_{\gamma_I}^{\lambda_I}$ l'intégrale orbitale $O_{\iota(\gamma_I)}(\mathbf{1}_{\mathbf{u}(n)(\mathcal{O})})$, qui ne dépend bien sûr pas du choix de ι . Alors on a

$$\begin{aligned} O_{\gamma_I}^{\lambda_I} &= |\{x \in \mathbf{U}(n)(\mathcal{O}) \setminus \mathbf{U}(n)(F), \text{ t.q. } \text{Ad}(x)(\iota(\gamma_I)) \in \mathbf{u}(n)(\mathcal{O})\}| \\ &= |\{\text{réseaux autoduaux } M \subset (F')^n, \text{ t.q. } \iota(\gamma_I)M \subseteq M\}| \\ &= |\{\text{réseaux } M_I \subset F'_I, \text{ t.q. } (M_I)^{\perp_{c_I}} = M_I \text{ et } \gamma_I M_I \subset M_I\}|. \end{aligned}$$

Précisons la dernière ligne : on compte les $\mathcal{O}_{F'}$ -réseaux M_I dans le F' -espace vectoriel $F'_I := \oplus_i F'_i$, autoduaux pour une forme hermitienne $\oplus_i \Psi_{i,c_i}$ comme dans 1.8.2, et stables par γ_I . Cette formulation se prête bien à l'approche géométrique initiée dans [11] via les fibres de Springer affines.

L'approche de Laumon [28] et Laumon-Ngô commence par la réinterprétation suivante du comptage ci-dessus. Un réseau M_I de F'_I stable par γ_I est simplement un idéal fractionnaire de l'anneau $A_I := \mathcal{O}_{F'}[\gamma_I] \subset \mathcal{O}_{F'_I}$ (qui a pour anneau total de fractions F'_I). En termes géométriques, c'est un point rationnel de la jacobienne compactifiée du germe de courbe $\text{Spf}(A_I)$. Soit $c_I^0 \in F_I$ un générateur du module dualisant relatif

de A_I sur $\mathcal{O}_{F'}$. Un calcul montre que l'orthogonal de A_I pour la forme hermitienne $\Psi_{I, c_I} := \oplus_i \Psi_{i, c_i}$ est donné par la formule $(A_I)^{\perp_{c_I}} = c_I^0 c_I^{-1} A_I$. On a donc la formule

$$O_{\gamma_I}^{\lambda_I} = |\{\text{idéaux fractionnaires } M_I \text{ de l'anneau } A_I \text{ tels que } M_I = c_I^0 c_I^{-1} M_I^{-1}\}|.$$

1.8.5. Groupes endoscopiques et lemme fondamental. — Les groupes endoscopiques non-ramifiés elliptiques de $\mathbf{U}(n)$ sont tous de la forme $\mathbf{U}(n_1) \times \mathbf{U}(n_2)$ pour une partition $n = n_1 + n_2$. Fixons donc un tore F -anisotrope $\mathbf{T}_{I, (F_i)}$ qui se décompose en un produit $\mathbf{T}_{I_1, (F_i)} \times \mathbf{T}_{I_2, (F_i)}$ pour une partition $I = I_1 \sqcup I_2$ telle que $\sum_{i \in I_1} \deg(F_i|F) = n_1$, puis un élément $\gamma_I = (\gamma_{I_1}, \gamma_{I_2})$ de son algèbre de Lie. Par les paragraphes précédents, cet élément détermine une classe stable de $\mathbf{U}(n_1) \times \mathbf{U}(n_2)$ et une classe stable de $\mathbf{U}(n)$.

D'après le calcul ci-dessus, l'intégrale orbitale stable de γ_I dans $\mathbf{U}(n_1) \times \mathbf{U}(n_2)$ est

$$SO_{\gamma_I} = \sum_{\lambda_{I_1}, \lambda_{I_2}} O_{\gamma_{I_1}}^{\lambda_{I_1}} O_{\gamma_{I_2}}^{\lambda_{I_2}}$$

où $(\lambda_{I_1}, \lambda_{I_2}) \in \partial^1(F, \mathbf{T}_{I_1, (F_i)}, \mathbf{U}(n_1)) \times \partial^1(F, \mathbf{T}_{I_2, (F_i)}, \mathbf{U}(n_2))$.

Par ailleurs le caractère endoscopique κ de $\partial^1(F, \mathbf{T}_{I, (F_i)}, \mathbf{U}(n))$ déterminé par le groupe endoscopique $\mathbf{U}(n_1) \times \mathbf{U}(n_2)$ est le caractère $\partial^1(F, \mathbf{T}_{I, (F_i)}, \mathbf{U}(n)) \rightarrow \mathbb{Z}/2\mathbb{Z}$ de noyau $\partial^1(F, \mathbf{T}_{I_1, (F_i)}, \mathbf{U}(n_1)) \times \partial^1(F, \mathbf{T}_{I_2, (F_i)}, \mathbf{U}(n_2))$. Pour former la κ -intégrale orbitale, il faut choisir un point-base dans la classe stable. Laumon et Ngô choisissent un point dans le noyau de κ . On pose donc

$$O_{\gamma_I}^{\kappa} = \sum_{\lambda_I} \kappa(\lambda_I) O_{\gamma_I}^{\lambda_I}$$

où $\lambda_I \in \partial^1(F, \mathbf{T}_{I, (F_i)}, \mathbf{U}(n))$.

Introduisons maintenant l'ingrédient principal pour définir le facteur de transfert : pour $i \neq j$ dans I on définit un entier r_{ij} par une des égalités

$$\begin{aligned} r_{ij} &:= \text{val}_{F'}(\text{Res}(P_i, P_j)) = f_i \text{val}_{E'_i}(P_j(\gamma_i)) = f_j \text{val}_{E'_j}(P_i(\gamma_j)) \\ &= [\mathcal{O}_{F'}[\gamma_i] \oplus \mathcal{O}_{F'}[\gamma_j] : \mathcal{O}_{F'}[\gamma_i \oplus \gamma_j]] \end{aligned}$$

où P_i désigne le polynôme minimal de γ_i et f_i le degré résiduel de F_i .

THÉORÈME 1.6 (Laumon, Ngô). — *Supposons $p > n$. Avec les notations ci-dessus, on a l'égalité*

$$O_{\gamma_I}^{\kappa} = (-1)^r q^r SO_{\gamma_I}$$

où

$$r = \sum_{i_1 \in I_1, i_2 \in I_2} r_{i_1 i_2}.$$

Notons que l'entier r s'interprète aussi comme le nombre d'intersection des germes de courbes $\text{Spf}(A_{I_1} = \mathbb{F}_{q^2}[[X]](T)/P_{I_1}(T))$ et $\text{Spf}(A_{I_2} = \mathbb{F}_{q^2}[[X]](T)/P_{I_2}(T))$ dans le germe de surface $\text{Spf}(\mathbb{F}_{q^2}[[X, T]])$.

2. FIBRATION DE HITCHIN ET FORMULE DES TRACES

2.1. Quelques mots sur les champs algébriques

Il est hors de question de donner une définition en forme, pour laquelle nous renvoyons à [29], mais d'en donner un début d'intuition, indispensable pour la suite. Rappelons qu'un groupoïde est une (petite) catégorie dont toutes les flèches sont inversibles. Soient S un schéma et Sch/S la catégorie des S -schémas que l'on munit de la topologie étale. Un S -champ \mathfrak{X} est un « faisceau en groupoïdes » sur Sch/S . Ceci veut signifier qu'à tout S -schéma $T \rightarrow S$ on associe un groupoïde $\mathfrak{X}(T)$, à tout S -morphisme $T' \rightarrow T$ un foncteur $\mathfrak{X}(T) \rightarrow \mathfrak{X}(T')$ avec des compatibilités « à isomorphisme canonique près » pour la composition, et de sorte que pour tout S -morphisme étale surjectif $T' \xrightarrow{f} T$, les données de descente sur les morphismes et sur les objets de $\mathfrak{X}(T')$ relativement à f soient effectives. Les S -champs sont les objets d'une 2-catégorie dont les 1 et 2-morphismes sont faciles à deviner.

Champs quotients : Lorsque G est un S -schéma en groupes opérant sur un S -schéma X , on peut définir un champ quotient $[X/G]$. Si T est un S -schéma, $[X/G](T)$ est la catégorie dont les objets sont les couples (P, α) , où P est un G_T -torseur sur T et $\alpha : P \rightarrow X_T$ est un T -morphisme G -équivariant. Un morphisme $(P, \alpha) \rightarrow (P', \alpha')$ est un morphisme de toseurs β tel que $\alpha'\beta = \alpha$. Enfin si $T' \rightarrow T$, « le » foncteur $[X/G](T) \rightarrow [X/G](T')$ est donné par produit fibré. Comme cas particulier, on a $X = S$ avec action triviale ; le champ quotient $[S/G]$ est alors appelé le classifiant et parfois noté $B(G/S)$.

Remarque : Se donner une flèche α comme ci-dessus revient à se donner une section $T \rightarrow X_T \times_T^G P$ de T dans le produit G -contracté de X_T et P .

Exemples : Si $G = \mathbb{G}_m$, se donner un morphisme $T \rightarrow B(\mathbb{G}_m) = B(\mathbb{G}_m/\text{Spec}(\mathbb{Z}))$ revient à se donner un fibré inversible sur T . Si $G = GL(n)$, un morphisme $T \rightarrow B(GL(n))$ est représenté par un fibré vectoriel de rang n .

La notion de champ *algébrique* (au sens d'Artin) est plus délicate à introduire rapidement. Contentons-nous de dire qu'un S -champ algébrique est à un S -champ ce qu'un S -espace algébrique est à un faisceau sur Sch/S : un objet « de nature géométrique », auquel on peut associer un topos muni d'un faisceau d'anneaux structural, des modules quasi-cohérents, des groupes de Chow, de K -théorie... et surtout un formalisme de cohomologie étale semblable à celui des schémas, mais sensiblement moins complet⁽⁵⁾. Il y a une notion renforcée d'algébricité due à Deligne-Mumford, où l'on demande grosso modo que les groupes d'automorphismes des objets soient finis.

Exemples : La définition d'algébricité montre immédiatement qu'un quotient $[X/G]$ est algébrique (resp. de Deligne-Mumford) si G est lisse (resp. étale), séparé et de présentation finie. Plus difficile est le résultat suivant (qui requiert la théorie des schémas Quot de Grothendieck) : supposons de plus que G se plonge dans $GL(n)_S$

⁽⁵⁾Signalons ici que Laumon et Ngô admettent certaines propriétés de ce formalisme qui ne se trouvent pas nécessairement dans la littérature.

et soit $S \xrightarrow{\pi} Z$ un morphisme *projectif* tel qu'après tout changement de base $Z' \rightarrow Z$, on ait $\mathcal{O}_{Z'} \xrightarrow{\sim} \pi_*(\mathcal{O}_{S'})$. Alors, le Z -champ «poussé en avant» $\pi_*[X/G]$ est algébrique. *Mise en garde* : supposons que le préfaisceau quotient X/G soit représentable par un schéma. On a alors une flèche canonique $[X/G] \rightarrow X/G$ qui n'est un isomorphisme (*i.e.* une équivalence de catégories) *que si G agit librement*.

2.2. L'espace de Hitchin

2.2.1. Les objets. — On pose $k = \mathbb{F}_q$ et on en fixe une clôture algébrique \bar{k} . Soit X une courbe projective lisse et géométriquement connexe sur k . On note $\eta = \text{Spec}(F)$ son point générique. Les places de F correspondent aux points fermés de X ; on notera donc $|X|$ à la place de $\mathcal{P}(F)$. On se donne ensuite un schéma en groupes lisse à fibres réductives connexes G sur X . Son algèbre de Lie, $\text{Lie}(G)$, est un \mathcal{O}_X -module localement libre dont on note \mathfrak{g} le fibré vectoriel associé. Enfin, on se donne un diviseur effectif $D = \sum_{v \in |X|} d_v[v]$ sur X dont on note T_D «le» \mathbb{G}_m -torseur associé.

Le groupe multiplicatif \mathbb{G}_m opère par homothéties sur \mathfrak{g} , ce qui permet de tordre \mathfrak{g} par le diviseur D , en posant $\mathfrak{g}_D := \mathfrak{g} \times^{\mathbb{G}_m} T_D$ qui est le fibré vectoriel associé à $\text{Lie}(X) \otimes \mathcal{O}_X(D)$. Comme cette action de \mathbb{G}_m commute à l'action adjointe de G , on tord de la même manière le X -champ quotient $[\mathfrak{g}/G]$ par D et on note $[\mathfrak{g}/G]_D$ le X -champ ainsi obtenu.

Exemple : Dans le cas $G = GL(n) \times_k X$, le groupoïde $[\mathfrak{g}/G]_D(S)$ pour $S \rightarrow X$ classe les paires (\mathcal{E}, ϕ) où \mathcal{E} est un \mathcal{O}_S -module localement libre de rang n et $\phi : \mathcal{E} \rightarrow \mathcal{E}(D)$ un morphisme \mathcal{O}_S -linéaire.

2.2.2. Un comptage remarquable. — En vertu d'un théorème de Lang selon lequel l'ensemble pointé $H^1(\mathcal{O}_v, G) = 0$ est trivial, la suite d'ensembles pointés

$$H^1(X, G) \longrightarrow H^1(F, G) \longrightarrow \bigoplus_{v \in |X|} H^1(F_v, G)$$

est exacte. Nous supposons dorénavant pour simplifier certains énoncés que le noyau $\ker(H^1(F, G) \rightarrow \bigoplus_{v \in |X|} H^1(F_v, G))$ est trivial. Ceci est par exemple vérifié pour les groupes semi-simples adjoints ou simplement connexes, et pour les groupes unitaires.

Le point de départ de l'approche de Ngô est l'«observation» suivante :

PROPOSITION 2.1 (Ngô [33]). — *Le groupoïde $[\mathfrak{g}/G]_D(X)$ est équivalent au groupoïde \mathcal{C}_D dont les objets sont les couples $(\gamma, (g_v)_{v \in |X|})$ où*

- $g_v \in G(F_v)/G(\mathcal{O}_v)$ presque partout triviaux,
- $\gamma \in \mathfrak{g}(F)$, tel que $ad(g_v)^{-1}\gamma \in \varpi_v^{-d_v}\mathfrak{g}(\mathcal{O}_v)$

et où $\text{Hom}_{\mathbb{C}}((\gamma, (g_v)_{v \in |X|}), (\gamma', (g'_v)_{v \in |X|})) := \{x \in G(F), \gamma' = ad(x)\gamma \text{ et } g'_v = xg_v\}$, la composition étant donnée par la multiplication dans $G(F)$.

Démonstration. — Soit $(\gamma, (g_v)_{v \in |X|})$ un couple comme dans l'énoncé. Si U est un ouvert de X tel que g_v soit trivial pour $v \in |U|$, la donnée des g_v permet de recoller «à la Beauville-Lazslo» le G_U -torseur trivial G_U aux toseurs triviaux $G_v, v \in X \setminus U$. On

obtient ainsi un G -torseur E sur X et la condition sur les g_v permet d'identifier γ à une section ϕ de $\text{Lie}(\text{Aut}(E)) \otimes \mathcal{O}_X(D)$, c'est-à-dire à une section de $\mathfrak{g}_D \times^G E$. La paire (E, ϕ) ainsi obtenue est un objet de $[\mathfrak{g}/G]_D(X)$, qui est bien défini à isomorphisme unique près puisque E est par construction trivialisé sur F . Le procédé est donc fonctoriel et on vérifie facilement qu'il est pleinement fidèle et essentiellement surjectif. \square

Pour voir se dessiner une conséquence remarquable, notons \mathcal{C}_D^{ssreg} , resp. \mathcal{C}_D^{ell} , la sous-catégorie pleine de \mathcal{C}_D correspondant à la condition $\gamma \in \mathfrak{g}(F)^{ssreg}$, resp. $\gamma \in \mathfrak{g}(F)^{ell}$. Alors si le centre de G_η est anisotrope, \mathcal{C}_D^{ell} est un groupoïde essentiellement fini de cardinal

$$|\mathcal{C}_D^{ell}| = \sum_{\gamma \in \mathfrak{g}(F)_{/ad}^{ell}} \int_{G(\mathbb{A}_F)/G_\gamma(F)} \mathbf{1}_D(\text{ad}(x)\gamma) dx$$

si la mesure sur $G(\mathbb{A}_F)$ vaut 1 sur $\prod_v G(\mathcal{O}_v)$ et $\mathbf{1}_D$ désigne la fonction $\otimes_v \mathbf{1}_{\varpi_v^{-d_v} \mathcal{O}_v}$. On retrouve donc l'analogie pour l'algèbre de Lie de la partie elliptique de la formule des traces, telle qu'elle apparaît dans les travaux de Waldspurger [35] !

Pour rendre utilisable cette formule, posons $\mathcal{M}_D := (\pi_X)_*([\mathfrak{g}/G]_D)$ le k -champ poussé en avant par $X \xrightarrow{\pi_X} \text{Spec}(k)$. Il associe donc à tout k -schéma S le groupoïde des paires (E, ϕ) où E est un G -torseur sur $X \times_k S$ et ϕ une section de $\text{Lie}(\text{Aut}(E)) \otimes \mathcal{O}_X(D)$. Une telle paire est dite « de Hitchin ». \mathcal{M}_D est un champ algébrique, cf. 2.1, appelé « espace de Hitchin ». Il n'est pas de type fini mais on aimerait définir un sous- k -champ de type fini \mathcal{M}_D^{ell} tel que l'équivalence de la proposition 2.1 induise une équivalence $\mathcal{C}_D^{ell} \xrightarrow{\sim} \mathcal{M}_D^{ell}(k)$. Ceci permettrait d'utiliser une formule cohomologique de type Grothendieck-Lefschetz pour étudier les termes elliptiques de la formule des traces ci-dessus.

2.3. La fibration de Hitchin

2.3.1. Un théorème de Chevalley. — Donnons-nous un groupe réductif connexe déployé \mathbf{G}_0 sur k et choisissons un tore maximal \mathbf{T}_0 de groupe de Weyl W_0 défini sur k . Nous supposons toujours que la caractéristique p de k est première à l'ordre du groupe de Weyl de \mathbf{G}_0 . Chevalley a montré que l'inclusion d'algèbres de Lie $\mathfrak{t}_0 \hookrightarrow \mathfrak{g}_0$ induit un isomorphisme de \bar{k} -algèbres $\text{Gal}(\bar{k}/k)$ -équivariant $\bar{k}[\mathfrak{t}_0]^{W_0(\bar{k})} \xrightarrow{\sim} \bar{k}[\mathfrak{g}_0]^{\mathbf{G}_0(\bar{k})}$. Cet isomorphisme définit un k -morphisme \mathbf{G}_0 -équivariant $\mathfrak{g}_0 \rightarrow \mathfrak{t}_0/W_0$ qui induit à son tour pour tout corps algébriquement clos K contenant k une bijection $\mathfrak{g}_0(K)_{/ad}^{ssreg} \xrightarrow{\sim} (\mathfrak{t}_0^{reg}/W_0)(K)$, où $\mathfrak{t}_0^{reg} = \mathfrak{t}_0 \cap \mathfrak{g}_0^{reg}$ désigne l'ouvert des éléments W_0 -réguliers. Pour un corps *non*-algébriquement clos, il subsiste une bijection $\mathfrak{g}_0(K)_{/st}^{ssreg} \xrightarrow{\sim} (\mathfrak{t}_0^{reg}/W_0)(K)$ où la source désigne l'ensemble des classes *stables* semi-simples régulières de $\mathfrak{g}_0(K)$.

Exemple : Dans le cas $\mathbf{G}_0 = GL(n)$, on a des isomorphismes

$$k[\mathfrak{t}_0]^{W_0} \simeq k[X_1, \dots, X_n]^{\mathfrak{S}_n} = k[\Sigma_1, \dots, \Sigma_n]$$

où les Σ_i sont les polynômes symétriques « élémentaires ». Soit $x \in GL_n(K)$, et $P_x(T) = T^n + a_1 T^{n-1} + \dots + a_n \in K[T]$ son polynôme caractéristique, alors l'élément de

$(\mathfrak{t}_0/W_0)(K)$ correspondant est défini par $\Sigma_i \mapsto a_i$ pour tout i . Ainsi \mathfrak{t}_0/W_0 peut être vu comme «l’espace des polynômes caractéristiques».

2.3.2. Hypothèses sur G . — Pour mettre en famille la construction de Chevalley, nous supposons dorénavant que G est une forme extérieure d’un groupe constant $\mathbf{G}_0 \times_k X$. Une telle forme est déterminée par un revêtement étale galoisien fini $X' \rightarrow X$ et une action de $\text{Gal}(X'/X)$ sur la donnée radicielle basée $(\mathbb{X}_*, \Delta, \mathbb{X}^*, \Delta^\vee)$ associée à \mathbf{G}_0 . En effet, celle-ci équivaut à une action de $\text{Gal}(X'/X)$ sur $\mathbf{G}_0 \times_X X'$ qui preserve un épingleage, laquelle permet de descendre $\mathbf{G}_0 \times_X X'$ et son épingleage de X' à X . On fixera un tore maximal T . Le groupe $W := \mathcal{N}_G(T)/T$ est étale sur X , isomorphe à $W_0 \times_k X'$ sur X' . Nous dirons par la suite que G est «constant» si $X' = X$.

Exemple des groupes unitaires : Supposons k de caractéristique $\neq 2$. Soit $X' \xrightarrow{\pi} X$ un revêtement étale de degré 2 avec X' géométriquement connexe, et τ l’élément non-trivial de $\text{Gal}(X'/X)$. On définit un X -schéma en groupes du type décrit ci-dessus avec $\mathbf{G}_0 = GL(n)$ en posant pour tout $S \rightarrow X$

$$\mathbf{U}_{X'|X}(n)(S) = \{g \in GL_n(H^0(X' \times_X S, \mathcal{O}_{X' \times_X S})), \Phi_n \cdot \tau({}^t g^{-1}) \cdot \Phi_n^{-1} = g\}$$

où Φ_n est la matrice définie en 1.8.1. Dans ce cas $\mathcal{M}_D(S)$ classifie les triplets $(\mathcal{E}, \Phi, \theta)$ où \mathcal{E} est un fibré vectoriel de rang n sur $X' \times_k S$ muni d’une forme hermitienne, *i.e.* un isomorphisme $\Phi : \mathcal{E} \xrightarrow{\sim} \tau^* \mathcal{E}^\vee$ (dual) tel que ${}^t \Phi = \tau^* \Phi$, et $\theta : \mathcal{E} \rightarrow \mathcal{E}(2D)$ est un morphisme hermitien, *i.e.* tel que $\Phi(2D) \circ \theta + \tau^*({}^t \theta)(2D) \circ \Phi = 0$.

2.3.3. Définition du morphisme de Hitchin. — Sous les hypothèses précédentes et en notant \mathfrak{t} l’algèbre de Lie du tore T , on obtient par descente du théorème de Chevalley de 2.3.1 un X -morphisme G -équivariant $\mathfrak{g} \rightarrow \mathfrak{t}/W$ qui se factorise par le champ quotient $[\mathfrak{g}/G] \rightarrow \mathfrak{t}/W$ ⁽⁶⁾. Faisant agir \mathbb{G}_m par homothéties sur \mathfrak{t} , action qui commute à celle de W , on a aussi une forme tordue $[\mathfrak{g}/G]_D \rightarrow \mathfrak{t}_D/W$. Posons alors $\mathcal{A}_D := \pi_{X*}(\mathfrak{t}_D/W)$ (où $\pi_X : X \rightarrow \text{Spec}(k)$ est le morphisme structural). Puisque X est propre et \mathfrak{t}_D/W est un fibré affine sur X , \mathcal{A}_D est un k -espace affine. En poussant en avant le morphisme de Chevalley par π_X , on obtient la *fibration de Hitchin* $\mathcal{M}_D \xrightarrow{f} \mathcal{A}_D$ qui est un morphisme de k -champs.

Exemple des groupes unitaires (suite) : en globalisant et tordant la description de \mathfrak{t}_0/W_0 pour $\mathbf{G}_0 = GL(n)$, on constate que $(\mathfrak{t}_D/W) \times_X X' \rightarrow X'$ est le fibré affine associé au $\mathcal{O}_{X'}$ -module localement libre $\bigoplus_{i=1}^n \mathcal{O}_{X'}(2iD)$. Pour descendre à X , introduisons le \mathcal{O}_X -module inversible $\mathcal{L}_D := \pi_* (\mathcal{O}_{X'})^{\tau^* = -1} \otimes \mathcal{O}_X(2D)$. Alors on vérifie que $(\mathfrak{t}_D/W) \rightarrow X$ est le fibré affine associé au \mathcal{O}_X -module localement libre $\bigoplus_{i=1}^n \mathcal{L}_D^{\otimes i}$.

On en déduit que \mathcal{A}_D est le k -schéma affine associé au k -espace vectoriel

$$\bigoplus_{i=1}^n H^0(X, \mathcal{L}_D^{\otimes i}) = \bigoplus_{i=1}^n H^0(X', \mathcal{O}_{X'}(2iD))^{\tau^* = (-1)^i}$$

⁽⁶⁾Noter que \mathfrak{t}/W est un X -schéma plus grossier que le champ quotient $[\mathfrak{t}/W]$ et qu’il n’est pas possible de relever ce morphisme en $[\mathfrak{g}/G] \rightarrow [\mathfrak{t}/W]$.

et que le morphisme de Hitchin envoie le triplet hermitien $(\mathcal{E}, \Phi, \theta)$ dans $\mathcal{M}_D(k)$ sur les coefficients (a_1, \dots, a_n) du polynôme caractéristique $P_\theta(T)$ de θ .

2.3.4. *Ouvert des points génériquement semi-simples réguliers.* — Soient \mathcal{A}_D^{reg} l’ouvert de \mathcal{A}_D défini par

$$\mathcal{A}_D^{reg}(S) = \{X \times_k S \xrightarrow{a} \mathfrak{t}_D/W, \text{ t.q. } a|_{\eta \times_k S} \in (\mathfrak{t}_\eta^{reg}/W_\eta)(\eta \times_k S)\}$$

pour tout S -schéma k , et \mathcal{M}_D^{reg} son «image réciproque» par f . Alors l’équivalence de groupoïdes de la proposition 2.1 se restreint à une équivalence $\mathcal{C}_D^{ssreg} \xrightarrow{\sim} \mathcal{M}_D^{reg}(k)$.

2.3.5. *Exemple des groupes unitaires : courbes spectrales.* — Nous avons vu qu’un élément $a \in \mathcal{A}_D(k)$ est un polynôme dont les coefficients sont des sections des puissances d’un \mathcal{O}_X -module inversible \mathcal{L}_D . La courbe spectrale Y_a est par définition le lieu des zéros de ce polynôme dans l’espace total du fibré en droites $\Sigma_D^0 := \mathbb{V}_X(\mathcal{L}_D^{\otimes -1})$ sur X .

Plus généralement et plus précisément, soit $S \xrightarrow{a} \mathfrak{t}_D/W$ un S -point de \mathfrak{t}_D/W . Il est donné par un n -uplet $a = (a_1, \dots, a_n) \in \bigoplus_{i=0}^n H^0(S, \mathcal{O}_S \otimes_{\mathcal{O}_X} \mathcal{L}_D^{\otimes i})$. Complétons-le par la section identité a_0 pour obtenir un élément de $\text{Hom}_{\mathcal{O}_X}(\mathcal{L}_D^{\otimes -n}, \mathcal{O}_S \otimes (\bigoplus_{i=0}^n \mathcal{L}_D^{\otimes i-n}))$. On définit alors le «revêtement spectral» Y_a de S comme le sous- S -schéma fermé de $\mathbb{V}(\mathcal{L}_D^{-1}) \times_X S$ dont l’idéal est engendré par l’image de ce morphisme dans $\mathcal{O}_S \otimes \text{Sym}_{\mathcal{O}_X}(\mathcal{L}_D^{\otimes -1})$. La flèche $Y_a \rightarrow S$ est donc un revêtement fini de degré n . En particulier, pour $a = \text{Id}_{\mathfrak{t}_D/W}$, on a un revêtement spectral universel $Y_{\mathfrak{t}_D/W} \rightarrow \mathfrak{t}_D/W$.

La flèche canonique $X \times_k \mathcal{A}_D \rightarrow \mathfrak{t}_D/W$ fournit donc le revêtement spectral $\Sigma_D^0 \times_k \mathcal{A}_D \supset Y_D \rightarrow X \times_k \mathcal{A}_D$, fini de degré n , et la composée $Y_D \rightarrow \mathcal{A}_D$ est une courbe projective relative, appelée «courbe spectrale universelle». Pour $a \in \mathcal{A}_D(\bar{k})$, on en déduit la courbe spectrale Y_a sur \bar{k} , qui est un revêtement fini de degré n sur $X \times_k \bar{k}$.

On vérifie alors que $a \in \mathcal{A}_D^{reg}(\bar{k})$ si et seulement si Y_a est génériquement étale au-dessus de $X \times_k \bar{k}$, et ceci équivaut encore, lorsque $p > n$, à ce que Y_a soit réduite.

2.4. Centralisateurs et champs de Picard

2.4.1. *Centralisateurs réguliers.* — Reprenons le contexte de 2.3.1 et notons $\chi : \mathfrak{g}_0 \rightarrow \mathfrak{t}_0/W_0$ le morphisme de Chevalley. Le schéma en groupes relatif sur \mathfrak{g}_0 des centralisateurs

$$I_0 := \{(x, g) \in \mathfrak{g}_0 \times \mathbf{G}_0, \text{ad}(g)x = x\}$$

n’est pas plat sur \mathfrak{g}_0 mais sa restriction I_0^{reg} au-dessus de \mathfrak{g}_0^{reg} est lisse et commutative. Par descente fidèlement plate pour le morphisme $\chi|_{\mathfrak{g}_0^{reg}} : \mathfrak{g}_0^{reg} \rightarrow \mathfrak{t}_0/W_0$, on construit un schéma en groupes relatif lisse J_0 sur \mathfrak{t}_0/W_0 muni d’un isomorphisme $\chi^*(J_0)|_{\mathfrak{g}_0^{reg}} \xrightarrow{\sim} I_0^{reg}$. Ce dernier se prolonge de manière unique, «par normalité», en un morphisme $\chi^*(J_0) \rightarrow I_0$. Par ailleurs, Kostant a construit des sections du morphisme de Chevalley $\mathfrak{t}_0/W_0 \rightarrow \mathfrak{g}_0$ d’image contenue dans \mathfrak{g}_0^{reg} ; ainsi pour chaque telle section ξ , on a un isomorphisme $J_0 \xrightarrow{\sim} \xi^* I_0$.

Exemple : Dans le cas $\mathbf{G}_0 = GL(n)$, si $a \in (\mathfrak{t}_0/W_0)(K)$ et $P_a(T) = (T^n + a_1 T^{n-1} + \dots + a_0) \in K[T]$ est le polynôme correspondant, alors $J_{0,a}$ est le K -schéma en groupes

associé au groupe des inversibles de l'algèbre $K[T]/P_a(T)$. En d'autres termes, si $Y_a = \text{Spec}(K[T]/P_a(T)) \xrightarrow{\beta_a} \text{Spec}(K)$, alors $J_{0,a} = \beta_{a*} \mathbb{G}_m$.

Revenons maintenant au contexte de 2.3.2. Notons encore $I \subset \mathfrak{g} \times G$ le schéma des centralisateurs sur lequel G agit par $g(\gamma, h) = (\text{Ad}(g)\gamma, ghg^{-1})$. On obtient en passant au quotient un morphisme représentable de champs $[I/G] \rightarrow [\mathfrak{g}/G]$ que l'on pourrait appeler «classifiant des automorphismes de $[\mathfrak{g}/G]$ » puisque si $S \xrightarrow{e} [\mathfrak{g}/G]$ est un objet de $[\mathfrak{g}/G](S)$ donné par un couple (P, α) comme dans 2.1, alors la fibre $e^*[I/G] \rightarrow S$ est le S -schéma des automorphismes de la paire (P, α) . Notons aussi $[\chi] : [\mathfrak{g}/G] \rightarrow \mathfrak{t}/W$ le morphisme de Chevalley descendu au champ quotient. On construit comme dans le cas constant un schéma en groupes lisse relatif J sur \mathfrak{t}/W , muni d'un morphisme $[\chi]^*(J) \rightarrow [I/G]$ qui est un isomorphisme au-dessus de $[\mathfrak{g}^{reg}/G]$. On peut aussi tout tordre par D pour obtenir J_D sur \mathfrak{t}_D/W et $[\chi]_D^*(J_D) \rightarrow [I/G]_D$. Enfin, on peut adapter [33, par. 2] la construction des sections de Kostant à notre situation relative tordue et obtenir ainsi des sections $\mathfrak{t}_D/W \rightarrow \mathfrak{g}_D^{reg}$; cela suppose l'existence et le choix d'une racine carrée de D .

Exemple des groupes unitaires : La description qu'on a donnée de J_0 dans le cas $GL(n)$ suggère d'utiliser le revêtement spectral. Notons en effet β la flèche $Y'_{\mathfrak{t}_D/W} = Y_{\mathfrak{t}_D/W} \times_X X' \rightarrow \mathfrak{t}_D/W$; alors $J_D = \beta_*(\mathbb{G}_m)^{\tau=-1}$.

2.4.2. Action du champ de Picard relatif. — Commençons par un peu d'*abstract nonsense*. Si \mathcal{T} est un faisceau en groupes commutatifs au-dessus d'un schéma \mathcal{X} , le \mathcal{X} -champ classifiant $B(\mathcal{T}/\mathcal{X})$ est muni d'une structure supplémentaire induite par le produit contracté des \mathcal{T} -torseurs : pour $S \rightarrow \mathcal{X}$, l'ensemble des objets de $B(\mathcal{T}/\mathcal{X})(S)$ est ainsi muni d'une structure de groupe, compatible en un certain sens avec les morphismes. Une telle structure a été formalisée par Deligne et appelée «catégorie de Picard». Une fois convenablement faisceautisée, on obtient la notion de «champ de Picard» dont $B(\mathcal{T}/\mathcal{X})$ est un exemple. On peut aussi définir la notion d'action d'une catégorie de Picard sur une catégorie, qui au niveau des objets revient à la notion habituelle d'action d'un groupe sur un ensemble. Cette notion se prolonge à son tour en celle d'action d'un champ de Picard sur un champ.

Dans notre situation, le morphisme $[\chi]^*(J) \rightarrow [I/G]$ induit une action naturelle du champ de Picard $B(J/(\mathfrak{t}/W))$ sur le champ $[\mathfrak{g}/G]$, au-dessus de \mathfrak{t}/W . Décrivons-la sur les objets. Fixons un X -schéma S , une section $S \xrightarrow{e} [\mathfrak{g}/G]$ donnée par une paire (P, α) comme en 2.1, et une section $S \rightarrow B(J/(\mathfrak{t}/W))$ donnée par un $([\chi] \circ e)^*(J)$ -torseur \mathcal{J} . Le morphisme $[\chi]^*(J) \rightarrow [I/G]$ permet de transférer le $([\chi] \circ e)^*J$ -torseur \mathcal{J} en un $e^*[I/G]$ -torseur \mathcal{I} et puisque $e^*[I/G]$ opère sur la paire (P, α) , on peut tordre celle-ci par \mathcal{I} . Ceci définit l'action sur les objets. Comme d'habitude, on peut tordre la construction par D et pousser par π_X ; on obtient un champ de Picard relatif $\mathcal{P}_D \rightarrow \mathcal{A}_D$ qui agit sur \mathcal{M}_D , relativement au-dessus de \mathcal{A}_D .

2.4.3. Exemple des groupes unitaires. — Dans le cas unitaire, la courbe spectrale permet d'interpréter géométriquement le champ de Picard \mathcal{P}_D et son action sur \mathcal{M}_D . En

effet, l'isomorphisme $J_D \simeq \beta_*(\mathbb{G}_m)^{\tau=-1}$ donné plus haut induit un isomorphisme de champs de Picard $\mathcal{P}_D \simeq (\text{Pic}_{Y'_D/\mathcal{A}_D})^{\tau=\otimes^{-1}}$ où $\text{Pic}_{Y'_D/\mathcal{A}_D}$ désigne le champ de Picard relatif « usuel » des $\mathcal{O}_{Y'_D}$ -modules inversibles. Ainsi, si $S \rightarrow \mathcal{A}_D$, la catégorie $\mathcal{P}_D(S)$ a pour objets les couples (\mathcal{F}, ι) où \mathcal{F} est un $\mathcal{O}_{Y'_D \times_{\mathcal{A}_D} S}$ -module inversible et ι est un isomorphisme $\mathcal{F} \xrightarrow{\sim} \tau^* \mathcal{F}^{-1}$ tel que $\iota = \tau^*(\iota^{\otimes -1})$.

Pour décrire l'action sur le champ \mathcal{M}_D , on va se restreindre à l'ouvert $\mathcal{A}_D^{\text{reg}}$ au-dessus duquel la courbe Y_D est à fibres géométriquement réduites. On note encore Y'_D pour $Y'_D \times_{\mathcal{A}_D} \mathcal{A}_D^{\text{reg}}$ de sorte que $\mathcal{P}_D^{\text{reg}} := \mathcal{P}_D \times_{\mathcal{A}_D} \mathcal{A}_D^{\text{reg}} = (\text{Pic}_{Y'_D/\mathcal{A}_D^{\text{reg}}})^{\tau=\otimes^{-1}}$. Introduisons alors le champ de Picard « compactifié » $\overline{\text{Pic}}_{Y'_D/\mathcal{A}_D^{\text{reg}}}$ qui classifie les $\mathcal{O}_{Y'_D}$ -modules cohérents sans torsion de rang 1 relatifs (*i.e.* qui sont $\mathcal{O}_{\mathcal{A}_D^{\text{reg}}}$ -plats et dont la restriction à chaque fibre de $Y'_D \rightarrow \mathcal{A}_D^{\text{reg}}$ est sans torsion et de rang 1 en tout point générique). Malgré son nom, celui-ci n'est pas un champ de Picard au sens général précédent mais est muni d'une action (par produit tensoriel) du champ de Picard $\text{Pic}_{Y'_D/\mathcal{A}_D^{\text{reg}}}$. Si l'on munit $\overline{\text{Pic}}_{Y'_D/\mathcal{A}_D^{\text{reg}}}$ de la dualité $\mathcal{F} \mapsto \mathcal{F}^\vee := \mathcal{H}om_{\mathcal{O}_{Y'_D}}(\mathcal{F}, \omega_{Y'_D/(X' \otimes \mathcal{A}_D^{\text{reg}})})$ (module dualisant relatif), on peut définir le champ $\overline{\mathcal{P}}_D^{\text{reg}} := (\overline{\text{Pic}}_{Y'_D/\mathcal{A}_D^{\text{reg}}})^{\tau=\vee}$ dont les objets au-dessus de $S \rightarrow \mathcal{A}_D^{\text{reg}}$ sont les couples (\mathcal{F}, ι) avec $\mathcal{F} \in \overline{\text{Pic}}_{Y'_D/\mathcal{A}_D^{\text{reg}}}(S)$ et ι un isomorphisme $\mathcal{F} \xrightarrow{\sim} \tau^* \mathcal{F}^\vee$ tel que $\iota = \tau^*(\iota^\vee)$. L'action précédente par produit tensoriel induit une action de $\mathcal{P}_D^{\text{reg}}$ sur $\overline{\mathcal{P}}_D^{\text{reg}}$.

Maintenant on remarque que si $S \xrightarrow{a} \mathcal{A}_D$ et \mathcal{F} est un $\mathcal{O}_{Y'_a}$ -module sans torsion de rang 1 relatif, alors en notant $p'_a : Y'_a \rightarrow X' \times_k S$ le revêtement spectral de degré n , le $\mathcal{O}_{X' \times_k S}$ -module $\mathcal{E} := p'_{a*}(\mathcal{F})$ est localement libre de rang n . De plus, si $\iota : \mathcal{F} \xrightarrow{\sim} \tau^* \mathcal{F}^\vee$ est comme ci-dessus, alors $\Phi := p'_{a*}(\iota)$ est une structure hermitienne sur \mathcal{E} . Enfin, on définit un morphisme hermitien $\theta : \mathcal{E} \rightarrow \mathcal{E}(2D)$ grâce au morphisme

$$\mathcal{O}_{X' \times_k S}(-2D) \rightarrow p'_{a*}(\mathcal{O}_{Y'_a}) \rightarrow \text{End}_{\mathcal{O}_{X' \times_k S}}(\mathcal{E}),$$

la flèche de gauche venant de la définition de Y'_a comme courbe tracée sur le fibré en droites $\mathbb{V}(\mathcal{O}_{X' \times_k S}(-2D))$. On a donc défini un morphisme $\overline{\mathcal{P}}_D^{\text{reg}} \rightarrow \mathcal{M}_D^{\text{reg}}$.

PROPOSITION 2.2 ([30], prop. 2.6.1). — *Le morphisme de $\mathcal{A}_D^{\text{reg}}$ -champs $\overline{\mathcal{P}}_D^{\text{reg}} \rightarrow \mathcal{M}_D^{\text{reg}}$ défini ci-dessus est un isomorphisme équivariant sous l'action de $(\text{Pic}_{Y'_D/\mathcal{A}_D^{\text{reg}}})^{\tau=\otimes^{-1}}$ identifié à $\mathcal{P}_D^{\text{reg}}$.*

Il s'agit d'une variante d'un résultat original de Hitchin et de Beauville, Narasimhan et Ramanan [5].

2.4.4. Quotients. — Voici une étape délicate des travaux de Laumon et Ngô. On voudrait définir un champ quotient $[\mathcal{M}_D/\mathcal{P}_D]$ au-dessus de \mathcal{A}_D . Rappelons comment on peut construire le champ quotient $[X/G]$ associé à l'action d'un faisceau en groupes sur un faisceau : pour tout schéma test S on considère la catégorie quotient $X(S)/G(S)$ dont les objets sont $X(S)$ et les flèches sont données par l'action de $G(S)$. La règle $S \mapsto X(S)/G(S)$ définit un « pré-champ » dont le champ associé peut se décrire comme on l'a fait en 2.1. Lorsque l'action de G est libre, les catégories quotients sont « discrètes », *i.e.* équivalentes à des ensembles, et le champ quotient est essentiellement un faisceau.

Supposons maintenant que X , resp. G , soit un champ, resp. un champ de Picard. Par analogie ou généralisation⁽⁷⁾ de l'exemple précédent, le quotient sera un 2-champ, au sens de [9], associé au 2-préchamp qui à S associe la 2-catégorie (2-groupeïde, en fait) quotient de la catégorie $X(S)$ par la catégorie de Picard $G(S)$. Mais lorsque G agit «librement» en un certain sens sur X , les 2-catégories quotients sont équivalentes à des 1-catégories et le 2-champ quotient est essentiellement un 1-champ. C'est par exemple le cas si pour tous objets x, g de $X(S) \times G(S)$, le morphisme $Aut_{G(S)}(g) \longrightarrow Aut_{X(S)}(gx)$ est injectif. La catégorie quotient $X(S)/G(S)$ se décrit alors comme suit : les objets sont ceux de $X(S)$ et les flèches $x_1 \longrightarrow x_2$ sont les classes d'équivalence de paires (g, α) où $g \in G(S)$ et $\alpha : gx_1 \xrightarrow{\sim} x_2$, la relation d'équivalence étant induite par les (iso)morphismes de $G(S)$.

Cette condition est vérifiée pour l'action de \mathcal{P}_D^{reg} sur \mathcal{M}_D^{reg} et on peut donc définir un champ quotient $[\mathcal{M}_D^{reg}/\mathcal{P}_D^{reg}]$.

2.5. Étude ponctuelle et pré-stabilisation

Dans cette section, on fixe une «caractéristique» $a \in \mathcal{A}_D^{reg}(k)$ et on s'intéresse à la fibre \mathcal{M}_a de \mathcal{M}_D au-dessus de a . C'est un k -champ algébrique, localement de type fini sur lequel agit le k -champ de Picard $\mathcal{P}_a = \mathcal{P}_D \times_{\mathcal{A}_D, a} k$, lui aussi algébrique et localement de type fini. Si \mathcal{C}_a désigne la sous-catégorie pleine de la catégorie \mathcal{C}_D de la proposition 2.1 dont les objets sont les couples $(\gamma, (g_v)_{v \in |X|})$ tels que $\gamma \in \mathfrak{g}(F)$ s'envoie sur $a|_\eta \in (t^{reg}/W)(F)$ par le morphisme de Chevalley (rappelons que de tels γ forment une classe de conjugaison *stable*), alors l'équivalence de catégories de ce même lemme induit une équivalence $\mathcal{C}_a \xrightarrow{\sim} \mathcal{M}_a(k)$.

On dit que a est *géométriquement elliptique* si la classe stable γ qui lui correspond est elliptique dans $\mathfrak{g}(F \otimes_k \bar{k})$ (et pas seulement dans $\mathfrak{g}(F)$). Avec cette définition, on vérifie [30, 2.8.1] et [33, par. 7]⁽⁸⁾ que de tels a sont les k -points d'un sous- k -schéma ouvert $\mathcal{A}_D^{ell} \subset \mathcal{A}_D^{reg}$, qui est non-vide dès que le degré de D est suffisamment grand. Lorsque $a \in \mathcal{A}_D^{ell}(k)$ et le centre de $G|_F$ est F -anisotrope, le groupeïde $\mathcal{M}_a(k)$ est donc essentiellement fini et son cardinal est donné par

$$|\mathcal{M}_a(k)| = \sum_{\{\gamma \mapsto a\}_{/ad}} \int_{G(\mathbb{A}_F)/G_\gamma(F)} \mathbf{1}_D(ad(x)\gamma) dx,$$

où $\{\gamma \mapsto a\}$ désigne la classe de conjugaison *stable* (régulière elliptique) associée à a . Nous allons voir comment l'action de \mathcal{P}_a permet de pré-stabiliser comme en 1.4 l'expression ci-dessus.

Nous verrons plus loin que lorsque a est elliptique et le centre de G_η est anisotrope, le champ \mathcal{P}_a est de type fini, de sorte que son cardinal $|\mathcal{P}_a(k)|$ est fini. À partir de la

⁽⁷⁾Pour laquelle l'auteur ne connaît pas de référence...

⁽⁸⁾Les références à [33] concernent la version de février 2005 de cet article dont la version définitive sera probablement sensiblement différente.

définition des (2)-catégories quotients, on vérifie alors la formule agréable

$$|\mathcal{M}_a(k)| = |\mathcal{P}_a(k)| |\mathcal{M}_a(k)/\mathcal{P}_a(k)|.$$

Par analogie avec la discussion de 1.4, celle-ci invite donc à étudier le foncteur canonique $\mathcal{M}_a(k)/\mathcal{P}_a(k) \xrightarrow{\text{can}} [\mathcal{M}_a/\mathcal{P}_a](k)$. Celui-ci est pleinement fidèle, puisque $S \mapsto \mathcal{M}_a(S)/\mathcal{P}_a(S)$ est déjà un préchamp. Il suffit donc de déterminer son image essentielle. Introduisons l'ensemble $H^1(k, \mathcal{P}_a)$ des classes d'équivalences de \mathcal{P}_a -torseurs sur k . Le morphisme de 2-champs $[\mathcal{M}_a/\mathcal{P}_a] \longrightarrow [k/\mathcal{P}_a]$ associe à tout objet $m \in [\mathcal{M}_a/\mathcal{P}_a](k)$ un invariant $\text{inv}(m) \in H^1(k, \mathcal{P}_a)$ et l'objet m est dans l'image essentielle de can si et seulement si cet invariant est « nul ». Encore une fois, nous verrons ci-dessous que lorsque a est elliptique et le centre de $G|_F$ est anisotrope, $H^1(k, \mathcal{P}_a)$ est un groupe abélien *fini*. Sous ces hypothèses et suivant la même idée présentée en 1.4, on peut donc effectuer une transformation de Fourier

$$(9) \quad |\mathcal{M}_a(k)| = \frac{|\mathcal{P}_a(k)|}{|H^1(k, \mathcal{P}_a)|} \sum_{\kappa \in H^1(k, \mathcal{P}_a)^*} \sum_{m \in [\mathcal{M}_a/\mathcal{P}_a](k)_{/\sim}} \frac{\langle \text{inv}(m), \kappa \rangle}{|Aut(m)(k)|}$$

où la notation $_{/\sim}$ désigne l'ensemble des classes d'isomorphisme d'objets. La ressemblance avec la formule (4) n'est pas seulement formelle. Pour nous en convaincre, nous allons calculer $H^1(k, \mathcal{P}_a)$, décomposer en produit de facteurs locaux et vérifier que pour $\kappa = 1$ on obtient bien l'intégrale orbitale *stable*.

Commençons par le groupe $H^1(k, \mathcal{P}_a)$. Rappelons que si H est un groupe algébrique lisse sur k , alors l'application canonique $H^1(k, H) \longrightarrow H^1(k, \pi_0(H))$ est un isomorphisme, car k est de dimension cohomologique ≤ 1 . De même le champ de Picard \mathcal{P}_a a une composante neutre \mathcal{P}_a^0 et un faisceau quotient « des composantes connexes » $\pi_0(\mathcal{P}_a) := \mathcal{P}_a/\mathcal{P}_a^0$. Ngô montre alors l'isomorphisme $H^1(k, \mathcal{P}_a) \xrightarrow{\sim} H^1(k, \pi_0(\mathcal{P}_a)) = \pi_0(\mathcal{P}_a)(\bar{k})_\sigma$ où σ désigne le Frobenius et la notation en indice les coinvariants. En particulier, pour a elliptique, on en déduit que $\frac{|\mathcal{P}_a(k)|}{|H^1(k, \mathcal{P}_a)|} = |\mathcal{P}_a^0(k)|$. Il reste donc à décrire $\pi_0(\mathcal{P}_a)$.

Notons $J_a := a^*J$ le schéma en groupes sur X déduit de $a : X \longrightarrow \mathfrak{t}/W$; sa fibre générique $J_{a,\eta}$ est F -isomorphe au tore centralisateur d'un élément quelconque de la classe stable dans $\mathfrak{g}(F)$ associée à a . Pour un groupe algébrique H défini sur un corps K dont on a fixé une clôture algébrique \bar{K} , convenons de désigner par $X_*(H)$ le groupe des \bar{K} -cocaractères de H , qui est donc muni d'une action de $Gal(\bar{K}/K)$. Notons $\bar{\Gamma} := Gal(\bar{F}|F \otimes_k \bar{k}) \subset \Gamma = Gal(\bar{F}/F)$.

PROPOSITION 2.3 ([33], sect. 6). — *Il existe un morphisme surjectif*

$$\xi_a : X_*(J_{a,\eta})_{\bar{\Gamma}} \longrightarrow \pi_0(\mathcal{P}_a)(\bar{k})$$

qui induit donc un morphisme surjectif $X_(J_{a,\eta})_{\Gamma} \longrightarrow H^1(k, \mathcal{P}_a)$. Ces morphismes sont des isomorphismes si le X -schéma J_a est à fibres géométriques connexes, ce qui est par exemple vérifié lorsque \mathbf{G}_0 est semi-simple de type adjoint, ou lorsque $\mathbf{G}_0 = GL(n)$.*

Démonstration. — Puisque J_a est lisse sur X , il admet une composante neutre J_a^0 . Considérons le k -champ ${}^0\mathcal{P}_a$ poussé en avant par π_X du classifiant des J_a^0 -torseurs. Ngô montre qu'on a une suite exacte de groupes abéliens

$$H^0(X \times_k \bar{k}, \pi_0(J_a)) \longrightarrow \pi_0({}^0\mathcal{P}_a)(\bar{k}) \longrightarrow \pi_0(\mathcal{P}_a)(\bar{k}) \longrightarrow 0.$$

Considérons ensuite le modèle de Néron *connexe* J'_a qui prolonge $J_{a,\eta}$ sur X et \mathcal{P}'_a le k -champ obtenu en poussant en avant son classifiant par π_X . On montre que le morphisme canonique $J_a^0 \longrightarrow J'_a$ induit un isomorphisme $\pi_0({}^0\mathcal{P}_a)(\bar{k}) \xrightarrow{\sim} \pi_0(\mathcal{P}'_a)(\bar{k})$. Il fait ensuite appel à un résultat de Kottwitz pour prouver l'existence d'un isomorphisme « canonique » $X_*(J_{a,\eta})_{\bar{\Gamma}} \xrightarrow{\sim} \pi_0(\mathcal{P}'_a)(\bar{k})$. \square

On en déduit comme annoncé plus haut que $\pi_0(\mathcal{P}_a)$ et $H^1(k, \mathcal{P}_a)$ sont des groupes abéliens finis si et seulement si $a \in \mathcal{A}_D^{ell}(k)$ et le centre de G_η est F -anisotrope.

Pour décomposer la formule (9) en produit de facteurs locaux, introduisons les analogues locaux $\mathcal{M}_{a,v}$ et $\mathcal{P}_{a,v}$ pour $v \in |X|$, définis de la même manière que \mathcal{M}_a et \mathcal{P}_a mais en remplaçant X par son complété formel X_v en v . On a alors le spectaculaire

THÉORÈME 2.4 ([33], th. 4.5). — *Pour $a \in \mathcal{A}_D^{reg}(k)$, le morphisme de k -champs « de restriction »*

$$[\mathcal{M}_a/\mathcal{P}_a] \longrightarrow \prod_v [\mathcal{M}_{a,v}/\mathcal{P}_{a,v}]$$

est un isomorphisme. Plus précisément, pour tout \bar{k} -schéma S , le foncteur de « restriction » $\mathcal{M}_a(S)/\mathcal{P}_a(S) \longrightarrow \prod_v \mathcal{M}_{a,v}(S)/\mathcal{P}_{a,v}(S)$ est une équivalence de catégories. Dans les deux cas, les facteurs associés aux $v \in |X|$ dont l'image par $a : X \longrightarrow \mathfrak{t}/W$ est dans \mathfrak{t}_v^{reg}/W_v (presque tous puisque $a \in \mathcal{A}_D^{reg}(k)$), sont triviaux.

Démonstration. — (idée) Expliquons d'abord la dernière assertion. Une section de Kostant $\mathfrak{t}/W \xrightarrow{\xi} \mathfrak{g}^{reg}$ induit un \mathfrak{t}/W -isomorphisme $(G \times (\mathfrak{t}/W))/J \xrightarrow{\sim} \mathfrak{g}^{reg}$, d'où un isomorphisme $[(\mathfrak{t}/W)/J] = B(J/(\mathfrak{t}/W)) \xrightarrow{\sim} [\mathfrak{g}^{reg}/G]$. Ce dernier est $B(J/(\mathfrak{t}/W))$ -équivariant et montre donc que l'action du champ de Picard $B(J/(\mathfrak{t}/W))$ sur $[\mathfrak{g}^{reg}/G]$ au-dessus de \mathfrak{t}/W est « simplement transitive ». Ainsi lorsque $a(v)$ est dans \mathfrak{t}^{reg}/W , le champ $\mathcal{M}_{a,v}$ est un sous-champ de $\pi_{X*}([\mathfrak{g}^{reg}/G])$ et par conséquent $\mathcal{P}_{a,v}$ agit « simplement transitivement » sur $\mathcal{M}_{a,v}$ de sorte que $[\mathcal{M}_{a,v}/\mathcal{P}_{a,v}]$ est *trivial* (ce genre d'arguments se trouve aussi dans [10]).

Passons à l'étude de $\mathcal{M}_a(S)/\mathcal{P}_a(S) \longrightarrow \prod_v \mathcal{M}_{a,v}(S)/\mathcal{P}_{a,v}(S)$. La catégorie de gauche a pour objets les paires de Hitchin (E, ϕ) sur $X \times_k S$ de caractéristique a et pour morphismes $(E, \phi) \longrightarrow (E', \phi')$ les classes d'équivalences de couples (\mathcal{J}, α) où \mathcal{J} est un J_a -torseur et $\alpha : \mathcal{J}(E, \phi) \xrightarrow{\sim} (E', \phi')$. Idem pour les termes du produit en remplaçant X par X_v . Dans la preuve de Ngô, il faut d'abord *choisir un point-base* dans $\mathcal{M}_a(S)$, par exemple l'image de a par une section de Kostant. On introduit ensuite une variante $\mathcal{M}_{a,v}^\bullet$ de $\mathcal{M}_{a,v}$ où l'on rajoute une rigidification par ce point-base au point générique de X_v . Comme on tue ainsi les automorphismes, on obtient un faisceau, qui n'est autre que celui qui définit la fibre de Springer affine associée au point-base.

De même on définit une version rigidifiée $\mathcal{P}_{a,v}^\bullet$ de $\mathcal{P}_{a,v}$. On a alors un foncteur d'oubli des rigidifications $\mathcal{M}_{a,v}^\bullet(S)/\mathcal{P}_{a,v}^\bullet(S) \rightarrow \mathcal{M}_{a,v}(S)/\mathcal{P}_{a,v}(S)$ qui est une équivalence de catégories. Par recollement formel au point-base choisi, on a un foncteur dans l'autre sens $\prod_v \mathcal{M}_{a,v}^\bullet(S)/\mathcal{P}_{a,v}^\bullet(S) \rightarrow \mathcal{M}_a(S)/\mathcal{P}_a(S)$, qui est pleinement fidèle. Lorsque S est un \bar{k} -schéma, on montre qu'il est aussi essentiellement surjectif. \square

Ce théorème permet de réécrire la formule (9) sous la forme

$$|\mathcal{M}_a(k)| = |\mathcal{P}_a^0(k)| \sum_{\kappa \in (\pi_0(\mathcal{P}_a)(\bar{k}))^\sigma} \prod_{v \in |X|} \left(\sum_{m_v \in [\mathcal{M}_{a,v}/\mathcal{P}_{a,v}](k)_{/\sim}} \frac{\langle \text{inv}(m_v), \kappa \rangle}{|\text{Aut}(m_v)(k)|} \right),$$

où la définition de l'accouplement local $\langle \text{inv}(m_v), \kappa \rangle$ dépend du choix d'une «section» $[\mathcal{M}_{a,v}/\mathcal{P}_{a,v}] \rightarrow [\mathcal{M}_a/\mathcal{P}_a]$ qui est obtenue comme dans la preuve ci-dessus en introduisant des rigidifications génériques par un point-base dans $\mathcal{M}_a(k)$. Pour faire un tel choix de manière «uniforme en a », on utilise une section de Kostant. Un comptage comme dans [11] montre alors que les termes locaux sont les κ -intégrales orbitales des fonctions $\mathbf{1}_{\varpi_v^{-d_v} \mathfrak{g}(\mathcal{O}_v)}$ sur l'orbite associée à a :

$$\sum_{m_v \in [\mathcal{M}_{a,v}/\mathcal{P}_{a,v}](k)_{/\sim}} \frac{\langle \text{inv}(m_v), \kappa \rangle}{|\text{Aut}(m_v)(k)|} = O_a^\kappa(\mathbf{1}_{\varpi_v^{-d_v} \mathfrak{g}(\mathcal{O}_v)}),$$

à condition de normaliser la mesure de Haar sur $J_a(F_v)$ par $\text{vol}(J_a(\mathcal{O}_v)) = 1$.

L'action de \mathcal{P}_a a donc à la fois permis de pré-stabiliser et de rendre «locale» l'expression de $|\mathcal{M}_a(k)|$. Ce qu'on a gagné par rapport à la situation de 1.4, c'est la possibilité d'utiliser des outils cohomologiques, comme par exemple :

PROPOSITION 2.5. — Soit $\mathcal{M}_{\bar{a}} := \bar{k} \otimes_k \mathcal{M}_a$.

1. L'action de \mathcal{P}_a sur \mathcal{M}_a induit une action de $\pi_0(\mathcal{P}_a)(\bar{k})$ sur les groupes de cohomologies $H^n(\mathcal{M}_{\bar{a}}, \mathbb{Q}_l)$.
2. Pour tout caractère $\kappa : \pi_0(\mathcal{P}_a)(\bar{k}) \rightarrow \bar{\mathbb{Q}}_l^*$, notons $H^n(\mathcal{M}_{\bar{a}}, \bar{\mathbb{Q}}_l)_\kappa$ la partie κ -équivariante de la cohomologie. Si $\kappa = \sigma(\kappa)$, on a

$$\sum_n (-1)^n \text{Tr}(\sigma, H^n(\mathcal{M}_{\bar{a}}, \bar{\mathbb{Q}}_l)_\kappa) = |\mathcal{P}_a^0(k)| \sum_{m \in [\mathcal{M}_a/\mathcal{P}_a](k)_{/\sim}} \frac{\langle \text{inv}(m), \kappa \rangle}{|\text{Aut}(m)(k)|}$$

et si $\kappa \neq \sigma(\kappa)$, alors la somme de gauche est nulle.

Démonstration. — Le i) est un «lemme d'homotopie» [30, 3.2.3] qui montre que la partie connexe de \mathcal{P}_a agit trivialement sur la cohomologie. Le ii) est une variation sur la formule de points fixes de Grothendieck-Lefschetz [30, App. C] (notons que \mathcal{M}_a est un k -champ de Deligne-Mumford, comme on le verra ci-dessous). \square

Néanmoins, en ayant fixé a comme on l'a fait, on se retrouve au même point que dans les approches locales de Goresky, Kottwitz, MacPherson [11] et Laumon [28] : leurs stratégies pour étudier la cohomologie et la comparer aux analogues endoscopiques requièrent une propriété de *pureté* qui pour l'instant est hors de portée. Pour contourner

le problème, l'idée introduite par Ngô consiste à faire une étude locale au voisinage de a dans \mathcal{A}_D^{ell} .

2.6. Étude locale et endoscopie

On suppose toujours le centre de G_η anisotrope. La proposition suivante est espérée en toute généralité, mais est pour l'instant prouvée au cas par cas pour les groupes classiques.

PROPOSITION 2.6. — 1. \mathcal{M}_D^{reg} est lisse sur k ([30, 2.5.2] et [33]).
 2. $\mathcal{M}_D^{ell} \xrightarrow{f^{ell}} \mathcal{A}_D^{ell}$ est propre de type fini (cas par cas, [30, 2.8.1], [33] et [32]).
 3. \mathcal{M}_D^{ell} est un k -champ de Deligne-Mumford (cas par cas [30, 2.8.1], [33] et [32]).

La preuve de la lissité utilise des travaux de Biswas et Ramanan [6], et les deux autres points apparaissent aussi chez Faltings, [10].

Considérons maintenant $f_*^{ell} \mathbb{Q}_l := Rf_*^{ell} \mathbb{Q}_l$. Par le troisième point, c'est un objet de $D_c^b(\mathcal{A}_D^{ell}, \mathbb{Q}_l)$. D'après les deux premiers points, c'est un complexe *pur* de poids 0 au sens de Deligne. En particulier, ses faisceaux de cohomologie *perverse* ${}^p\mathcal{H}^n(f_*^{ell} \mathbb{Q}_l)$, $n \in \mathbb{N}$ sont *purs* de poids n .

Soit maintenant R , resp \bar{R} , l'enséhlisé, resp. l'enséhlisé strict, de \mathcal{A}_D^{ell} en a ; notons $\mathcal{M}_R, \mathcal{P}_R$, etc. les restrictions à R de $\mathcal{M}_D, \mathcal{P}_D$, etc. Par le même lemme d'homotopie que pour le (1) de la proposition 2.5, on a une action de $\pi_0(\mathcal{P}_R)(\bar{R}) = \pi_0(\mathcal{P}_a)(\bar{k})$ sur les faisceaux pervers ${}^p\mathcal{H}^n(f_{R,*} \mathbb{Q}_l)_{|\bar{R}}$ et on peut donc définir leur partie κ -isotypique. Lorsque $\kappa = \sigma(\kappa)$, ceci définit un facteur direct ${}^p\mathcal{H}^n(f_{R,*} \mathbb{Q}_l)_\kappa$ de ${}^p\mathcal{H}^n(f_{R,*} \mathbb{Q}_l)$. La considération de ces faisceaux pervers est toute aussi bonne pour le but final, en vertu du

LEMME 2.7 ([30], 3.10.1). — Soit $\kappa : \pi_0(\mathcal{P}_a)(\bar{k}) \longrightarrow \mathbb{C}^*$ fixe par σ . On a l'égalité

$$\sum_n (-1)^n \text{Tr}(\sigma, H^n(\mathcal{M}_{\bar{a}}, \bar{\mathbb{Q}}_l)_\kappa) = \sum_{n,m} (-1)^{n+m} \text{Tr}(\sigma, H^n(({}^p\mathcal{H}^m(f_{R,*} \mathbb{Q}_l)_\kappa)_{\bar{a}})).$$

Un des points-clef pour le lien avec les groupes endoscopiques est l'estimation (élémentaire) du support de ${}^p\mathcal{H}^n(f_{R,*} \mathbb{Q}_l)_\kappa$.

LEMME 2.8 ([33], cor 8.2). — Soit $\kappa : \pi_0(\mathcal{P}_a)(\bar{k}) \longrightarrow \mathbb{C}^*$ fixe par σ . Le support de ${}^p\mathcal{H}^n(f_{R,*} \mathbb{Q}_l)_\kappa$ est le fermé de R formé des points $b \in R$ tels que κ se factorise par le morphisme de restriction

$$\pi_0(\mathcal{P}_a)(\bar{k})_\sigma = \pi_0(\mathcal{P}_R)(\bar{R})_\sigma \longrightarrow \pi_0(\mathcal{P}_b)(\bar{k}(b))_\sigma$$

où $k(b)$ est le corps résiduel en b et $\bar{k}(b) = \bar{k} \otimes k(b)$.

2.6.1. *Groupes endoscopiques.* — Reprenons les notations de 1.5 et adoptons la définition simplifiée de donnée endoscopique qu'on y trouve. Disons qu'une telle donnée (s, ρ) est *géométrique* si ρ se factorise par $\Gamma = \text{Gal}(\bar{\eta}/\eta) \longrightarrow \pi_1(\bar{\eta}, X)$. Le groupe endoscopique $\mathbf{G}_{s,\rho}$ se prolonge alors en un groupe $G_{s,\rho}$ sur X du type décrit en 2.3.2. On peut donc le munir d'une paire $(T_{s,\rho}, B_{s,\rho})$, définir sa fibration de Hitchin, etc. Par construction du groupe endoscopique, on a un morphisme $\mathfrak{t}_{s,\rho}/W_{s,\rho} \longrightarrow \mathfrak{t}/W$, qui est simplement la version géométrique pour les algèbres de Lie du procédé de transfert des classes de conjugaison stables de $\mathbf{G}_{s,\rho}$ vers \mathbf{G} . Ce morphisme induit un morphisme $\mathcal{A}_{D,(s,\rho)} \xrightarrow{\iota_{s,\rho}} \mathcal{A}_D$. Par commodité nous omettrons l'indice D dans la suite.

PROPOSITION 2.9 ([33], prop. 10.1). — *Soit $\mathcal{A}_{s,\rho}^{G\text{-reg}}$ l'image réciproque de \mathcal{A}^{reg} . Le morphisme*

$$\iota_{s,\rho} : \mathcal{A}_{s,\rho}^{G\text{-reg}} \longrightarrow \mathcal{A}^{\text{reg}}$$

est fini et non-ramifié. Les images de ces morphismes sont deux à deux disjointes, lorsque (s, ρ) parcourt les classes d'isomorphisme de données endoscopiques géométriques.

Fixons maintenant un caractère $\kappa : \pi_0(\mathcal{P}_a)(\bar{k}) \longrightarrow \mathbb{C}^*$ invariant par σ , qui induit donc, grâce à la proposition 2.3, un caractère κ de $X_*(J_{a,\eta})_\Gamma$. Nous avons expliqué à la fin de la section 1.5 comment la paire (a, κ) détermine (à isomorphisme près) une donnée endoscopique (s_κ, ρ_a) . Il n'est pas du tout évident qu'une telle donnée soit «géométrique» (ce serait en général faux si κ ne se factorisait pas par $\pi_0(\mathcal{P}_a)(\bar{k})$).

PROPOSITION 2.10 ([33], lem. 10.4). — *La donnée endoscopique (s_κ, ρ_a) est géométrique.*

Démonstration. — On se ramène d'abord par descente au cas où G est constant sur X (donc égal à $\mathbf{G}_0 \times_k X$). Dans le cas où \mathbf{G}_0 est adjoint, on sait (on peut le vérifier sur la définition 1.2 en utilisant le fait que $(W_G)_s = W_{G_s}$) que les groupes endoscopiques de $G_\eta = \mathbf{G}_0 \times_k F$ sont déployés et se prolongent donc en des groupes constants sur X . La difficulté de la preuve de Ngô consiste à se ramener au cas adjoint. Cela utilise un lemme de Kottwitz. \square

Abrégeons $(s, \rho) := (s_\kappa, \rho_a)$, fixons $a' \in \mathcal{A}_{s,\rho}^{G\text{-reg}}(k)$ d'image a dans $\mathcal{A}^{\text{ell}}(k)$ et notons S l'hensélisé de $\mathcal{A}_{s,\rho}^{G\text{-reg}}$ en a' . Par la proposition 2.9, le morphisme $\iota_{s,\rho} : S \longrightarrow R$ est une immersion fermée.

THÉORÈME 2.11 ([33], th. 10.2). — *Le support des faisceaux pervers ${}^p\mathcal{H}^n(f_{R,*}\mathbb{Q}_l)_\kappa$ sur R est inclus dans l'image $\iota_{s,\rho}(S)$.*

Démonstration. — Comme dans le résultat précédent, la preuve est plus simple dans le cas adjoint et une des difficultés est de se ramener à ce cas. Nous n'esquissons les arguments que dans ce cas \mathbf{G} adjoint, de sorte qu'on a l'isomorphisme de la proposition 2.3 $X_*(J_{a,\eta})_\Gamma \xrightarrow{\sim} \pi_0(\mathcal{P}_a)(\bar{k})_\sigma$. Plus généralement, pour un point b pas nécessairement fermé de R , on a encore un isomorphisme $X_*(J_{b,\eta})_{\Gamma_b} \xrightarrow{\sim} \pi_0(\mathcal{P}_b)(\bar{k}(b))_\sigma$ qui est indépendant

du choix de la clôture algébrique de $F \otimes k(b)$ telle que $\Gamma_b := \text{Gal}(\overline{F \otimes_k k(b)} / F \otimes_k k(b))$. De plus le morphisme de restriction $\pi_0(\mathcal{P}_a)(\overline{k})_\sigma \rightarrow \pi_0(\mathcal{P}_b)(\overline{k}(b))_\sigma$ correspond via ces isomorphismes au morphisme canonique $X_*(J_{a,\eta})_\Gamma \rightarrow X_*(J_{b,\eta})_{\Gamma_b}$. Ainsi d’après le lemme 2.8, si b est dans le support de ${}^p\mathcal{H}^n(f_{R,*}\mathbb{Q}_l)_\kappa$, alors κ , vu comme caractère de $X_*(J_{b,\eta})$, doit être fixe par Γ_b .

Traduisons en langage endoscopique : si $X_*(J_{b,\eta}) \xrightarrow{\sim} \mathbb{X}_*$ est un isomorphisme dans la W_G -classe canonique qui envoie κ sur s , alors le cocycle $\Gamma_b \rightarrow W_G \rtimes \Gamma_b$ qui décrit l’action de Γ_b transportée sur \mathbb{X}_* a son image dans le fixateur $(W_G \rtimes \Gamma_b)_s$ de s , et ce dernier coïncide avec son sous-groupe $W_{G_s} \rtimes \Gamma_b$ car le groupe est supposé adjoint.

Ceci équivaut à dire que le $F \otimes k(b)$ -point $b|_{\eta \times k(b)}$ de \mathfrak{t}/W se relève à $\mathfrak{t}_{s,\rho}/W_{s,\rho}$. Par adhérence schématique, on en déduit que la $k(b)$ -section $b : X \times k(b) \rightarrow \mathfrak{t}/W$ se relève en une section $b' : X \times k(b) \rightarrow \mathfrak{t}_{s,\rho}/W_{s,\rho}$. \square

Voici une conséquence importante : le faisceau pervers $i_{s,\rho}^*({}^p\mathcal{H}^n(f_{R,*}\mathbb{Q}_l)_\kappa)$ est pur de poids n . La conjecture suivante est un analogue géométrique de la conjecture de transfert pour les fonctions-test $\mathbf{1}_D$:

CONJECTURE 2.12 (Laumon, Ngô). — *(Pour les groupes classiques.) Il existe un système local \mathcal{L} sur S en \mathbb{Z} -modules libres de rang 1 et quadratique, un entier r et un isomorphisme de faisceaux pervers gradués*

$${}^p\mathcal{H}^\bullet(f_{S,*}^{s,\rho}\mathbb{Q}_l)_1 \xrightarrow{\sim} \mathcal{L} \otimes {}^p\mathcal{H}^\bullet(f_{R,*}\mathbb{Q}_l)_\kappa[-2r](-r).$$

Ici ${}^p\mathcal{H}^\bullet$ désigne la somme des ${}^p\mathcal{H}^n$ et est gradué par le degré n ; le crochet désigne un décalage dans la graduation. Le lien avec la conjecture de transfert est le suivant : en prenant la trace alternée de Frobenius en la fibre géométrique fermée de chaque côté, on obtient la formule

$$|\mathcal{P}_{s,\rho,a'}^0(k)|SO_{a'}(\mathbf{1}_D) = |\mathcal{P}_a^0(k)|\varepsilon_{\mathcal{L}}q^r O_a^\kappa(\mathbf{1}_D)$$

où $\varepsilon_{\mathcal{L}}$ est un signe. En principe, les termes $|\mathcal{P}^0|$ se simplifient en compensant les deux normalisations différentes des mesures de Haar sur $J_{a,\eta}(\mathbb{A}_F)$ (rappelons que l’une dépend du modèle entier $J_{a'}$ de $J_{a',\eta} = J_{a,\eta}$ et l’autre dépend de J_a). Le facteur $\varepsilon_{\mathcal{L}}q^r$ dépend du diviseur D selon la formule

$$\varepsilon_{\mathcal{L}}q^r = \prod_{v \in |X|} \frac{\Delta_v(\varpi_v^{d_v}\gamma', \varpi_v^{d_v}\gamma)}{\Delta_v(\gamma', \gamma)}$$

où $\gamma' \mapsto a'$, $\gamma \mapsto a$ sont des éléments des classes de conjugaisons respectives et les Δ_v sont les facteurs de transferts locaux. On peut le calculer explicitement à partir de [36]. Le seul cas résolu est celui des groupes unitaires, par Laumon et Ngô.

3. GROUPES UNITAIRES

Dans cette partie nous précisons l'énoncé de la conjecture 2.12 pour les groupes unitaires, et en esquissons la preuve. On expliquera ensuite comment s'en déduit le lemme fondamental. Auparavant, il nous faut expliciter les notions introduites précédemment.

3.1. Énoncé du théorème

3.1.1. Lien avec les deux sections précédentes. — Nous fixerons une fois pour toutes le diviseur D de degré suffisamment grand, et l'omettrons des notations. Soient $a \in \mathcal{A}^{reg}(k)$ et \bar{a} le point géométrique associé. On a alors la courbe spectrale $Y_a \subset \Sigma^0$, voir 2.3.5, qui est un revêtement de degré n de X et son changement de base Y'_a au-dessus de X' . On vérifie que a est géométriquement elliptique si et seulement si l'application $Irr(Y'_a) \rightarrow Irr(Y_{\bar{a}})$ induite sur les ensembles de composantes irréductibles géométriques est *bijjective*. Ensuite, soit par application de la proposition 2.3, soit par un raisonnement plus géométrique [30, 2.8.2], on vérifie que $\pi_0(\mathcal{P}_a)(\bar{k}) \simeq (\mathbb{Z}/2\mathbb{Z})^{Irr(Y_{\bar{a}})}$. Un caractère κ fixe sous σ est donc de la forme $(d_i)_{i \in Irr(Y_a)} \mapsto (-1)^{\sum_{i \in I} d_i}$ pour un sous-ensemble I de $Irr(Y_a)$. Si n_1 désigne la somme des degrés des composantes $Y_{a,i}$, $i \in I$, alors le groupe endoscopique H associé à κ est $U(n_1) \times U(n - n_1 =: n_2)$ et le morphisme $\mathcal{A}_H \rightarrow \mathcal{A}$ s'identifie au morphisme

$$\begin{aligned} \bigoplus_{i=1}^{n_1} H^0(X, \mathcal{L}_D^{\otimes i}) \times \bigoplus_{i=1}^{n_2} H^0(X, \mathcal{L}_D^{\otimes i}) &\rightarrow \bigoplus_{i=1}^n H^0(X, \mathcal{L}_D^{\otimes i}) \\ (a_1, a_2) &\mapsto a \end{aligned}$$

où les composantes de a sont données par l'égalité de polynômes

$$(T^{n_1} + a_{1,1}T^{n_1-1} + \cdots + a_{1,n_1})(T^{n_2} + a_{2,1}T^{n_2-1} + \cdots + a_{2,n_2}) = (T^n + a_1T^{n-1} + \cdots + a_n).$$

3.1.2. Contexte et notations. — Nous nous restreindrons dorénavant au cas où $|Irr(Y_{\bar{a}})| = |Irr(Y_a)| = 2$, et κ est un des deux caractères $\kappa_i : (d_1, d_2) \mapsto (-1)^{d_i}$ pour $i = 1, 2$. Tout ce qui suit peut se généraliser sans difficultés au cas général, et l'était d'ailleurs dans une première version de [30].

Ainsi Y_a est réunion de deux composantes géométriquement irréductibles Y_{a_1} et Y_{a_2} de degrés respectifs n_1 et n_2 avec $n_1 + n_2 = n$. Le groupe endoscopique est donc $H := U(n_1) \times U(n_2)$. Pour retrouver les notations de [30], nous noterons $f : \mathcal{M} \rightarrow \mathcal{A}$, resp. $g : \mathcal{N} \rightarrow \mathcal{A}_H$, la fibration de Hitchin du groupe $G = U(n)$, resp. de son groupe endoscopique $H = U(n_1) \times U(n_2)$, ainsi que $\mathcal{P} \rightarrow \mathcal{A}$, resp. $\mathcal{Q} \rightarrow \mathcal{A}_H$, les champs de Picards respectifs. En fait, comme dans la dernière section de la partie 2, nous fixerons (a_1, a_2) dans la préimage de a et noterons S , resp. R , l'hensélisé de \mathcal{A}_H en (a_1, a_2) , resp. de \mathcal{A} en a . On sait que le morphisme $S \rightarrow R$ est une immersion fermée. Les changements de base à S seront notés par un indice. On a donc les S -champs $f_S : \mathcal{M}_S \rightarrow S$, $g_S : \mathcal{N}_S \rightarrow S$, \mathcal{P}_S , \mathcal{Q}_S et les courbes spectrales relatives $Y_S = Y_{1,S} \cup Y_{2,S}$ tracées sur la même surface réglée $\Sigma_S^0 := \mathbb{V}(\mathcal{L}_D) \times_k S$ au-dessus de $X_S := X \times_k S$.

Nous allons préciser l'énoncé de la conjecture 2.12 pour les groupes unitaires.

3.1.3. *L'entier r et le système local.* — Soit Z_S l'intersection de $Y_{1,S}$ et $Y_{2,S}$ dans Y_S . C'est un sous-schéma fermé de Y_S qui est fini et plat sur S , de degré

$$r = 2n_1n_2\deg(D).$$

On a aussi le revêtement étale quadratique Z'_S au-dessus de Z_S , déduit de $X' \rightarrow X$. Ce revêtement définit un système local $L_{Z'/Z}$ en \mathbb{Z} -modules libres de rang 1 et de carré trivial sur Z_S comme le conoyau de la flèche d'adjonction $\mathbb{Z}_{Z_S} \rightarrow pr_{Z'_S \rightarrow Z_S,*}(\mathbb{Z}_{Z'_S})$. Comme Z_S est hensélien, $L_{Z'/Z}$ est géométriquement constant (*i.e.* constant sur $\bar{k} \times_k Z$).

Soit S_{\natural} l'ouvert de S dont les points géométriques b sont ceux où les courbes $Y_{1,b}$ et $Y_{2,b}$ se coupent *transversalement* dans Σ_S^0 et où, en tout point de l'intersection Z_b , les revêtements $Y_{i,b} \rightarrow X_b$ sont *étales*. Cet ouvert est non-vide d'après [30, 2.9.1]. En particulier, l'intersection $Z_{S_{\natural}}$ est *étale* au-dessus de S_{\natural} . On définit alors

$$L_{Z'_S/Z_S/S_{\natural}} := \left(\bigwedge^r pr_{Z_{S_{\natural}} \rightarrow S_{\natural},*} L_{Z'_S/Z_S} \right)^{\otimes -1} \otimes \left(\bigwedge^r pr_{Z_{S_{\natural}} \rightarrow S_{\natural},*} \mathbb{Z}_{Z_S} \right).$$

C'est un système local en \mathbb{Z} -modules libres de rang 1 de carré trivial qui est géométriquement constant sur S_{\natural} et qui donc se prolonge en un objet similaire $L_{Z'/Z/S}$ sur S tout entier. Un tel objet est entièrement déterminé par la valeur propre de Frobenius, qui est simplement un signe.

Pour comprendre pourquoi une telle chose apparaît dans le théorème ci-dessous, il faudra rentrer un peu dans la preuve.

THÉORÈME 3.1 ([30], th. 3.9.3). — *Pour chacun des deux caractères κ considérés en 3.1.2, il existe un isomorphisme de faisceaux pervers gradués :*

$${}^p\mathcal{H}^\bullet(f_{S,*}\mathbb{Q}_l)_\kappa \xrightarrow{\sim} L_{Z'/Z/S} \otimes_{\mathbb{Z}_S} {}^p\mathcal{H}^\bullet(g_{S,*}\mathbb{Q}_l)_\kappa[-2r](-r).$$

3.2. Éléments de la preuve

Après avoir souligné quelques spécificités du cas unitaire, nous introduirons «le» tore, son action, et la cohomologie équivariante. Ces outils fondamentaux ont été introduits par Goresky, Kottwitz et MacPherson [11] dans leur approche du lemme fondamental via les fibres de Springer. Ici, on a besoin de variantes «relatives» au-dessus de S . Viendra ensuite la «glissade» qui est un analogue de l'argument de déformation que Laumon avait introduit dans [28]. Cet argument ramènera le problème de S à S_{\natural} . Pour résoudre le problème au-dessus de S_{\natural} , on verra apparaître des idées réminiscentes de [31].

3.2.1. *Spécificités du cas unitaire.* — Rappelons en termes imprécis le contenu de la proposition 2.2 :

- Le champ \mathcal{P}_S classe les Y'_S -modules inversibles «unitaires» et agit par produit tensoriel sur le champ \mathcal{M}_S qui classe les Y'_S -modules sans torsion de rang 1 «unitaires».

- Le champ \mathcal{Q}_S classe les $(Y'_{1,S} \sqcup Y'_{2,S})$ -modules inversibles «unitaires» et agit par produit tensoriel sur le champ \mathcal{N}_S qui classe les $(Y'_{1,S} \sqcup Y'_{2,S})$ -modules sans torsion «unitaires» de rang 1.

Notons alors $\nu : Y_{1,S} \sqcup Y_{2,S} \longrightarrow Y_S$ le morphisme de la courbe spectrale endoscopique vers la courbe spectrale. On a donc un morphisme $\mathcal{N}_S \xrightarrow{\nu_*} \mathcal{M}_S$ qui envoie $((\mathcal{F}_1, \iota_1), (\mathcal{F}_2, \iota_2))$ sur $(\mathcal{F}_1 \oplus \mathcal{F}_2, \iota_1 \oplus \iota_2)$ et un morphisme «dans l'autre sens» $\mathcal{P}_S \xrightarrow{\nu^*} \mathcal{Q}_S$ qui consiste simplement à tirer en arrière le Y_S -module inversible unitaire (\mathcal{F}, ι) . De plus, le morphisme ν_* est ν^* -équivant au sens où il existe un isomorphisme canonique

$$\nu_*(\nu^*(\mathcal{F}, \iota_{\mathcal{F}}) \otimes (\mathcal{G}, \iota_{\mathcal{G}})) \xrightarrow{\sim} (\mathcal{F}, \iota_{\mathcal{F}}) \otimes \nu_*(\mathcal{G}, \iota_{\mathcal{G}}).$$

Le morphisme ν_* peut aussi s'expliciter en termes de triplets de Hitchin $(\mathcal{E}, \Phi, \theta)$. Il consiste simplement à envoyer le couple de fibrés unitaires (\mathcal{E}_i, Φ_i) de rangs respectifs n_1 et n_2 sur leur somme directe et le couple de morphismes θ_i sur leur somme directe. Cette possibilité de plonger l'espace de Hitchin endoscopique dans l'espace de Hitchin, qui vient du fait qu'on peut plonger le groupe endoscopique dans le groupe, est très particulière aux groupes unitaires.

On vérifie aisément que le morphisme ν_* est une immersion fermée. Un des points-clef est qu'on peut aussi caractériser son image de manière agréable.

3.2.2. Le tore. — Nous voulons construire un tore T_S dans le noyau du morphisme $\mathcal{P}_S \longrightarrow \mathcal{Q}_S$. Posons $K_S := pr_{Z'_S \rightarrow S, *}(G_{mZ'_S})^{r=-1}$. La suite exacte $0 \longrightarrow G_{mY'_S} \longrightarrow G_{mY'_{1,S}} \times G_{mY'_{2,S}} \longrightarrow G_{mZ'_S} \longrightarrow 0$ fournit une flèche de bord $K_S \longrightarrow \mathcal{P}_S$ dont la composée avec $\mathcal{P}_S \longrightarrow \mathcal{Q}_S$ est nulle : une section de K_S fournit une donnée de recollement des modules inversibles triviaux sur $Y_{1,S}$ et $Y_{2,S}$ avec structures unitaires triviales. Comme S est hensélien, on peut prolonger le sous-tore maximal de la fibre fermée et noter $T_S \hookrightarrow K_S$ ce tore.

PROPOSITION 3.2 ([30], prop. 3.3.1). — *L'image de l'immersion fermée $\nu_* : \mathcal{N}_S \longrightarrow \mathcal{M}_S$ est le lieu des points fixes de T_S agissant sur \mathcal{M}_S via le morphisme $T_S \longrightarrow \mathcal{P}_S$.*

Démonstration. — Le terme «lieu des points fixes» dans notre contexte catégorique désigne «la sous-catégorie pleine des objets isomorphes à leurs translatés». On va tenter d'expliquer la preuve pour $GL(n)$, en négligeant les structures unitaires. Soit b un point géométrique de S . Pour tout $z \in Z_b$ et $i = 1, 2$, notons $\mathcal{O}_{Y_{i,b}}^{[[z]]}$ le complété formel de l'anneau local de $Y_{i,b}$ en z et $\mathcal{O}_{Y_{i,b}}^{((z))}$ son anneau total des fractions. Alors la donnée de $\mathcal{F} \in \mathcal{M}_S(b)$ est équivalente à la donnée de ses restrictions \mathcal{F}_i^0 , $i = 1, 2$, aux ouverts $Y_{i,b} \setminus Z_b$ et, pour chaque $z \in Z_b$, d'un $\mathcal{O}_{Y_b}^{[[z]]}$ -réseau F_z de $(\mathcal{F}_1^0 \otimes \mathcal{O}_{Y_{1,b}}^{((z))}) \oplus (\mathcal{F}_2^0 \otimes \mathcal{O}_{Y_{2,b}}^{((z))})$. Un tel objet \mathcal{F} est isomorphe à un objet de $\mathcal{N}_S(b)$ si et seulement si les F_z sont des $\mathcal{O}_{Y_{1,b}}^{[[z]]} \times \mathcal{O}_{Y_{2,b}}^{[[z]]}$ -réseaux, *i.e.* si et seulement si les F_z sont de la forme $F_z = F_{1,z} \oplus F_{2,z}$ pour des $\mathcal{O}_{Y_{i,b}}^{[[z]]}$ -réseaux $F_{i,z}$.

Par définition de T , il existe un isomorphisme $(k(b)^\times)^{Z_a} \longrightarrow T_b(k(b))$ de sorte que le terme $t_y \in k(b)^\times$ associé à $y \in Z_a$ envoie, pour tout $z \in Z_b$ se spécialisant en y , le réseau

F_z sur le réseau $t.F_z := (t_y, 1)F_z$. Donc les deux modules \mathcal{F} et $t.\mathcal{F}$ sont isomorphes si et seulement si , pour tout $z \in Z_b$, le réseau F_z est décomposé en $F_z = F_{1,z} \oplus F_{2,z}$. \square

3.2.3. Cohomologie équivariante. — Il y a plusieurs manières de la définir, mais dans le contexte présent, la plus naturelle utilise le langage des champs. On a ici besoin d’une notion relative (sur S). On s’intéresse donc aux complexes $f_{S,*}^T \mathbb{Q}_l$ et $g_{S,*}^T \mathbb{Q}_l$, où $f_S^T : [\mathcal{M}_S/T_S] \rightarrow S$ et $g_S^T : [\mathcal{N}_S/T_S] \rightarrow S$ sont les morphismes structuraux des champs quotients $[\mathcal{M}_S/T_S]$ et $[\mathcal{N}_S/T_S]$. Remarquons que ces complexes ne sont en général pas cohomologiquement bornés.

Notons $\mathcal{E}_{\mathcal{N}/T}$ le T_S -torseur « canonique » au-dessus de $[\mathcal{N}_S/T_S]$, image réciproque du toseur universel par le morphisme $[g] : [\mathcal{N}_S/T_S] \rightarrow B(T_S/S)$ déduit de g par passage au quotient. Si μ est une section sur S du faisceau étale $X^*(T_S)$ des caractères de T_S , alors μ permet de pousser $\mathcal{E}_{\mathcal{N}/T}$ en un \mathbb{G}_m -torseur dont la classe de Chern est un morphisme $\mathbb{Q}_l[\mathcal{N}_S/T_S] \rightarrow \mathbb{Q}_l[\mathcal{N}_S/T_S][2](1)$ de faisceaux étales sur $[\mathcal{N}_S/T_S]$. En appliquant $g_{S,*}^T$, en passant à la cohomologie perverse, et en faisceautisant, on obtient un morphisme

$$X^*(T_S) \otimes_{\mathbb{Z}_S} {}^p\mathcal{H}^\bullet(g_{S,*}^T \mathbb{Q}_l) \rightarrow {}^p\mathcal{H}^\bullet(g_{S,*}^T \mathbb{Q}_l)[2](1)$$

qui munit la somme formelle de faisceaux pervers ${}^p\mathcal{H}^\bullet(g_{S,*}^T \mathbb{Q}_l)$ d’une structure de module gradué sur l’anneau étale gradué $\text{Sym}^\bullet(X^*(T_S) \otimes \mathbb{Q}_l(-1))$ (à condition de graduer ce dernier par l’indice $2\bullet$). De même ${}^p\mathcal{H}^\bullet(f_{S,*}^T \mathbb{Q}_l)$ est muni d’une telle structure.

Comme T_S agit trivialement sur \mathcal{N}_S , on a un isomorphisme canonique

$${}^p\mathcal{H}^\bullet(g_{S,*}^T \mathbb{Q}_l) \xrightarrow{\sim} {}^p\mathcal{H}^\bullet(g_{S,*} \mathbb{Q}_l) \otimes_{\mathbb{Q}_l} \text{Sym}^\bullet(X^*(T_S) \otimes \mathbb{Q}_l(-1))$$

qui montre que la cohomologie ordinaire et la cohomologie équivariante se déterminent l’une l’autre. Du côté de f_S^T et f_S , on a :

PROPOSITION 3.3. — *Grâce à la pureté de ${}^p\mathcal{H}^\bullet(f_{S,*} \mathbb{Q}_l)_\kappa$ (propositions 2.6 et 2.11),*

1. [30, A.1.1] *il existe un isomorphisme $\text{Sym}^\bullet(X^*(T_S) \otimes \mathbb{Q}_l(-1))$ -équivariant*

$${}^p\mathcal{H}^\bullet(f_{S,*}^T \mathbb{Q}_l)_\kappa \xrightarrow{\sim} {}^p\mathcal{H}^\bullet(f_{S,*} \mathbb{Q}_l)_\kappa \otimes \text{Sym}^\bullet(X^*(T_S) \otimes \mathbb{Q}_l(-1)).$$

2. [30, A.1.3] *la flèche de restriction ${}^p\mathcal{H}^\bullet(f_{S,*}^T \mathbb{Q}_l)_\kappa \rightarrow {}^p\mathcal{H}^\bullet(g_{S,*}^T \mathbb{Q}_l)_\kappa$ est injective.*

Le premier point dit que la κ -partie de la cohomologie ordinaire de f_S et celle de la cohomologie équivariante se déterminent l’une l’autre. Le deuxième point est une variante l -adique, champêtre, et perverse d’un théorème d’Atiyah-Borel-Segal. Il montre en particulier que les ${}^p\mathcal{H}^n(f_{S,*}^T \mathbb{Q}_l)_\kappa$ sont eux aussi purs.

3.2.4. « Glissade » et réduction à l’ouvert S_{\natural} . — La proposition suivante permet de ramener l’énoncé du théorème 3.1 à l’énoncé analogue où l’on remplace S par S_{\natural} . On notera $f_{S_{\natural}}, g_{S_{\natural}}$, etc. les restrictions à S_{\natural} des objets correspondants relatifs à S .

PROPOSITION 3.4. — *Notons j l’immersion ouverte $S_{\natural} \hookrightarrow S$. Pour tout $n \in \mathbb{Z}$, il y a des isomorphismes canoniques de faisceaux pervers*

$$j_{!*} {}^p\mathcal{H}^n(f_{S_{\natural},*} \mathbb{Q}_l)_\kappa \xrightarrow{\sim} {}^p\mathcal{H}^n(f_{S,*} \mathbb{Q}_l)_\kappa \quad \text{et} \quad j_{!*} {}^p\mathcal{H}^n(g_{S_{\natural},*} \mathbb{Q}_l)_\kappa \xrightarrow{\sim} {}^p\mathcal{H}^n(g_{S,*} \mathbb{Q}_l)_\kappa.$$

($j_{!*}$ désigne le prolongement intermédiaire de Deligne).

Démonstration. — Grâce au (1) de la proposition précédente, on peut déduire l'énoncé ci-dessus de l'énoncé analogue pour $f_{S_{\mathfrak{h}},*}^T$ et $g_{S_{\mathfrak{h}},*}^T$ par réduction modulo l'idéal d'augmentation de l'anneau $\mathrm{Sym}^\bullet(X^*(T_S) \otimes \mathbb{Q}_l(-1))$. D'après [1, 1.4.25], un faisceau pervers K sur S est isomorphe au prolongement intermédiaire de sa restriction $j_{!*}j^*K$ si et seulement si il n'a ni quotient ni sous-objet supporté par $S \setminus S_{\mathfrak{h}}$. Sur l'hensélisé strict de S , les faisceaux pervers ${}^p\mathcal{H}^n(f_{S_{\mathfrak{h}}}^T \mathbb{Q}_l)_\kappa$ et ${}^p\mathcal{H}^n(g_{S_{\mathfrak{h}}}^T \mathbb{Q}_l)_\kappa$ sont semi-simples, puisqu'ils sont purs. Comme le premier s'injecte dans le second par le (2) de la proposition précédente, il suffit donc de prouver que ${}^p\mathcal{H}^n(g_{S_{\mathfrak{h}}}^T \mathbb{Q}_l)_\kappa$ n'a pas de sous-objet supporté par $S \setminus S_{\mathfrak{h}}$.

Ici intervient l'argument de « glissade » ou de « déformation ». En termes imprécis, il consiste à remarquer que pour $b = (b_1, b_2)$ un point géométrique de S , la fibre de Hitchin \mathcal{N}_b ne dépend que des courbes Y_{b_1} et Y_{b_2} , et pas de la manière dont elles s'intersectent. On peut donc glisser, « à cohomologie constante », du point $b \in S$ où l'intersection peut être compliquée à un point de $S_{\mathfrak{h}}$, où elle est le plus simple possible. En formalisant ceci [30, 3.9.1], on en déduit que le support de tout sous-objet de ${}^p\mathcal{H}^n(g_{S_{\mathfrak{h}}}^T \mathbb{Q}_l)_\kappa$ rencontre $S_{\mathfrak{h}}$. \square

3.2.5. Stratégie sur $S_{\mathfrak{h}}$. — Rappelons que le tore T_S est par définition contenu dans le S -schéma en groupes commutatifs K_S de 3.2.2. Au-dessus de $S_{\mathfrak{h}}$, $K_{S_{\mathfrak{h}}} := K_S \times_S S_{\mathfrak{h}}$ est lui-même un tore, de dimension r , qui contient en général strictement $T_{S_{\mathfrak{h}}}$. De même que dans le (1) de la proposition 3.3, on récupère la κ -partie de la cohomologie ${}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*} \mathbb{Q}_l)_\kappa$ par réduction de celle de la cohomologie équivariante ${}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa$ modulo l'idéal d'augmentation de l'algèbre $\mathrm{Sym}^\bullet(X^*(K_{S_{\mathfrak{h}}}) \otimes \mathbb{Q}_l(-1))$. Ainsi pour prouver l'énoncé du théorème 3.1, il suffit de prouver l'énoncé analogue où l'on remplace $f_{S,*}$ par $f_{S_{\mathfrak{h}},*}^K$ et $g_{S,*}$ par $g_{S_{\mathfrak{h}},*}^K$ (et bien sûr $L_{Z'/Z/S}$ par $L_{Z'_{\mathfrak{h}}/Z_{\mathfrak{h}}/S_{\mathfrak{h}}}$).

De même que dans le (2) de la proposition 3.3, la flèche restriction ${}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa \longrightarrow {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa$ est injective, et le problème est donc de décrire son image. En fait, la preuve de ce point (2) fournit un conoyau au morphisme de restriction, à savoir le morphisme cobord ${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa \longrightarrow {}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*}^K ([j]! \mathbb{Q}_l))_\kappa[1]$ où $[j]$ désigne l'immersion ouverte du complémentaire de $[\mathcal{N}_{S_{\mathfrak{h}}}/K_{S_{\mathfrak{h}}}]$ dans $[\mathcal{M}_{S_{\mathfrak{h}}}/K_{S_{\mathfrak{h}}}]$. Pour étudier ce conoyau, Laumon et Ngô introduisent en [30, 3.5] un certain $K_{S_{\mathfrak{h}}}$ -torseur \mathcal{E}_{12} sur $\mathcal{N}_{S_{\mathfrak{h}}}$, qui apparaissait déjà dans [31]. Définissons ce torseur ; soit $U_{\mathfrak{h}}$ un $S_{\mathfrak{h}}$ -schéma et $((\mathcal{F}_1, \iota_1), (\mathcal{F}_2, \iota_2))$ un objet de $\mathcal{N}_{S_{\mathfrak{h}}}(U_{\mathfrak{h}})$. Comme $Z_{U_{\mathfrak{h}}}$ est étale sur $U_{\mathfrak{h}}$, les restrictions $\mathcal{F}_i|_{Z_{U_{\mathfrak{h}}}}$ sont des $\mathcal{O}_{Z_{U_{\mathfrak{h}}}}$ -modules inversibles et le produit $\mathcal{E} := \mathcal{F}_1|_{Z_{U_{\mathfrak{h}}}} \otimes \mathcal{F}_2|_{Z_{U_{\mathfrak{h}}}}^{\otimes -1}$ est muni d'un isomorphisme $\iota_1 \otimes \iota_2^{-1} : \tau^* \mathcal{E} \xrightarrow{\sim} \mathcal{E}^{\otimes -1}$. Un tel $\mathcal{O}_{Z_{U_{\mathfrak{h}}}}$ -module inversible \mathcal{E} définit un $K_{U_{\mathfrak{h}}}$ -torseur. On obtient de la sorte un morphisme $\mathcal{N}_{S_{\mathfrak{h}}} \longrightarrow B(K_{S_{\mathfrak{h}}}/S_{\mathfrak{h}})$. En tirant en arrière le torseur universel, on obtient le $K_{S_{\mathfrak{h}}}$ -torseur \mathcal{E}_{12} sur $\mathcal{N}_{S_{\mathfrak{h}}}$ annoncé.

À l'aide de ce torseur \mathcal{E}_{12} , nous allons construire une flèche injective

$$(10) \quad L_{Z'_{\mathfrak{h}}/Z_{\mathfrak{h}}/S_{\mathfrak{h}}} \otimes_{\mathbb{Z}_{S_{\mathfrak{h}}}} {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa[-2r](-r) \longrightarrow {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa,$$

puis il faudra expliquer pourquoi l'image de cette flèche coïncide avec l'image de la restriction ${}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa \longrightarrow {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa$, ce qui achèvera la preuve du théorème.

3.2.6. *Construction de la flèche* (10). — Le formalisme des classes de Chern (*cf.* 3.2.3) appliqué à \mathcal{E}_{12} munit ${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)$ d'une action e_{12} de l'anneau gradué $\mathrm{Sym}^\bullet(X^*(K_{S_{\mathfrak{h}}}) \otimes \mathbb{Q}_l(-1))$. En quotientant \mathcal{E}_{12} par l'action de $K_{S_{\mathfrak{h}}}$ on obtient un $K_{S_{\mathfrak{h}}}$ -torseur $\mathcal{E}_{12}/K_{S_{\mathfrak{h}}}$, au-dessus de $[\mathcal{N}_{S_{\mathfrak{h}}}/K_{S_{\mathfrak{h}}}]$ (différent du $K_{S_{\mathfrak{h}}}$ torseur « canonique » $\mathcal{E}_{[N/K]}$), qui munit donc ${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)$ d'une action e_{12}^K de $\mathrm{Sym}^\bullet(X^*(K_{S_{\mathfrak{h}}}) \otimes \mathbb{Q}_l(-1))$. Le lien entre cette action et l'action canonique est le suivant : l'isomorphisme canonique

$${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l) \xrightarrow{\sim} {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l) \otimes_{\mathbb{Q}_l} \mathrm{Sym}^\bullet(X^*(K_{S_{\mathfrak{h}}}) \otimes \mathbb{Q}_l(-1))$$

est $\mathrm{Sym}^\bullet(X^*(K_{S_{\mathfrak{h}}}) \otimes \mathbb{Q}_l(-1))$ -équivariant pour l'action e_{12}^K sur le terme de gauche et l'action $(e_{12} \otimes \mathrm{mult}) \circ \Delta$ sur le terme de droite, en notant Δ la comultiplication $\mathrm{Sym}^\bullet \longrightarrow \mathrm{Sym}^\bullet \otimes \mathrm{Sym}^\bullet$, *cf.* [30, A.2.1]. On en déduit que ${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)$ est sans torsion pour l'action de e_{12}^K .

L'action e_{12}^K fournit en particulier un morphisme

$$\mathrm{Sym}^r(X^*(K_{S_{\mathfrak{h}}})) \otimes_{\mathbb{Z}_{S_{\mathfrak{h}}}} {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l) \longrightarrow {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)[2r](r).$$

La flèche (10) que l'on veut définir sera la κ -partie de la composée du morphisme ci-dessus avec le plongement

$$(11) \quad L_{Z'_{\mathfrak{h}}/Z_{\mathfrak{h}}/S_{\mathfrak{h}}} \longrightarrow \mathrm{Sym}^r(X^*(K_{S_{\mathfrak{h}}}))$$

que nous définissons maintenant.

Rappelons que $K_{S_{\mathfrak{h}}}$ est la partie « τ -impair» de $pr_{Z'_{\mathfrak{h}} \rightarrow S_{\mathfrak{h}},*} \mathbb{G}_m$, et introduisons temporairement sa partie τ -paire $K'_{S_{\mathfrak{h}}}$. La structure d'anneau de $X_*(\mathbb{G}_m) = \mathcal{E}nd_{Z'_{\mathfrak{h}}}(\mathbb{G}_m)$ induit une action $X_*(K_{S_{\mathfrak{h}}}) \otimes X_*(K'_{S_{\mathfrak{h}}}) \longrightarrow X_*(K_{S_{\mathfrak{h}}})$ de la partie paire sur la partie impaire. Par dualité et puissance tensorielle, on a une action $\mathrm{Tens}^r X^*(K_{S_{\mathfrak{h}}}) \otimes \mathrm{Tens}^r X_*(K'_{S_{\mathfrak{h}}}) \longrightarrow \mathrm{Tens}^r X^*(K_{S_{\mathfrak{h}}})$ qui envoie $\bigwedge^r X^*(K_{S_{\mathfrak{h}}}) \otimes \bigwedge^r X_*(K'_{S_{\mathfrak{h}}})$ dans $\mathrm{Sym}^r(X^*(K_{S_{\mathfrak{h}}}))$ (on a plongé \bigwedge^r et Sym^r dans Tens^r respectivement par les morphismes « trace alternée » et « trace » sous le groupe symétrique \mathfrak{S}_r). Maintenant, l'isomorphisme canonique $\mathbb{Z} \xrightarrow{\sim} X_*(\mathbb{G}_m)$ d'anneaux étales sur $Z'_{S_{\mathfrak{h}}}$ induit, par construction, un isomorphisme $L_{Z'_{\mathfrak{h}}/Z_{\mathfrak{h}}/S_{\mathfrak{h}}} \xrightarrow{\sim} \bigwedge^r X^*(K_{S_{\mathfrak{h}}}) \otimes \bigwedge^r X_*(K'_{S_{\mathfrak{h}}})$. Ceci achève de définir la flèche (11) et donc la flèche (10). Cette flèche (10) est injective puisque l'action e_{12}^K est sans torsion.

3.2.7. *Fin de la preuve.* — Il s'agit de prouver que la flèche injective (10) et la flèche injective de restriction ${}^p\mathcal{H}^\bullet(f_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa \longrightarrow {}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa$ ont la même image dans ${}^p\mathcal{H}^\bullet(g_{S_{\mathfrak{h}},*}^K \mathbb{Q}_l)_\kappa$.

Supposons un instant que $r = 1$, de sorte que $Z_{\mathfrak{h}} \simeq S_{\mathfrak{h}}$, $K_{S_{\mathfrak{h}}} \simeq U_{Z'_{\mathfrak{h}}/S_{\mathfrak{h}}}(1)$ et $L_{Z'_{\mathfrak{h}}/Z_{\mathfrak{h}}/S_{\mathfrak{h}}} \simeq X^*(K_{\mathfrak{h}})$. Posons $S'_{\mathfrak{h}} := Z'_{\mathfrak{h}}$ et changeons tout le monde de base. Le choix d'une section $S'_{\mathfrak{h}} \xrightarrow{z'} Z'_{\mathfrak{h}} \times_{S_{\mathfrak{h}}} S'_{\mathfrak{h}}$ induit un isomorphisme $K_{S'_{\mathfrak{h}}} \xrightarrow{\sim} \mathbb{G}_m$, de sorte que $\mathcal{E}'_{12} := \mathcal{E}_{12} \times_{S_{\mathfrak{h}}} S'_{\mathfrak{h}}$ est simplement un \mathbb{G}_m -torseur au-dessus de $\mathcal{N}_{S'_{\mathfrak{h}}}$. Notons \mathbb{P}'_{12} la compactification projective relative de \mathcal{E}'_{12} , qui est donc un fibré en droites projectives au-dessus de $\mathcal{N}_{S'_{\mathfrak{h}}}$. Si $U'_{\mathfrak{h}}$ est

un $S'_\mathfrak{h}$ -schéma, une section $U'_\mathfrak{h} \longrightarrow \mathbb{P}_{12}$ est la donnée d'une paire $((\mathcal{F}_1, t_1), (\mathcal{F}_2, t_2))$ dans $\mathcal{N}_{S'_\mathfrak{h}}(U'_\mathfrak{h})$ plus un facteur direct de rang 1 de $\mathcal{F}_{1,z'} \oplus \mathcal{F}_{2,z'}$. Ce facteur direct détermine un sous- $\mathcal{O}_{Y_{U'_\mathfrak{h}}}$ -module sans torsion de rang 1 $\mathcal{F} \hookrightarrow \mathcal{F}_1 \oplus \mathcal{F}_2$. L'amalgamé de \mathcal{F} et $\tau^*(\mathcal{F}^\vee)$ au-dessus de $\mathcal{F}_1 \oplus \mathcal{F}_2$ est muni d'une structure unitaire évidente et définit un objet de $\mathcal{M}_{S'_\mathfrak{h}}(U'_\mathfrak{h})$. On obtient ainsi un morphisme $\mathbb{P}'_{12} \longrightarrow \mathcal{M}_{S'_\mathfrak{h}}$ qui envoie isomorphiquement \mathcal{E}'_{12} sur le complémentaire de $\mathcal{N}_{S'_\mathfrak{h}}$ et qui identifie les sections 0 et ∞ de \mathbb{P}'_{12} . L'image de l'immersion $\mathcal{N}_{S'_\mathfrak{h}} \hookrightarrow \mathcal{M}_{S'_\mathfrak{h}}$ est justement cette section double. On a donc une prise sur le complémentaire de $\mathcal{N}_{S'_\mathfrak{h}}$ dans $\mathcal{M}_{S'_\mathfrak{h}}$. La cohomologie de cette situation géométrique assez simple est étudiée dans l'appendice A.2 de [30]. Après descente subtile de $S'_\mathfrak{h}$ à $S_\mathfrak{h}$, on en déduit le résultat cherché dans ce cas $r = 1$.

Le jeu est alors de se ramener à cette situation $r = 1$. L'idée est de faire intersecter les courbes Y_1 et Y_2 «petit à petit» en considérant une suite de courbes intermédiaires entre $Y_1 \sqcup Y_2$ et Y . La réalisation formelle de ce programme est assez technique et occupe les six pages de la section 3.6. de [30].

3.3. Le lemme fondamental

Comme on l'a mentionné après la conjecture 2.12, on obtient en prenant les traces de Frobenius des deux côtés de l'isomorphisme du théorème 3.1 une égalité entre une intégrale orbitale stable sur $H = U_{X'|X}(n_1) \times U_{X'|X}(n_2)$ et une κ -intégrale orbitale sur $G = U_{X'|X}(n)$. Cela fournit un transfert explicite des fonctions $\mathbf{1}_D$ sur G vers son groupe endoscopique H et il est possible que l'on puisse en déduire le lemme fondamental par des arguments de formule des traces et d'approximation du local par le global. Toutefois, la stratégie suivie par Laumon et Ngô est de nature plus géométrique.

3.3.1. Conséquence numérique du théorème 3.1. — On fixe $a \in \mathcal{A}_D^{ell}(k)$. Le théorème 2.4 montre comment le quotient $[\mathcal{M}_a/\mathcal{P}_a]$ se décompose en un produit d'analogues locaux indexés par les points fermés de X . Dans le cas unitaire, on peut raffiner cette décomposition en un produit indexé par les points fermés de la courbe spectrale Y_a . Notons en effet A_y l'anneau semi-local avec involution τ complété de Y'_a le long de l'image réciproque de $y \in |Y_a|$ et introduisons les classifiants $\mathcal{M}_{a,y}$, resp. $\mathcal{P}_{a,y}$, des A_y -modules sans torsion de rang 1 «unitaires», resp. inversibles «unitaires». Lorsque y est un point lisse, le quotient de $\mathcal{M}_{a,y}$ par $\mathcal{P}_{a,y}$ est bien sûr trivial. La même preuve que pour le théorème 2.4 fournit alors un isomorphisme de k -champs $[\mathcal{M}_a/\mathcal{P}_a] \xrightarrow{\sim} \prod_{y \in Y^{sing}} [\mathcal{M}_{a,y}/\mathcal{P}_{a,y}]$. Lorsque $a = (a_1, a_2)$ et $y \in Y_{a_1}$ ou $y \in Y_{a_2}$, on définit de même $A_{1,y}$ et $A_{2,y}$, puis les classifiants $\mathcal{N}_{a,y}$ et $\mathcal{Q}_{a,y}$, et on a une décomposition $[\mathcal{N}_a/\mathcal{Q}_a] \xrightarrow{\sim} \prod_{y \in Y^{sing}} [\mathcal{N}_{a,y}/\mathcal{Q}_{a,y}]$.

De même que dans la discussion qui suit le théorème 2.4, on a besoin de sections $[\mathcal{M}_{a,y}/\mathcal{P}_{a,y}] \longrightarrow [\mathcal{M}_a/\mathcal{P}_a]$ pour définir les accouplements $\langle \text{inv}(m_y), \kappa \rangle$ pour m_y un objet de $[\mathcal{M}_{a,y}/\mathcal{P}_{a,y}]$. Ces sections s'obtiennent en introduisant des versions génériquement rigidifiées $\mathcal{M}_{a,y}^\bullet$ et $\mathcal{P}_{a,y}^\bullet$ par un *point-base bien choisi*. Laumon et Ngô utilisent un

point de Kostant du groupe endoscopique qui se décrit concrètement comme $(\mathcal{K}, \iota_{\mathcal{K}}) := \nu_*((\mathcal{K}_1, \iota_{\mathcal{K}_1})(\mathcal{K}_2, \iota_{\mathcal{K}_2}))$ où, pour $\alpha = 1, 2$, on prend $\mathcal{K}_\alpha := pr_{Y'_{a_\alpha} \rightarrow X'}^* \mathcal{O}_{X'}((n-1)D)$ qui est une racine carrée du dualisant relatif $\omega_{Y'_{a_\alpha}/X'}$ et qui provient de X , donc est muni d'une structure unitaire ι_α (de descente).

COROLLAIRE 3.5 ([30], th. 3.10.6). — *Sous les mêmes hypothèses que le théorème 3.1, notons Z^{inerte} l'ensemble des points fermés de Z qui sont inertes dans Z' et, pour chaque tel point, r_z la longueur de la $k(z)$ -algèbre artinienne $\mathcal{O}_{Z,z}$. Alors on a l'égalité*

$$\begin{aligned} & \prod_{z \in Z^{\text{inerte}}} \left(|\mathcal{P}_{a,z}^0(k)| \sum_{m_z \in [\mathcal{M}_{a,z}/\mathcal{P}_{a,z}](k)/\sim} \frac{\langle \text{inv}(m_z), \kappa \rangle}{\text{Aut}(m_z)} \right) \\ &= \prod_{z \in Z^{\text{inerte}}} \left((-|k(z)|)^{r_z} |\mathcal{Q}_{a,z}^0(k)| \sum_{n_z \in [\mathcal{N}_{a,z}/\mathcal{Q}_{a,z}](k)/\sim} \frac{1}{\text{Aut}(n_z)} \right) \end{aligned}$$

Démonstration. — En prenant la trace de Frobenius du théorème 3.1, on obtient la formule

$$|\mathcal{P}_a^0(k)| \prod_{y \in Y^{\text{sing}}} \left(\sum_{m_y} \frac{\langle \text{inv}(m_y), \kappa \rangle}{\text{Aut}(m_y)} \right) = \varepsilon_{Z'/Z/S} q^r |\mathcal{Q}_a^0(k)| \prod_{y \in Y^{\text{sing}}} \left(\sum_{n_y} \frac{1}{\text{Aut}(n_y)} \right)$$

où $\varepsilon_{Z'/Z/S}$ est le signe de Frobenius qui détermine $L_{Z'/Z/S}$. Le lemme [30, 3.4.1] montre que ce signe est la parité de $\sum_{z \in Z^{\text{inerte}}} r_z$. On a aussi par définition $q^r = \prod_{y \in Y^{\text{sing}}} |k(y)|^{r_y}$ si on convient de poser $r_y = 0$ pour $y \in Y^{\text{sing}} \setminus Z$. Les termes $|\mathcal{P}_a^0(k)|$ et $|\mathcal{Q}_a^0(k)|$ ne se décomposent pas en produits de termes locaux, mais en explicitant le quotient $[\mathcal{Q}_a^0/\mathcal{P}_a^0]$, on montre que leur quotient se décompose en un tel produit.

Il reste alors à montrer que, pour $y \in Y^{\text{sing}} \setminus Z^{\text{inerte}}$, on a l'égalité

$$|\mathcal{P}_{a,y}^0(k)| \sum_{m_y} \frac{\langle \text{inv}(m_y), \kappa \rangle}{\text{Aut}(m_y)} = |k(y)| |\mathcal{Q}_{a,y}^0(k)| \sum_{n_y} \frac{1}{\text{Aut}(n_y)}.$$

Pour y hors de Z , c'est immédiat. Pour $y \in Z \setminus Z^{\text{inerte}}$, il faut faire un calcul, qui rappelle la descente des intégrales orbitales. \square

3.3.2. *Plongement d'une situation locale dans une situation globale.* — On fixe ici la courbe X , son revêtement X' que l'on suppose décomposé en un point x_∞ , et un diviseur de degré assez grand». On a alors la surface réglée $\Sigma_D^0 \rightarrow X$.

On se donne ensuite un contexte local comme dans 1.8.1. On a donc $F = \mathbb{F}_q[[x]]$, F' son extension quadratique non-ramifiée, des extensions séparables $(F_i)_{i \in I}$ de F disjointes de F' et des éléments entiers $\gamma_i \in F'_i$ tels que $\gamma_i + \tau(\gamma_i) = 0$. On se donne aussi une partition $I = I_1 \sqcup I_2$ et on pose $A_{I_\alpha} = \prod_{i \in I_\alpha} \mathcal{O}_{F'}[\gamma_i]$ pour $\alpha = 1, 2$.

PROPOSITION 3.6 ([30], prop. 4.6.1). — *Il existe des entiers non-nuls $N = N_1 + N_2$, une caractéristique $(a_1, a_2) \in \mathcal{A}_H(k)$ pour le groupe endoscopique $H := U_{X'|X}(N_1) \times U_{X'|X}(N_2)$ de $G = U_{X'|X}(N)$ telle que les courbes spectrales associées Y_{a_1} et Y_{a_2} soient*

géométriquement irréductibles, et un point rationnel z_0 de l'intersection Z de Y_{a_1} et Y_{a_2} dans la surface réglée Σ_D^0 , ayant les propriétés suivantes :

- en tout point fermé $z \in Z \setminus z_0$, les courbes spectrales se coupent transversalement.
- z_0 est inerte dans Z' , de préimage un point fermé noté $z'_0 \in Z'$. Alors x_0 est aussi inerte dans X' , et on note $x'_0 \in X'$ le point fermé au-dessus.
- Il existe des isomorphismes compatibles aux involutions et compatibles entre eux
 - $\widehat{\mathcal{O}}_{X',x'_0} \xrightarrow{\sim} \mathcal{O}_{F'}$
 - $A_{\alpha,z_0} = \widehat{\mathcal{O}}_{Y'_{a_\alpha},z'_0} \xrightarrow{\sim} A_{I_\alpha} = \prod_{i \in I_\alpha} \mathcal{O}_{F'}[\gamma_i]$ pour $\alpha = 1, 2$.

La preuve de cette proposition utilise un lemme d'approximation local-global [30, 4.2] et un théorème à la Bertini sur les corps finis, dû à Poonen [34].

3.3.3. *Fin de la preuve du lemme fondamental.* — Il faut étudier l'égalité donnée par le corollaire 3.5 dans la situation de la proposition ci-dessus. Une description explicite [30, 4.6] des catégories quotients $[\mathcal{M}_{a,z}/\mathcal{P}_{a,z}]$ montre que les termes indexés par $Z^{\text{inerte}} \setminus z_0$ des deux côtés sont égaux. Il subsiste donc une égalité des deux termes indexés par z_0 . La même description montre aussi les égalités suivantes (tout a été fait pour) :

$$|\mathcal{Q}_{a,z_0}^0(k)| \sum_{n_{z_0} \in [\mathcal{N}_{a,z_0}/\mathcal{Q}_{a,z_0}](k)/\sim} \frac{1}{\text{Aut}(n_{z_0})} = SO_{\gamma_I}^H$$

et

$$|\mathcal{P}_{a,z_0}^0(k)| \sum_{m_{z_0} \in [\mathcal{M}_{a,z_0}/\mathcal{P}_{a,z_0}](k)/\sim} \frac{\langle \text{inv}(m_{z_0}), \kappa \rangle}{\text{Aut}(m_{z_0})} = O_{\gamma_I}^\kappa.$$

Le lemme fondamental (théorème 1.6) en découle.

RÉFÉRENCES

- [1] J. Bernstein A. Beilinson and P. Deligne. *Analyse et topologie sur les espaces singuliers (I: Faisceaux Pervers)*. Astérisque 100.
- [2] J. Arthur. A stable trace formula II : Global descent. *Invent. Math.*, 143:157–220, 2001.
- [3] J. Arthur. A stable trace formula I : General expansions. *J. Inst. Math. Jussieu*, 1(2):175–277, 2002.
- [4] J. Arthur. A stable trace formula III : Proof of the main theorems. *Ann. of Math.*, 158, 2003.
- [5] A. Beauville, M.S. Narasimhan, et S. Ramanan. Spectral curve and the generalized Theta divisor. *J. reine angew. Math.*, 398:169–179, 1989.
- [6] I. Biswas et S. Ramanan. Infinitesimal study of Hitchin pairs. *J. London Math. Soc.*, 49:219–231, 1994.
- [7] A. Borel. Automorphic L -functions. Proc. Symp. Pure Math. 33(2), pages 27–63, 1979.

- [8] A. Borel and H. Jacquet. Automorphic forms and automorphic representations. Proc. Symp. Pure Math. 33(1), pages 189–202, 1979.
- [9] L. Breen. *On the classification of 2-gerbes and 2-stacks*, Astérisque 225. 1994.
- [10] G. Faltings. Stable G -bundles and projective connections. *J. Alg. Geom.*, 2:507–568, 1993.
- [11] M. Goresky, R. Kottwitz, et R. Macpherson. Homology of affine Springer fibers in the unramified case. *Duke Math. J.*, 121:509–561, 2004.
- [12] T.C. Hales. A simple definition of transfer factors for unramified groups. *Contemporary Math*, 145:109–134, 1993.
- [13] T.C. Hales. On the fundamental lemma for standard endoscopy : reduction to unit elements. *Canad. J. Math.*, 47:974–994, 1995.
- [14] T.C. Hales. A statement of the fundamental lemma. *ArXiv:math.RT/0312227*, 2003.
- [15] N. Hitchin. Stable bundles and integrable connections. *Duke Math. J.*, 54:91–11, 1987.
- [16] D. Kazhdan et G. Lusztig. Fixed points varieties on affine flag manifolds. *Israel. J. Math.*, 62:129–168, 1988.
- [17] R.E. Kottwitz. Stable trace formula : Cuspidal tempered terms. *Duke Math. J.*, 51:611–650, 1984.
- [18] R.E. Kottwitz. Stable trace formula : Elliptic singular terms. *Math. Ann.*, 275:365–399, 1986.
- [19] R.E. Kottwitz. Tamagawa numbers. *Ann. of Maths*, 127:629–646, 1988.
- [20] R.E. Kottwitz. Transfer factors for Lie algebras. *Represent. Theory*, 3:127–138, 1999.
- [21] R.E. Kottwitz et D. Shelstad. *Foundations of twisted endoscopy*. Astérisque 225, 1999.
- [22] J.-P. Labesse. Stable twisted trace formula : elliptic terms. *J. Inst. Math. Jussieu*, 3(4).
- [23] J.-P. Labesse. *Cohomologie, stabilisation et changement de base*, Astérisque 257, 1999.
- [24] J.-P. Labesse et R.P. Langlands. L -indistinguishability for $SL(2)$. *Canad. J. Math*, 31, 1979.
- [25] R.P. Langlands. *Les débuts d’une formule des traces stable*. Publ. Math. Univ. Paris 7, 1979.
- [26] R.P. Langlands et D. Shelstad. On the definition of transfer factors. *Math. Ann.*, 278:219–271, 1987.
- [27] R.P. Langlands et D. Shelstad. Descent for transfer factors. *The Grothendieck Festschrift*, vol II, 485–563, Prog. Math. 87, 1990.
- [28] G. Laumon. Sur le lemme fondamental pour les groupes unitaires. *math AG/0212245*, 2002.

- [29] G. Laumon et L. Moret-Bailly. *Champs algébriques*. Ergebnisse der Mathematik. Springer, 2000.
- [30] G. Laumon et B.C. Ngô. Le lemme fondamental pour les groupes unitaires. *math.AG/0404454*, 2004.
- [31] G. Laumon et M. Rapoport. A geometric approach to the fundamental lemma for unitary groups. *math AG/9711021*, 1997.
- [32] B.C. Ngô. Sur l'endoscopie des groupes orthogonaux et symplectiques et leur fibration de Hitchin. *En préparation*.
- [33] B.C. Ngô. Fibration de Hitchin et endoscopie. *math AG/0406599*, 2004.
- [34] B. Poonen. Bertini theorems over finite fields. *math AG/0204002*, 2002.
- [35] J.-L. Waldspurger. Le lemme fondamental implique le transfert. *Compositio Math.*, 105:153–236, 1997.
- [36] J.-L. Waldspurger. *Intégrales orbitales nilpotentes et endoscopie pour les groupes non ramifiés*, Astérisque 269, 2001.
- [37] J.-L. Waldspurger. Endoscopie et changement de caractéristique. Prépublication disponible sur la page <http://www.math.jussieu.fr/waldspur>, 2004.
- [38] J.-L. Waldspurger. Sur les intégrales orbitales tordues des groupes linéaires. Un lemme fondamental. *Canad. J. Math.* 4:852–896, 1991.

Jean-François DAT

Université Paris-Nord

Département de Mathématiques

Avenue J.-B. Clément

F-93430 Villetaneuse

E-mail : dat@math.univ-paris13.fr

FORMES QUADRATIQUES ET CYCLES ALGÈBRIQUES
[d'après Rost, Voevodsky, Vishik, Karpenko...]

par **Bruno KAHN**

Table des matières

Introduction.....	1
Partie I. Formes quadratiques.....	3
1. Définitions et notations.....	3
2. La théorie de Witt.....	4
3. Le théorème de Springer.....	6
4. La théorie de Pfister : puissances de I	7
5. Corps de fonctions de quadriques.....	8
6. La théorie de Knebusch : déploiement générique.....	9
7. Équivalence birationnelle stable.....	14
8. Quatre résultats fondamentaux.....	15
9. Trois applications.....	18
Partie II. Cycles algébriques.....	20
10. Formes quadratiques et motifs : résultats de base.....	20
11. Formes quadratiques et motifs : théories de Rost et de Vishik.....	31
12. Quelques démonstrations.....	44
Références.....	46

INTRODUCTION

La théorie algébrique des formes quadratiques (par opposition à la théorie arithmétique dans la lignée de Legendre, Gauss, Hermite, Minkowski, Hasse...) est l'étude des formes quadratiques sur un corps quelconque : l'article fondateur est celui de Witt ([69], 1937). Elle a connu un développement par à-coups, impulsé par les idées nouvelles et fondamentales introduites successivement par un petit nombre de mathématiciens. Il s'agit maintenant d'une théorie foisonnante, en pleine clarification, où cependant les méthodes les plus sophistiquées voisinent toujours de manière un

peu mystérieuse avec les plus élémentaires, donnant parfois des démonstrations très différentes d'un même théorème.

Qu'est-ce que la théorie des formes quadratiques sur un corps F ? Sous un angle, il s'agit de l'étude des polynômes homogènes de degré 2. Le point de vue de Witt était qu'on peut munir ces polynômes d'une somme et d'un produit, en considérant la somme directe et le produit tensoriel des espaces vectoriels sous-jacents : on obtient ainsi l'anneau de Witt (ou plus exactement de Witt-Grothendieck) de F . La tentation de placer ainsi la théorie dans le contexte plus général de l'étude des formes de degré d est trompeuse : pour $d \geq 3$, la situation devient trop rigide et l'anneau obtenu est essentiellement inintéressant (*cf.* [20]⁽¹⁾).

D'un autre point de vue, la quadrique projective des zéros d'une forme quadratique q est un cas particulier de variété projective homogène (ici, sous l'action du groupe $SO(q)$) ; en particulier c'est une variété rationnelle. L'étude de l'isotropie de cette quadrique sur les extensions de F est centrale dans la théorie. Quoique d'autres variétés projectives homogènes soient naturellement attachées à des structures algébriques (par exemple les variétés de Severi-Brauer), il n'y a pas d'autre famille de telles variétés qui soit représentée par des structures algébriques qu'on puisse additionner et multiplier comme les formes quadratiques. Ceci donne sa richesse et sa spécificité à la théorie, qui se trouve naturellement au confluent de deux mondes (formes et variétés de drapeaux généralisées).

Pendant longtemps, cette théorie a pu passer pour une curiosité à mi-chemin entre l'« arithmétique des corps » et une géométrie algébrique relativement élémentaire. Elle est en train de trouver sa vraie place : d'une part elle intervient de manière essentielle dans la démonstration de la conjecture de Milnor par Voevodsky ([66], voir [30] pour un rapport dans ce Séminaire). D'autre part, elle est intrinsèquement présente dans la théorie homotopique des schémas de Morel-Voevodsky [49] : ce fait, anticipé par Rost, est illustré par le théorème fondamental de Morel (*cf.* [48, rem. 6.4.2]) selon lequel l'anneau des endomorphismes de l'objet unité de la catégorie homotopique stable des F -schémas n'est autre que l'anneau de Witt-Grothendieck de F , lorsque F est parfait de caractéristique $\neq 2$.

Par ailleurs, la théorie des formes quadratiques sur un schéma, qui était longtemps restée embryonnaire après le travail de fondements de Knebusch dans les années 1970, connaît depuis une dizaine d'années un développement spectaculaire grâce aux résultats de Balmer, Walter et d'autres, obtenus à l'aide des catégories triangulées à dualité [7, 8]. Ce n'est pas le lieu d'en parler ici, mais cela confirme que la théorie arrive à maturité.

Dans cet exposé, j'ai voulu d'abord offrir un survol de la théorie telle qu'elle s'est développée jusqu'au milieu des années 1990 : mal connue des non spécialistes, elle présente une élégance et une originalité qui, j'espère, séduiront certains lecteurs comme

⁽¹⁾Par exemple, d'après un résultat remontant à Camille Jordan, si F est de caractéristique zéro, le groupe des automorphismes d'une telle forme est fini dès que l'hypersurface projective correspondante est lisse ; voir [55] pour une démonstration moderne et un peu plus générale.

elles m'ont séduit moi-même. (Ne pouvant être exhaustif dans un tel exposé, je me suis néanmoins limité aux aspects directement liés aux derniers développements : ainsi, les travaux d'Elman-Lam ou ceux sur les corps ordonnés ne sont pas mentionnés.) Ensuite expliquer les développements de nature motivique (ou pour être plus terre à terre, faisant intervenir les correspondances algébriques) qui la révolutionnent depuis un peu moins de 10 ans, et les illustrer par des applications frappantes qui semblaient hors de portée avant l'apparition de ces méthodes.

J'ai essayé de donner un traitement aussi géométrique que possible de la théorie ; en particulier, chaque fois que cela a été possible je me suis efforcé de la placer dans le contexte plus général des variétés projectives homogènes (voir ci-dessus).

Ce rapport étant déjà excessivement long, j'ai été conduit à faire des choix cornéliens. En particulier, j'ai renoncé à exposer la partie de la théorie faisant intervenir les invariants supérieurs des formes quadratiques, c'est-à-dire la seconde conjecture de Milnor ([50] ; cf. [30, (2) à la fin de l'introduction]). On n'y trouvera que de brèves allusions ici ou là, voir notamment remarques 6.9 et 11.32. Un avantage de ce choix est que les résultats expliqués ici sont tous de démonstrations « élémentaires », n'utilisant pas la théorie homotopique des schémas.

Je remercie Yves André, Hélène Esnault, Detlev Hoffmann et Alexander Vishik pour leur aide dans la préparation de ce texte, ainsi que Nikita Karpenko, Jean-Pierre Serre et Tamás Szamuely pour diverses remarques le concernant. Je remercie également le CIMAT de Guanajuato, où il a été en partie conçu, pour son hospitalité et pour les excellentes conditions de travail dont j'ai bénéficié.

On travaille sur un corps F de caractéristique $\neq 2$. On note \bar{F} une clôture séparable de F .

PARTIE I

FORMES QUADRATIQUES

1. DÉFINITIONS ET NOTATIONS

Une *forme quadratique* sur F est un espace vectoriel V muni d'une application $q : V \rightarrow F$ telle que $q(\lambda x) = \lambda^2 q(x)$ pour $(\lambda, x) \in F \times V$ et que $\check{q}(x, y) := \frac{1}{2}(q(x + y) - q(x) - q(y))$ (la *forme polaire* de q) soit bilinéaire : V est l'*espace sous-jacent* à q ; sa dimension est appelée la *dimension* de q , et notée $\dim q$. On a les notions classiques de vecteurs orthogonaux et d'orthogonal X^\perp d'une partie X de V . On dit que q est *non dégénérée* si $V^\perp = \{0\}$.

À partir de maintenant, « forme quadratique » signifie forme quadratique de dimension finie, non dégénérée.

Un *morphisme* de (V, q) vers (V', q') est une application linéaire $f : V \rightarrow V'$ telle que $q' \circ f = q$: alors f est automatiquement injective et $f(V)$ est facteur direct orthogonal de V' . Si f est un isomorphisme, on dit que c'est une *isométrie*. On note $q \simeq q'$ s'il existe une isométrie entre q et q' , et $q \leq q'$ s'il existe un morphisme de q vers q' (*i.e.* si q est isométrique à une sous-forme de q'). On note $O(q)$ le groupe des isométries d'une forme quadratique q : c'est le *groupe orthogonal* de q .

On dit aussi que deux formes q, q' sur F sont *semblables* s'il existe $a \in F^*$ tel que $q \simeq aq'$.

On peut additionner et multiplier les formes quadratiques :

- *Somme directe orthogonale* : $(V, q) \perp (V', q') = (V \oplus V', q \perp q')$, où $(q \perp q')(x, x') = q(x) + q'(x')$.
- *Produit tensoriel* : en termes de formes polaires, $(V, \check{q}) \otimes (V', \check{q}') = (V \otimes V', \check{q} \otimes \check{q}')$, où $(\check{q} \otimes \check{q}')(x \otimes x', y \otimes y') = \check{q}(x, y)\check{q}'(x', y')$.

On peut aussi *étendre les scalaires* de F à une extension quelconque K de F : on notera cette opération $q \mapsto q_K$.

Si (V, q) est une forme quadratique, par le choix d'une base (e_i) de V , q correspond à un polynôme $Q = \sum_i a_i T_i^2 + \sum_{i < j} 2a_{ij} T_i T_j$, où $a_i = q(e_i)$ et $a_{i,j} = \check{q}(e_i, e_j)$. Si $\dim q \geq 3$, Q est irréductible et définit une hypersurface lisse $X_q \subset \mathbf{P}(V)$: la *quadrique* associée à q . On a $\dim X_q = \dim q - 2$. Si $\dim q = 2$, X_q n'est plus (géométriquement) irréductible, mais consiste en deux points rationnels ou un point quadratique de $\mathbf{P}(V)$. Deux quadriques X_q et $X_{q'}$ sont isomorphes si et seulement si q et q' sont semblables.

2. LA THÉORIE DE WITT

2.1. Base orthogonale et théorème de simplification

THÉORÈME 2.1. — a) *Toute forme quadratique q possède une base orthogonale.*

b) *Soient q, q_1, q_2 trois formes quadratiques sur F . Si $q \perp q_1 \simeq q \perp q_2$, alors $q_1 \simeq q_2$.*

La partie a) de ce théorème est bien connue et sa démonstration se perd dans la nuit des temps. La partie b) (théorème de simplification) est essentiellement équivalente au fait que, si $\dim q = n$, tout élément de $O(q)$ est produit de n réflexions. Elle est couramment attribuée à Witt [69] ; toutefois, Scharlau [54, §2]⁽²⁾ a fait observer qu'elle avait été obtenue 30 ans auparavant par Dickson [15] (très probablement sans que Witt en ait conscience).

Si (e_1, \dots, e_n) est une base orthogonale de q et $x = \sum x_i e_i$, on a $q(x) = a_1 x_1^2 + \dots + a_n x_n^2$ avec $a_i = q(e_i) \in F^*$. On résume ceci par la notation $q \simeq \langle a_1, \dots, a_n \rangle$.

⁽²⁾Je remercie Serre de m'avoir indiqué cette référence.

2.2. Indice de Witt

Soit q une forme quadratique d'espace vectoriel sous-jacent V . Un vecteur $x \in V$ est *isotrope* si $q(x) = 0$. Un sous-espace $W \subset V$ est *totalelement isotrope* si $q|_W = 0$. La forme q est *isotrope* s'il existe un vecteur isotrope $\neq 0$, *anisotrope* si elle n'est pas isotrope.

On appelle *plan hyperbolique*, et on note \mathbb{H} , la forme quadratique de dimension 2, d'espace vectoriel sous-jacent F^2 , définie par $q(x, y) = xy$ ($(x, y) \in F^2$). Pour tout $a \in F^*$, on a $\mathbb{H} \simeq \langle a, -a \rangle$; toute forme quadratique isotrope de dimension 2 est isométrique à \mathbb{H} . On dit qu'une forme h est *hyperbolique* si $h \simeq m\mathbb{H}$ pour m convenable.

THÉORÈME 2.2. — *Toute forme quadratique q se décompose de manière unique (à isométrie près) en somme directe orthogonale $q_{\text{an}} \perp h$, où q_{an} est anisotrope et h est hyperbolique.*

Ce théorème, dû à Witt [69], se déduit facilement du théorème 2.1. Il ramène dans une large mesure l'étude des formes quadratiques à celle des formes quadratiques anisotropes. On en déduit immédiatement que, si (V, q) est une forme quadratique, tous les sous-espaces totalelement isotropes maximaux de V ont la même dimension : c'est l'*indice de Witt* de q , noté $i(q)$. Avec la notation du théorème 2.2, on a $i(q) = \frac{1}{2} \dim h \leq \frac{1}{2} \dim q$. La forme q est hyperbolique si et seulement si $i(q) = \frac{1}{2} \dim q$. Notons le lemme suivant, fort utile et qui donne une idée de l'esprit du sujet :

LEMME 2.3. — a) *Soient q, q' deux formes quadratiques anisotropes, et soit $n = i(q \perp -q')$. Alors il existe des formes quadratiques φ, q_1, q'_1 , avec $\dim \varphi = n$, telles que*

$$q \simeq \varphi \perp q_1, \quad q' \simeq \varphi \perp q'_1.$$

b) *Soient q une forme quadratique sur F et $q' \leq q$, de codimension r . Alors $i(q') \geq i(q) - r$. En particulier, si $\dim q' > \dim q - i(q)$ alors q' est isotrope.*

Voici une démonstration de b) : soient V l'espace sous-jacent à q , W le sous-espace de V correspondant à q' et $H \subset V$ un sous-espace totalelement isotrope de dimension $i(q)$. Alors $\text{codim}_H(W \cap H) \leq r$.

2.3. Anneau de Witt

DÉFINITION 2.4. — *Deux formes quadratiques q, q' sont équivalentes au sens de Witt si $q_{\text{an}} \simeq q'_{\text{an}}$; on note cette relation $q \sim q'$.*

L'équivalence de Witt respecte la somme et le produit des formes quadratiques, d'où

DÉFINITION 2.5. — *L'anneau de Witt de F , noté $W(F)$, est l'anneau des classes d'équivalence de la relation \sim , pour l'addition et la multiplication induites respectivement par \perp et \otimes .*

D'après le théorème 2.2, une forme quadratique est caractérisée, à isométrie près, par sa dimension et sa classe dans $W(F)$. On a parfois besoin de considérer une variante de l'anneau de Witt, l'anneau de Witt-Grothendieck : c'est l'anneau des classes d'équivalence de la relation \simeq de l'introduction, pour l'addition et la multiplication induites respectivement par \perp et \otimes . Il est noté (ici) $\hat{W}(F)$. Il est clair qu'on a un homomorphisme surjectif $\hat{W}(F) \rightarrow W(F)$, de noyau isomorphe à \mathbf{Z} (engendré par la classe du plan hyperbolique). D'après le théorème 2.1 a), ces deux anneaux sont engendrés par les classes des formes quadratiques de dimension 1, et il est facile d'en donner en fait une *présentation* (cf. [45, ch. II]).

3. LE THÉORÈME DE SPRINGER

THÉORÈME 3.1 ([57]). — *Soit q une forme quadratique anisotrope sur F , et soit E/F une extension finie de degré impair. Alors q_E est encore anisotrope.*

Ce théorème avait été conjecturé par Witt. Il est contenu dans un théorème plus récent de Swan [58] et Karpenko [32, prop. 2.6] :

THÉORÈME 3.2. — *Soit X une quadrique anisotrope sur F , et soit $x \in X$ un point fermé de degré 2. Alors tout 0-cycle sur X est linéairement équivalent à un multiple de x .*

Utilisons l'argument de Springer pour démontrer ce théorème⁽³⁾ : soit $\langle a_0, \dots, a_{d+1} \rangle$ une équation de X , où $d = \dim X$, et $X \subset \mathbf{P}^{d+1}$ le plongement projectif associé. Pour commencer, on peut trouver une droite $l \subset \mathbf{P}^{d+1}$ telle que $l \cap X = \{x\}$. Ensuite, on voit que tout point fermé est linéairement équivalent à un point fermé à corps résiduel séparable (si la caractéristique est $p > 0$ avec p impair, utiliser l'endomorphisme « de Frobenius » $(y_0 : \dots : y_{d+1}) \mapsto (a_0^{\frac{p-1}{2}} y_0^p : \dots : a_{d+1}^{\frac{p-1}{2}} y_{d+1}^p)$). Soit maintenant $y \in X$ un point fermé à corps résiduel séparable, de degré n : le choix d'un élément primitif α de $F(y)$ fournit des polynômes p_0, \dots, p_{d+1} de degrés $< n$ tels que $(y_0 : \dots : y_{d+1}) = (p_0(\alpha) : \dots : p_{d+1}(\alpha))$; on peut d'ailleurs supposer les p_i premiers entre eux dans leur ensemble. Soit C la courbe rationnelle définie par les p_i : elle est de degré $m < n$. De plus C n'est pas contenue dans X , sans quoi X aurait un point rationnel. Par conséquent, $Z = C \cap X$ est un 0-cycle effectif de X , de degré $2m$ et contenant y . Comme $C \sim ml$, on a $Z \sim mx$ (\sim désigne l'équivalence linéaire). Mais $Z - y$ est un 0-cycle effectif de degré $2m - n < n$: si $n = 2$, on a nécessairement $Z = y$ et, si $n > 2$, les autres points fermés intervenant dans $Z - y$ sont de degrés $< n$. Par récurrence, $Z - y$ est linéairement équivalent à un multiple de x , donc aussi y .

⁽³⁾Je remercie Hélène Esnault pour une remarque éclairante à ce sujet.

4. LA THÉORIE DE PFISTER : PUISSANCES DE I

L'anneau $W(F)$ est muni d'une augmentation « dimension modulo 2 »

$$\overline{\dim} : W(F) \rightarrow \mathbf{Z}/2.$$

On note $IF = \text{Ker } \overline{\dim}$: c'est l'idéal fondamental de $W(F)$. On note traditionnellement ses puissances $I^n F$.

4.1. Formes de Pfister

Il est clair que IF est engendré additivement par les formes binaires $\langle 1, -a \rangle$, et donc que $I^n F$ est engendré additivement par les formes

$$\ll a_1, \dots, a_n \gg := \langle 1, -a_1 \rangle \otimes \cdots \otimes \langle 1, -a_n \rangle.$$

Ces formes sont appelées *n-formes de Pfister*. Le premier, Pfister a reconnu que ce sont les pierres de touche de la théorie des formes quadratiques. Il en a démontré des propriétés remarquables :

THÉORÈME 4.1 ([52, Theorem 2]). — *Pour toute n-forme de Pfister φ , $\varphi(x) \neq 0 \Rightarrow \varphi \simeq \varphi(x)\varphi$. En particulier, les valeurs non nulles de φ forment un sous-groupe de F^* .*

On a même mieux : toute forme anisotrope multiplicative en ce sens est une forme de Pfister. Cela se déduit du théorème 4.3 ci-dessous.

Pour $n = 1, 2, 3$, on dispose classiquement d'algèbres expliquant le théorème 4.1 (extensions quadratiques de F , quaternions, octonions) : ce n'est plus le cas pour $n > 3$. D'ailleurs, pour ces valeurs de n , les formules implicites dans le théorème 4.1 contiennent des dénominateurs.

Du théorème 4.1, Pfister déduit facilement :

COROLLAIRE 4.2. — a) *Une forme de Pfister isotrope est hyperbolique.*
b) *Deux formes de Pfister semblables sont isométriques.*

4.2. Théorèmes de Cassels-Pfister

Il s'agit de trois théorèmes dits de *représentation*. Soit (V, q) une F -forme quadratique et soit A une F -algèbre commutative : on dit que q représente $a \in A$ sur A s'il existe $\vec{a} \in A \otimes_F V$ tel que $q(\vec{a}) = a$. Notation : $a \in D(q_A)$. La partie a) du théorème ci-dessous avait été initialement démontrée par Cassels pour des sommes de carrés ; la version générale et la suite sont dus à Pfister [52].

THÉORÈME 4.3. — a) *Soient q une forme quadratique sur F et $f \in F[T] - \{0\}$. Si q représente f sur $F(T)$, alors q représente déjà f sur $F[T]$.*

b) *Soient $q = \langle a_1, \dots, a_n \rangle$ une forme anisotrope sur F ($n \geq 2$), $q' = \langle a_2, \dots, a_n \rangle$ et $d \in F^*$. Alors $d \in D(q')$ $\iff d + a_1 T^2 \in D(q_{F(T)})$.*

c) *Soient q, φ deux formes quadratiques sur F , avec q anisotrope. Soient V l'espace*

vectorel sous-jacent à φ et $K = F(V^*)$, de sorte que φ peut être vu comme élément de K . Alors $\varphi \in D(q_K) \iff \varphi \leq q$.

4.3. Le discriminant

À part la dimension modulo 2, c'est le seul invariant que nous utiliserons. Si $q = \langle a_1, \dots, a_n \rangle$, on pose

$$d_{\pm}q = (-1)^{\frac{n(n-1)}{2}} a_1 \dots a_n \in F^*/F^{*2}.$$

C'est le discriminant de q : il ne dépend que de la classe de q dans $W(F)$. Il interviendra implicitement à cause du lemme suivant :

LEMME 4.4. — a) L'application $q \mapsto d_{\pm}q$ induit un isomorphisme $IF/I^2F \xrightarrow{\sim} F^*/F^{*2}$.

b) Soient q, q' deux formes de dimension impaire. Alors il existe un scalaire a tel que $q \perp -aq' \in I^2F$. □

5. CORPS DE FONCTIONS DE QUADRIQUES

Soit q une forme quadratique sur F , et soit $X = X_q$ la quadrique projective associée : c'est une variété F -rationnelle si et seulement si q est isotrope (cela résulte facilement du théorème 2.2). On note en général $F(q)$ le corps des fonctions $F(X)$. C'est un invariant important de q (ou de X). On peut dire que, tenant compte des travaux antérieurs de Pfister, son utilisation systématique remonte véritablement à Arason [6, 4].

5.1. Théorème de la sous-forme

PROPOSITION 5.1. — Soient q, φ deux formes quadratiques sur F , avec φ anisotrope et $\dim q = n > 2$. Supposons $\varphi_{F(q)} \sim 0$. Alors, pour tout $b \in F^*$ représenté par q , on a

$$bq(T_1, \dots, T_n)\varphi_K \simeq \varphi_K$$

où $K = F(T_1, \dots, T_n)$.

Cette proposition se démontre facilement par réduction à une extension quadratique. Knebusch [42] en a démontré la réciproque.

Du théorème 4.3 c) et de la proposition 5.1, on déduit l'important théorème suivant, appelé théorème d'Arason-Cassels-Pfister ou simplement théorème de la sous-forme :

THÉORÈME 5.2. — Soient q, φ deux formes quadratiques sur F , avec φ anisotrope et $\dim q > 2$. Supposons $\varphi_{F(q)} \sim 0$. Alors, pour tout $(a, b) \in D(q) \times D(\varphi)$, on a $abq \leq \varphi$. En particulier, $\dim \varphi \geq \dim q$.

COROLLAIRE 5.3 (Arason [4, Satz 1.3]). — Si dans le théorème 5.2, on suppose que q est une forme de Pfister, alors $\varphi \simeq q \otimes \rho$ pour $\rho \in W(F)$ convenable.

Cela se voit par récurrence sur $\dim q$, en notant que $q_{F(q)} \sim 0$ par le corollaire 4.2.

5.2. Théorème d’Arason-Pfister

C’est le suivant :

THÉORÈME 5.4 ([6]). — *Soit q anisotrope sur F telle que $q \in I^n F$ pour $n \geq 0$. Alors, $\dim q \geq 2^n$.*

De ce théorème il résulte immédiatement que $\bigcap I^n F = \{0\}$. Donnons-en la démonstration : Écrivons $q = \pm\varphi_1 + \cdots + \pm\varphi_r$ dans $W(F)$, où les φ_i sont des n -formes de Pfister. Puisque q est anisotrope, on a $r > 0$. Procédons par récurrence sur r . Si $r = 1$ et $\dim q < 2^n$, on a $\pm\varphi_1 \simeq q \perp m\mathbb{H}$ pour $m > 0$ convenable ; alors φ_1 est hyperbolique (corollaire 4.2) et q aussi, absurde. Si $r > 1$, posons $K = F(\varphi_r)$. Distinguons deux cas :

1. q_K est hyperbolique. Par le corollaire 5.3, il existe ρ tel que $q \simeq \varphi_r \otimes \rho$. En particulier, $2^n = \dim \varphi_r \mid \dim q$, d’où $\dim q \geq 2^n$.
2. q_K n’est pas hyperbolique. Soit $q' = (q_K)_{\text{an}}$. Alors $q' = \pm\varphi_1 + \cdots + \pm\varphi_{r-1}$ dans $W(K)$. Par récurrence sur r , $\dim q' \geq 2^n$; alors $\dim q \geq 2^n$ également. \square

6. LA THÉORIE DE KNEBUSCH : DÉPLOIEMENT GÉNÉRIQUE

Dans [43] et [44], Knebusch développe une théorie de déploiement générique pour les formes quadratiques, qui peut se voir comme une conceptualisation et une extension des résultats précédents. Elle conduit à la définition de deux invariants numériques importants des formes quadratiques : la *hauteur* et le *degré*.

6.1. Tours de déploiement génériques ; hauteur

Considérons une forme quadratique q sur F . On associe à q un entier $h = ht(q)$ et une suite $(F_i, q_i)_{0 \leq i \leq h}$, où F_i est une extension de F et q_i est une forme quadratique sur F_i , de la manière suivante :

- (i) $q_0 = q_{\text{an}}, F_0 = F$.
- (ii) Supposons F_i et q_i définis. Si $\dim q_i = 0$ ou 1, alors $i = h$. Sinon, $F_{i+1} = F_i(q_i)$ et $q_{i+1} = ((q_i)_{F_{i+1}})_{\text{an}}$.

DÉFINITION 6.1. — *L’entier $ht(q)$ s’appelle la hauteur de q ⁽⁴⁾. La suite $F = F_0 \subset F_1 \subset \cdots \subset F_h$ s’appelle la tour de déploiement canonique de q . La forme q_i s’appelle le i -ième noyau de q . Le corps F_{h-1} (resp. F_h) s’appelle le corps dominant (resp. corps de déploiement canonique) de q . On note, pour $n > 0$*

$$i_n(q) = i((q_{n-1})_{F_n})$$

c’est le n -ième indice de Witt supérieur de q . La suite

$$(i_1(q), \dots, i_h(q))$$

⁽⁴⁾Nous utilisons ici la notation $ht(q)$ plutôt que $h(q)$ pour éviter toute confusion avec le motif de la quadrique X définie par q , qui sera noté $h(X)$.

s'appelle la suite de déploiement de q .

Si X est la quadrique définie par q , on notera aussi $i_n(X) := i_n(q)$.

Comme $\dim q \geq \dim q_0 > \dim q_1 > \dots$, l'entier h est bien défini; comme les $\dim q_i$ ont la même parité que $\dim q$, $ht(q) \leq \frac{1}{2} \dim q$.

« Suite de déploiement » est, faute de mieux, une traduction française de « Splitting pattern », terminologie introduite par Hurrelbrink et Rehmann. Signalons que leur définition diffère de la nôtre, qui est celle de Vishik : ils utilisent plutôt la suite $(i_1(q), i_1(q) + i_2(q), \dots)$. Quant à Hoffmann et Izhboldin, ils préfèrent utiliser la suite $(\dim q_0, \dots, \dim q_h)$. On prendra donc garde que la terminologie recouvre plusieurs définitions (aux contenus équivalents) dans la littérature.

Pour alléger, introduisons la notation

$$\dim_{\text{an}} q = \dim q_{\text{an}}.$$

THÉORÈME 6.2 (Knebusch [43, th. 5.1]). — *Soient q une forme quadratique sur F et K/F une extension quelconque. Alors il existe un unique $i \in \{0, \dots, ht(q)\}$ tel que $\dim_{\text{an}} q_K = \dim q_i$.*

L'unicité est claire. Knebusch démontre l'existence en utilisant sa théorie de la spécialisation des formes quadratiques [42]⁽⁵⁾ : donnons-en une démonstration directe. Soit i le plus petit entier tel que $\dim q_i \leq \dim_{\text{an}} q_K$. Il faut montrer que $\dim q_i = \dim_{\text{an}} q_K$. Posons, pour tout $j < h$, $K_j = KF_j$. Pour $j < i$, $(q_j)_{K_j} \sim q_{K_j}$ est isotrope; alors K_{j+1}/K_j est transcendante pure (voir début du §5). Il en résulte que K_i/K est transcendante pure, d'où il résulte facilement que $((q_K)_{\text{an}})_{K_i}$ est anisotrope. Mais alors

$$(6.1) \quad \dim_{\text{an}} q_{K_i} = \dim_{\text{an}} q_K \geq \dim q_i \geq \dim_{\text{an}} (q_i)_{K_i}$$

et

$$(q_{K_i})_{\text{an}} \simeq ((q_i)_{K_i})_{\text{an}}.$$

Par conséquent il y a partout égalité dans (6.1). □

DÉFINITION 6.3. — *Une suite $F = K_0 \subset \dots \subset K_h$ est une tour de déploiement générique de q si elle possède la propriété du théorème 6.2 avec $q_i = (q_{K_i})_{\text{an}}$ et si les extensions K_{i+1}/K_i sont régulières.*

(Dans [43], Knebusch demande de plus que, pour toute extension E/F , il existe une F -place de K_i vers E ayant « bonne réduction » en q_i , où i est l'entier tel que $\dim_{\text{an}} q_E = \dim q_i$, mais cette propriété est automatique puisque l'extension composée EK_i/E est transcendante pure, cf. remarque au début du §5.)

La définition suivante, due à Izhboldin, est particulièrement importante.

⁽⁵⁾Il démontre plus : il existe une F -place $\lambda : F_i \dashrightarrow K$ telle que q_i ait bonne réduction en λ et que $\lambda(q_i) \simeq q_K$.

DÉFINITION 6.4. — *Supposons q anisotrope. La dimension essentielle de X_q est*

$$\dim_{\text{es}} X_q = \dim X_q - i_1(q) + 1.$$

Nous poserons aussi

$$\dim_{\text{es}} q = \dim q - i_1(q) + 1.$$

Le lemme suivant est conséquence immédiate du lemme 2.3 b).

LEMME 6.5. — *Soient q une forme anisotrope et $q' \leq q$. Supposons que $\dim q' \geq \dim_{\text{es}} q$. Alors $q'_{F(q)}$ est isotrope.*

Nous verrons plus loin (§8.2) que *la réciproque est vraie.*

6.2. Interprétation en termes de groupes algébriques

Kersten et Rehmann [41] ont généralisé la notion de tour de déploiement générique dans le contexte des *groupes algébriques linéaires*.

Soit G un F -groupe réductif connexe, de diagramme de Dynkin Δ , et soit Θ un sous-ensemble de Δ . Une extension K/F est un Θ -corps de déploiement de G si G_K contient un K -sous-groupe parabolique P de type Θ ; K est dit *générique* si tout Θ -corps de déploiement de G est une F -spécialisation de K (au sens des F -places). L'existence d'un tel K se démontre ainsi : étant donné P (défini sur une clôture séparable \bar{F} de F), l'espace projectif homogène $V_\Theta = G_{\bar{F}}/P$ est défini sur une plus petite extension finie séparable F_Θ de F : l'extension minimale telle que la $*$ -action de $\text{Gal}(\bar{F}/F_\Theta)$ sur Δ laisse Θ invariant. On a alors $K = F_\Theta(V_\Theta)$.

Dans le cas d'une forme quadratique (V, q) , avec $\dim q = n \geq 3$, le groupe $G = SO(q)$ est de rang $i(q)$. Si $i(q) \geq i$, alors $SO(q)$ opère transitivement sur l'ensemble des sous-espaces totalement isotropes de V de dimension i ; si $i < n/2$, le stabilisateur de l'un d'entre eux est un sous-groupe parabolique P_i correspondant à $\Delta - \{\alpha_i\}$, où α_i est la i -ème racine. Réciproquement, si P_i est rationnel sur F on a $i(q) \geq i$. La variété $V_i = G/P_i$ correspondante est la « grassmannienne quadratique » des sous-espaces totalement isotropes de rang i de V . (Dans le cas $i = n/2$, c'est plus compliqué.) Sans hypothèse sur $i(q)$, une tour de déploiement générique de q est « contenue » dans la suite de corps $K_i = F(V_i)$: elle est plus « économique » que celle de la définition 6.1 en ce que ses degrés de transcendance sont plus petits, cf. [41, p. 61].

6.3. Formes de hauteur 1 ; degré

La proposition suivante est due indépendamment à Knebusch et Wadsworth ([43, démonstration du th. 5.8], [67]) :

PROPOSITION 6.6. — *Une forme q est de hauteur 1 si et seulement si elle est*

- semblable à une forme de Pfister au cas où $\dim q$ est paire ;
- semblable à une sous-forme de codimension 1 d'une forme de Pfister au cas où $\dim q$ est impaire.

Elle conduit immédiatement à la

DÉFINITION 6.7. — *Soit q une forme quadratique sur F .*

- a) *Si $\dim q$ est impaire, le degré de q est $\deg(q) = 0$.*
- b) *Si $\dim q$ est paire et non hyperbolique, soit F_{h-1} son corps dominant. Par la proposition 6.6, q_{h-1} est semblable à une forme de Pfister τ : τ est appelée forme dominante de q . Si $\dim \tau = 2^n$, le degré de q est l'entier $\deg(q) = n$.*
- c) *Si $q \sim 0$, $\deg(q) = \infty$.*

Cette notion est intéressante à cause du théorème suivant, dû à Knebusch [43, th. 6.4].

THÉORÈME 6.8. — *Pour tout $n \geq 0$, l'ensemble*

$$J_n(F) = \{q \in W(F) \mid \deg(q) \geq n\}$$

est un idéal de $W(F)$ contenant $I^n F$.

(Si $a \in F^*$ et $q \in J_n(F)$, il est clair que $\langle a \rangle q \in J_n(F)$, et de même il est clair que $J_n(F)$ contient les n -formes de Pfister ; la difficulté est de voir que $J_n(F)$ est stable par addition.)

Remarque 6.9. — En fait, on a $J_n(F) = I^n F$: ce théorème est dû à Orlov-Vishik-Voevodsky [50]. Leur démonstration repose sur les techniques de [66], donc sur la théorie homotopique des schémas. Il n'en sera pas fait usage ici.

Le théorème 6.8 fournit une nouvelle démonstration du théorème 5.4 : si $q \in I^n F - \{0\}$, alors $\deg(q) \geq n$, donc $\dim(q) \geq 2^n$. On a de plus le complément suivant :

COROLLAIRE 6.10. — *Soit q de dimension 2^n . Si $q \in J_n(F)$, alors q est semblable à une forme de Pfister. \square*

Comme application de la théorie de Knebusch, mentionnons le raffinement suivant du théorème de la sous-forme 5.2 :

THÉORÈME 6.11 (cf. [43, prop. 6.11], [5, Satz 18]). — *Soit φ une forme quadratique sur F de dimension paire, de corps dominant L et de forme dominante τ . Soit q une autre forme sur F . Alors $\deg(\varphi_{F(q)}) > \deg \varphi$ si et seulement si q_L est semblable à une sous-forme de τ . En particulier, en posant $n = \deg(\varphi)$:*

- (i) *Si $\dim q > 2^n$, on a $\deg(\varphi_{F(q)}) = \deg \varphi$.*
- (ii) *Si $\dim q = 2^n$, on a $\deg(\varphi_{F(q)}) > \deg \varphi$ si et seulement si q est semblable à une forme de Pfister τ_0 telle que $(\tau_0)_L \simeq \tau$.*

En particulier, $\varphi_{F(q)} \sim 0 \Rightarrow \dim q \leq 2^{\deg(\varphi)}$.

Dans la même direction on a le théorème sensationnel suivant, dû à Fitzgerald [17, th. 1.6] :

THÉORÈME 6.12. — Soient q, φ deux formes quadratiques sur F , avec $\varphi \not\sim 0$. Supposons que $\varphi_{F(q)} \sim 0$ et que φ ne soit pas semblable à une forme de Pfister. Alors $\dim \varphi - 2^{\deg(\varphi)} \geq 2 \dim q$.

6.4. Voisines de Pfister ; formes excellentes

Les notations suivantes sont très utiles :

NOTATION 6.13. — Pour un entier n , on note $l(n)$ l'unique entier tel que $2^{l(n)-1} < n \leq 2^{l(n)}$. Pour une forme quadratique q , on note $l(q) = l(\dim q)$.

La définition suivante est due à Knebusch [44].

DÉFINITION 6.14. — Soient q une forme quadratique et φ une n -forme de Pfister (éventuellement isotropes). La forme q est dite voisine de φ si

- (i) $\dim q > 2^{n-1}$
- (ii) q est semblable à une sous-forme de φ .

Une forme q' telle que $q \perp q'$ soit semblable à φ s'appelle forme complémentaire de q .

Noter que $n = l(q)$. En utilisant le théorème d'Arason-Pfister, on démontre facilement :

THÉORÈME 6.15 ([44, p. 3]). — Les formes φ et q' de la définition 6.14 sont uniquement déterminées par q .

Knebusch démontre ensuite :

THÉORÈME 6.16. — Pour une F -forme quadratique q , les conditions suivantes sont équivalentes :

- (i) Pour toute extension K/F , la forme $(q_K)_{\text{an}}$ est définie sur F .
- (ii) Pour tout i , la forme q_i de la définition 6.1 est définie sur F .
- (iii) Il existe des F -formes de Pfister τ_1, \dots, τ_h ($h = ht(q)$), telles que $\tau_j \mid \tau_{j-1}$ pour $j \in [1, h]$, des F -formes q'_0, \dots, q'_i et un scalaire a tels que $q'_0 = q_{\text{an}}$ et que, pour tout $j \in [1, h]$, on ait $q'_{j-1} \perp -q'_j \simeq (-1)^{j+1} a \tau_j$.

Lorsque ces conditions sont vérifiées, on a

$$[q] = a([\tau_1] - [\tau_2] + \dots + (-1)^{h+1} [\tau_h])$$

dans $W(F)$.

DÉFINITION 6.17. — Une forme vérifiant les conditions équivalentes du théorème 6.16 est dite excellente.

Les formes excellentes peuvent être considérées comme les plus simples des formes quadratiques. Par exemple, la suite de déploiement S d'une forme excellente q (en termes de dimensions anisotropes) ne dépend que de $n = \dim q$, à savoir :

$$S = \{n, c(n), c(c(n)), \dots\}$$

où, pour tout $k \in \mathbf{N}$, $c(k) = 2^{l(k)} - k$ (notation 6.13).

7. ÉQUIVALENCE BIRATIONNELLE STABLE

7.1. Généralités

Soit X une variété projective homogène. Les conditions suivantes sont équivalentes [9, th. 21.20] :

- (i) X a un point rationnel.
- (ii) Il existe une F -place de $F(X)$ vers F .
- (iii) X est une variété F -rationnelle, *i.e.* l'extension $F(X)/F$ est transcendante pure.

Soit Y une autre variété; de ce qui précède, il résulte que les conditions suivantes sont équivalentes :

- (i) X a un point rationnel sur $F(Y)$; autrement dit, il existe une application F -rationnelle de Y vers X .
- (ii) Il existe une F -place de $F(X)$ vers $F(Y)$.
- (iii) $X \times Y$ est Y -birationnel à $\mathbf{P}^n \times Y$ ($n = \dim X$).

Si ces conditions sont vérifiées, on dit que Y domine X et on écrit $Y \succcurlyeq X$ ou $X \preccurlyeq Y$. Cette relation est particulièrement intéressante quand Y est aussi projective homogène : la condition (ii) montre qu'elle définit une relation de *préordre* sur la famille de ces variétés. La relation d'équivalence associée est l'équivalence birationnelle stable : nous la noterons ici \approx (d'autres auteurs utilisent la notation $\overset{st}{\sim}$).

Nous utiliserons principalement les relations \preccurlyeq et \approx dans le cas des *quadriques*. On notera $q \preccurlyeq q'$ pour $X_q \preccurlyeq X_{q'}$, etc. On a les propriétés évidentes suivantes (la troisième résultant du lemme 6.5) :

LEMME 7.1. — a) La relation \preccurlyeq (resp. \approx) définit une relation de *préordre* (resp. d'*équivalence*) sur (l'ensemble sous-jacent à l'anneau) $W(F)$.

b) $q' \leq q \Rightarrow q' \preccurlyeq q$.

c) Si $q' \leq q$ et $\dim q' \geq \dim_{\text{es}} q$, alors $q' \approx q$.

De ce lemme, du corollaire 4.2 et du théorème 5.2, on déduit facilement :

THÉORÈME 7.2. — Soient π une forme de Pfister, q une voisine de π et q' une forme quadratique anisotrope sur F . Alors,

a) $q \approx \pi$.

b) $q' \preceq q \iff q'$ est semblable à une sous-forme de π .

c) $q' \approx q \iff q'$ est voisine de π .

En fait l'implication \Rightarrow dans c) est également vraie, mais c'est un résultat plus profond, cas particulier immédiat du théorème de Hoffmann qui suit.

7.2. Théorème de Hoffmann

C'est le suivant :

THÉORÈME 7.3. — Si $q \preceq q'$, alors $l(q) \leq l(q')$ (cf. notation 6.13). En d'autres termes, s'il existe n tel que $\dim q' \leq 2^n < \dim q$, alors $q'_{F(q)}$ est anisotrope.

Pour la démonstration originelle on pourra se reporter à [21]; une démonstration un peu plus simple, mais dans le même esprit, se trouve dans [23]. On peut aussi déduire ce théorème de la *formule du degré de Rost* [46, th. 5.3.1] (cette observation est due à Rost). Nous déduisons au §8 le théorème de Hoffmann de résultats plus fins obtenus ultérieurement (corollaire 8.3 et théorème 8.5).

COROLLAIRE 7.4. — Pour toute q anisotrope, on a $\dim_{\text{es}} q > 2^{l(q)-1}$.

Pour voir ceci, prendre une sous-forme $q' \leq q$ de dimension $2^{l(q)-1}$ et appliquer le lemme 7.1 c) et le théorème 7.3. On a même [21, 23] :

PROPOSITION 7.5. — Pour q anisotrope, les conditions suivantes sont équivalentes :

- (i) $\dim_{\text{es}} q = 2^{l(q)-1} + 1$;
- (ii) il existe K/F tel que q_K soit voisine d'une K -forme de Pfister et $ht(q_K) = ht(q)$.

Une forme vérifiant les propriétés de la proposition 7.5 est dite à *déploiement maximal* ($i_1(q)$ prend la plus grande valeur possible) : ces formes sont donc « stablement » des voisines de Pfister, mais on a des exemples de formes à déploiement maximal qui ne sont pas des voisines de Pfister (par exemple de dimension 5). Toutefois, quand $\dim q$ est « proche » de $2^{l(q)}$, une forme à déploiement maximal est une voisine de Pfister. Conjecturalement c'est le cas quand $\dim q > 5 \cdot 2^{l(q)-3}$, ce qui est connu pour $l(q) \leq 4$; pour $l(q) > 4$, c'est le cas quand $\dim q \geq 2^{l(q)} - 7$ [25, th. 1.7].

8. QUATRE RÉSULTATS FONDAMENTAUX

Jusqu'à maintenant, les résultats énoncés avaient été obtenus par des méthodes ne faisant intervenir que des résultats élémentaires d'algèbre commutative, voire de théorie des corps. Par contre, la démonstration des suivants utilise les correspondances algébriques, quoique de manière élémentaire pour une large part.

8.1. Dimension des formes dans I^n

THÉORÈME 8.1. — Soit $q \in I^n F$, anisotrope. Alors

- soit $\dim q \geq 2^{n+1}$ (et $\dim q$ est paire) ;
- soit $\dim q$ est de la forme $2^{n+1} - 2^i$ pour un entier $i \in [1, n]$.

De plus, toutes les dimensions visées sont atteintes.

Ce théorème généralise le théorème d’Arason-Pfister ; il avait été conjecturé par Vishik. La dernière partie est relativement facile. Il est dû à Vishik et à Karpenko, seule la démonstration de Karpenko étant rédigée [37]. Le cas particulier disant que $\dim q > 2^n \Rightarrow \dim q \geq 2^n + 2^{n-1}$ avait été démontré par Pfister pour $n = 3$, par Hoffmann pour $n = 4$ et par Vishik pour tout n [62, th. 6.4] en utilisant le théorème 11.31 ci-dessous. Voir aussi dans [40, th. 4.4] une démonstration de ce cas particulier due à Hoffmann et ne reposant que sur le corollaire 8.3 ci-dessous. Le théorème 8.1 rappelle de manière frappante le comportement des premières classes de Stiefel-Whitney non nulles d’un fibré quadratique (ou réel), cf. [28, prop. 1.1 (c)].

8.2. Dimension essentielle des quadriques

THÉORÈME 8.2. — Soient X une quadrique anisotrope et Y une F -variété propre (éventuellement singulière) dont tous les points fermés sont de degré pair. Supposons que $Y_{F(X)}$ ait un point fermé de degré impair. Alors

1. $\dim(Y) \geq \dim_{\text{es}}(X)$;
2. si $\dim(Y) = \dim_{\text{es}}(X)$, $X_{F(Y)}$ est isotrope.

COROLLAIRE 8.3. — Soient q, q' deux formes quadratiques anisotropes telles que $q \preceq q'$. Alors

1. $\dim_{\text{es}} q \leq \dim_{\text{es}} q'$;
2. $\dim_{\text{es}} q = \dim_{\text{es}} q' \Rightarrow q \approx q'$.

Ces résultats sont dus à Karpenko-Merkurjev [40] ; le corollaire 8.3 avait été conjecturé par Izhboldin. Mutatis mutandis, l’énoncé du théorème 8.2 est exactement le même que celui d’un résultat annoncé par Voevodsky dans une lettre à Rost [30, th. 9.3]⁽⁶⁾. Plus précisément, chez Voevodsky :

- Y est lisse ;
- F est de caractéristique 0 ;
- l’énoncé vaut pour un nombre premier l quelconque ;
- X est supposé être une « (v_n, l) -variété » (voir *loc. cit.*) et la conclusion de (1) est $\dim Y \geq \dim X$.

⁽⁶⁾Dans *loc. cit.*, il faut lire « tout morphisme *au-dessus de* $\mathbf{Z}_{(l)}$ ».

(Une quadrique de dimension d est une $(v_n, 2)$ -variété si et seulement si $d = 2^n - 1$: pour $l = 2$, le théorème 8.2 n'est donc pas recouvert par celui de Voevodsky, même pour $i_1(X) = 1$ et même sous les hypothèses additionnelles de ce dernier sur F et Y .)

Le cas particulier suivant avait été obtenu antérieurement par Vishik ([62, cor. 4.9], cf. §11.4 ci-dessous) :

COROLLAIRE 8.4. — $q \approx q' \Rightarrow \dim_{\text{es}} q = \dim_{\text{es}} q'$.

L'hypothèse est en particulier vérifiée si $q \leq q'$ et $q' \preceq q$, ce qui fournit la réciproque promise du lemme 6.5.

Pour la démonstration du théorème 8.2, voir §12.1.

8.3. Une borne pour $i_1(q)$

THÉORÈME 8.5. — *Pour toute forme anisotrope q , on a $i_1(q) \leq |\dim_{\text{es}} q - 1|_2^{-1}$, où $|\cdot|_2$ est la valuation dyadique.*

Ce théorème est dû à Karpenko [36] ; il avait été conjecturé par Hoffmann. En particulier, $i_1(q) = 1$ si $\dim_{\text{es}} q$ est paire. Une autre formulation est la suivante : il existe un entier n tel que $i_1(q) - 1$ soit le reste de la division de $\dim q - 1$ par 2^n .

Pour tout entier m , notons

$$m^{\text{es}} = m - 1 + |m - 1|_2^{-1}.$$

Un peu d'arithmétique élémentaire montre que $l(m) = l(m^{\text{es}})$ (cf. notation 6.13) : ainsi, le théorème 8.5 implique le corollaire 7.4. Il en découle que le corollaire 8.3 et le théorème 8.5 impliquent le théorème de Hoffmann 7.3. (Leurs démonstrations n'utilisent pas ce théorème!)

Pour la démonstration du théorème 8.5, voir §12.2.

8.4. Une relation entre les indices de Witt supérieurs

THÉORÈME 8.6. — *Soit q une forme quadratique de hauteur h , et soient i_1, \dots, i_h ses indices de Witt supérieurs. Notons v_2 la valuation dyadique. Alors, pour tout $q \in [1, h - 1]$, on a*

$$v_2(i_q) \geq \inf(v_2(i_{q+1}), \dots, v_2(i_h)) - 1.$$

Si de plus $\dim q$ est paire et l'entier $i_q + 2(i_{q+1} + \dots + i_h)$ n'est pas une puissance de 2, on a

$$v_2(i_q) \leq \sup(v_2(i_{q+1}), \dots, v_2(i_h)).$$

Ce théorème est dû à Karpenko [38]. Il en déduit une autre démonstration du théorème 8.1 : si q est un contre-exemple à ce théorème, on se ramène en montant dans sa tour de déploiement générique à supposer que q_1, \dots, q_h en vérifient la conclusion, ce qui contredit la borne inférieure du théorème 8.6.

9. TROIS APPLICATIONS

9.1. Dimension des formes de hauteur 2

THÉORÈME 9.1. — Soit q une forme anisotrope de hauteur 2 et de degré $n \geq 0$. Alors
 a) Si $n = 0$, $\dim q$ est de la forme $2^a - 2^b + 1$ et $i_1(q)$ est de la forme $2^{a-1} - 2^b + 1$ pour $a > b > 1$ (le cas $a = b + 1$ est permis).

b) Si $n > 0$, on a

$$\dim q \in \{2^n + 2^{n-1}, 2^{n+1}, 2^{r+1} - 2^n (r > n)\}.$$

Ce théorème est dû à Vishik [61, th. 3.1]. Donnons-en une autre démonstration, reposant sur les théorèmes 8.5 et 8.6 :

Démonstration. — Traitons b) : le cas de a) se traite de même. D’après la proposition 6.6, on a $\dim q_1 = \dim q - 2i_1(q) = 2^n$. Appliquons le théorème 8.5 : il existe r tel que $i_1(q) \leq 2^r$ et $\dim q - 1 \equiv i_1(q) - 1 \pmod{2^r}$. Deux cas se présentent :

1. $r > n$. Alors $i_1(q) = 2^r - 2^n$, donc $\dim q = 2^{r+1} - 2^n$.
2. $r \leq n$. On a $i_1(q) \equiv 0 \pmod{2^r}$, donc $i_1(q) = 2^r$ et $\dim q = 2^n + 2^{r+1}$.

Cette alternative avait été obtenue antérieurement par Vishik dans sa thèse, au moins quand -1 est un carré [60, Statement 6.2]. On applique maintenant le théorème 8.6, qui montre que (2) n’est possible que pour $r \geq n - 2$ (noter que, pour $r = n$, on retrouve le cas $r = n + 1$ de (1)). \square

Plus précisément, on conjecture (par exemple [60, Question 6.4]) :

CONJECTURE 9.2. — Une forme anisotrope q de hauteur 2 et de degré $n > 0$ est d’un des trois types suivants :

- (i) *excellente* ;
- (ii) $q \simeq \varphi \otimes \psi$, où φ est une $(n - 1)$ -forme de Pfister et $\dim \psi = 4$, $\psi \notin I^2 F$;
- (iii) $q \simeq \varphi \otimes \psi$, où φ est une $(n - 2)$ -forme de Pfister et $\dim \psi = 6$, $\psi \in I^2 F$ (on dit que ψ est une forme d’Albert).

Cette conjecture est connue pour $n = 1$ (Knebusch, [44, §10]) et pour $n = 2$ [29].

Remarque 9.3 (Cette remarque et le lemme qui la suit sont entièrement dus à Hoffmann)

Dans le cas $n = 0$, la situation est un peu différente. D’après Knebusch [43, §10], q est excellente dès que sa forme dominante est définie sur F : il obtient ainsi le théorème 9.1 a) dans ce cas particulier. D’autre part, si $\dim q = 5$, elle est de hauteur 2 mais pas excellente en général (par exemple si q est une sous-forme d’une forme d’Albert anisotrope, cf. conjecture 9.2 (iii)). Mais par ailleurs, le théorème 9.1 a) entraîne que q est à déploiement maximal si $a > b + 1$. Pour $a = b + 1$, q est excellente d’après le lemme 9.4 ci-dessous. La conjecture mentionnée à la fin du §7 implique ainsi que q doit être excellente en toute dimension $\neq 5$. C’est par exemple vrai pour $a \leq 4$ ou pour $b \leq 3$.

LEMME 9.4. — Soit q une forme de hauteur 2 et de dimension $2^b + 1$, avec $b \geq 3$. Alors q est excellente.

Démonstration. — Comme q est de hauteur 2, il existe un scalaire $u \in F_1^*$ tel que $q_1 \perp \langle u \rangle$ soit semblable à une b -forme de Pfister (proposition 6.6). En fait, la classe de carrés u est définie sur F (pour le voir, on peut l'interpréter comme le discriminant de q_1 , et donc de q). Soit $q' = q \perp \langle u \rangle$. Notons que $q' \in I^2F$, donc que $\deg(q') \geq 2$. Évidemment, on a $\deg(q') \leq b$. Supposons que $\deg(q') < b$. Alors il existe une extension L telle que

$$4 \leq \dim_{\text{an}}(q'_L) \leq 2^{b-1}$$

et donc

$$1 < 3 \leq \dim_{\text{an}}(q_L) \leq 2^{b-1} + 1 < 2^b - 1 \text{ (car } b \geq 3)$$

Or par hypothèse, les seules dimensions « anisotropes » possibles de q sont $1, 2^b - 1, 2^b + 1$. D'où une contradiction.

Par conséquent, $\deg(q') = b$. Comme $b \geq 3$, le théorème 8.1 implique que $\dim_{\text{an}} q' = 2^b$ (observer qu'a priori $\dim_{\text{an}}(q') \geq 2^b$ par construction de q'). Ainsi $q = q'' + \langle -u \rangle$ avec q'' semblable à une b -forme de Pfister. \square

9.2. Une borne pour l'indice de Witt

THÉORÈME 9.5. — Supposons que $q \approx q'$ où q et q' sont deux formes anisotropes. Alors

$$i(q'_{F(q)}) - i_1(q') \leq \dim_{\text{es}} q' - \dim_{\text{es}} q.$$

Ce théorème est dû à Karpenko-Merkurjev [40, cor. 4.2]; une variante tout aussi frappante est $\dim q' - i(q'_{F(q)}) + 1 \geq \dim_{\text{es}} q$. Voici leur démonstration : posons $X = X_q$ et $Y = X_{q'}$. Si $\dim_{\text{es}}(X) = 0$, l'énoncé est trivial. Sinon, soit Y' une sous-quadrique de Y de dimension $\dim_{\text{es}}(X) - 1$. Puisque $\dim_{\text{es}}(Y') < \dim(Y') < \dim_{\text{es}}(X)$, la quadrique Y' reste anisotrope sur $F(X)$ par la première partie du corollaire 8.3. Donc, d'après le lemme 2.3 b), on a $i(Y_{F(X)}) \leq \text{codim}_Y(Y') = \dim(Y) - \dim_{\text{es}}(X) + 1$, d'où l'inégalité.

9.3. Degré de transcendance d'un corps d'isotropie générique

Par définition, un corps d'isotropie générique d'une forme anisotrope q est un corps de type K_1 , où (K_i) est une tour de déploiement générique au sens de la définition 6.3.

THÉORÈME 9.6. — Le plus petit degré de transcendance d'un corps d'isotropie générique K de q est égal à $\dim_{\text{es}}(X_q) (= \dim_{\text{es}}(q) - 2)$.

Ce théorème est encore dû à Karpenko-Merkurjev [40, th. 4.3] (ils utilisent la terminologie « corps de déploiement générique », attribuée par Knebusch au corps K_h de la définition 6.3). Il le déduisent facilement du théorème 8.2 : nous renvoyons à *loc. cit.* pour les détails.

PARTIE II

CYCLES ALGÈBRIQUES

10. FORMES QUADRATIQUES ET MOTIFS : RÉSULTATS DE BASE

10.1. Motifs de Chow

Nous nous dispenserons de rappeler en détail la construction de la catégorie des motifs de Chow, celle-ci étant maintenant bien connue et ayant fait l'objet d'excellentes expositions, y compris dans ce Séminaire [47, 14, 56, 1]. Rappelons seulement que :

1. On part de la catégorie des F -variétés projectives lisses.
2. On « agrandit » celle-ci en gardant les mêmes objets mais en prenant comme morphismes les *correspondances de Chow* : Si X est purement de dimension d , les correspondances de X vers une autre variété Y sont les éléments du groupe de Chow $CH_d(X \times Y)$. La catégorie obtenue est additive. Elle est munie d'une structure monoïdale symétrique, induite par le produit des variétés, et bilinéaire par rapport à l'addition des correspondances (on dira qu'elle est *tensorielle*). La correspondance « graphe d'un morphisme » définit un foncteur (covariant, avec nos conventions) de la catégorie de (1) vers cette catégorie.
3. On agrandit la catégorie de (2) en adjoignant des noyaux aux endomorphismes idempotents (enveloppe pseudo-abélienne ou karoubienne). La catégorie obtenue est celle des *motifs de Chow effectifs*. Elle est encore tensorielle et notée $\text{Mot}^{\text{eff}}(F)$. Si X est une variété projective lisse, on note $h(X)$ son image dans $\text{Mot}^{\text{eff}}(F)$.
4. Dans $\text{Mot}^{\text{eff}}(F)$, on a une décomposition canonique $h(\mathbf{P}^1) = \mathbf{1} \oplus L$, où $\mathbf{1} = h(\text{Spec } F)$ et L est le *motif de Lefschetz*. On passe de $\text{Mot}^{\text{eff}}(F)$ à $\text{Mot}(F)$, catégorie des motifs de Chow, en inversant L pour la structure monoïdale. Si $M \in \text{Mot}(F)$ et $n \in \mathbf{Z}$, nous noterons ici

$$M(n) := M \otimes L^{\otimes n}.$$

La catégorie $\text{Mot}(F)$ est rigide, ce qui signifie qu'elle porte une dualité parfaite relativement à sa structure tensorielle : on notera M^\vee le dual d'un motif M ⁽⁷⁾. Dans les applications arithmético-géométriques des motifs, il est fréquent de tensoriser les groupes de morphismes par \mathbf{Q} : ici, au contraire, il est très important de les considérer à coefficients entiers. En fait nous aurons à considérer des variantes à coefficients finis : si p est un nombre premier, on note

$$\text{Mot}^{\text{eff}}(F, \mathbf{F}_p), \quad \text{Mot}(F, \mathbf{F}_p)$$

⁽⁷⁾On prendra garde au fait que, dans [62], Vishik utilise cette notation dans un sens différent : si N est un facteur direct du motif d'une quadrique de dimension d , ce qu'il note N^\vee correspond à ce que nous notons $N^\vee(d)$.

les catégories définies comme ci-dessus en prenant comme groupes de correspondances les groupes $CH_d(X \times Y)/p$. La catégorie $\text{Mot}(F, \mathbf{F}_p)$ est encore rigide.

NOTATION 10.1. — Pour $M \in \text{Mot}(F)$, on note $\delta(M) \in \mathbf{Z}$ la dimension de M (au sens des catégories rigides : c'est la trace de l'identité). De même pour $M \in \text{Mot}(F, \mathbf{F}_p)$; on a alors $\delta(M) \in \mathbf{F}_p$.

On sait que, si $M = h(X)$ pour une variété projective lisse X , $\delta(h(X))$ est la caractéristique d'Euler-Poincaré de X par rapport à une cohomologie de Weil quelconque.

10.2. Variétés cellulaires et motifs de Tate purs

Soit X une variété projective lisse de dimension d admettant une décomposition cellulaire : cela signifie que X admet une stratification par des espaces affines. Le motif de X a alors une description très simple : les groupes de Chow de X sont libres de type fini et

$$(10.1) \quad h(X) \simeq \bigoplus_{n=0}^d CH_n(X) \otimes L^n.$$

De plus, les accouplements $CH^n(X) \times CH^{d-n}(X) \rightarrow \mathbf{Z}$ donnés par le produit d'intersection sont parfaits. La première assertion (vraie sans supposer X projective ni lisse) se montre par dévissage et récurrence sur le nombre de cellules (c'est un cas très particulier de la proposition 10.7 ci-dessous) ; la seconde, qui revient à une formule de Künneth via le principe d'identité de Manin, se démontre de la même manière ; enfin la troisième se déduit de (10.1) par dualité (voir ci-dessous). Ainsi, $h(X)$ est somme directe de motifs de la forme L^n : on dit que $h(X)$ est un *motif de Tate pur*.

La sous-catégorie pleine des motifs de Tate purs est très simple : on a

$$(10.2) \quad \text{Hom}(L^m, L^n) = \begin{cases} \mathbf{Z} & \text{si } m = n \\ 0 & \text{si } m \neq n. \end{cases}$$

Ainsi, les Hom sont des groupes abéliens libres de type fini, et sont *invariants par extension des scalaires*. Il est également clair que

$$(10.3) \quad \delta(L^n) = 1 \text{ pour tout } n \text{ (cf. notation (10.1)).}$$

Plus généralement, soit X une variété projective lisse telle que $h(X)$ soit un motif de Tate pur. Alors les groupes de Chow de X sont des \mathbf{Z} -modules libres de type fini, comme le montre la formule

$$\text{Hom}(L^n, h(X)) = CH_n(X).$$

Ces groupes sont en dualité parfaite, comme il résulte des formules (10.2) et de la dualité sur $h(X)$. La formule (10.1) reste vraie.

Si F est séparablement clos et que l est un nombre premier différent de la caractéristique de F , on a $H_{\text{ét}}^j(X, \mathbf{Z}_l) = 0$ pour j impair et la classe de cycle $CH^i(X) \otimes \mathbf{Z}_l \rightarrow H_{\text{ét}}^{2i}(X, \mathbf{Z}_l(i))$ est bijective : c'est évident en considérant le foncteur de réalisation

l -adique. Les groupes de Chow forment donc (pour une telle X) une « cohomologie de Weil » à coefficients entiers. De plus, l'équivalence rationnelle coïncide avec l'équivalence numérique.

On peut pousser plus loin l'analogie : on a des isomorphismes de Künneth pour toute autre variété projective lisse Y (évidents en considérant une fois de plus le motif de X) :

$$(10.4) \quad CH_n(X \times Y) \simeq \bigoplus_{i+j=n} CH_i(X) \otimes CH_j(Y).$$

En tenant compte de la dualité sur les groupes de Chow de X donnée ci-dessus, on obtient ainsi une description canonique de $\text{Hom}(h(X), h(Y))$:

$$(10.5) \quad \text{Hom}(h(X), h(Y)) \simeq \prod_{i \geq 0} \text{Hom}(CH_i(X), CH_i(Y)).$$

10.3. Motifs géométriquement de Tate purs

DÉFINITION 10.2. — *Un motif $M \in \text{Mot}(F)$ est géométriquement de Tate pur si $\bar{M} \in \text{Mot}(\bar{F})$ est un motif de Tate pur. On note $\text{Mot}(F)_{\text{tate}}$ la sous-catégorie pleine de $\text{Mot}(F)$ formée des motifs géométriquement de Tate purs. Si p est un nombre premier, on définit de même la catégorie $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}}$.*

La catégorie $\text{Mot}(F)_{\text{tate}}$ est visiblement une sous-catégorie rigide de $\text{Mot}(F)$, stable par facteurs directs, de même que $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}}$ dans $\text{Mot}(F, \mathbf{F}_p)$.

De même que, dans le cas classique, on peut définir l'équivalence numérique dans $\text{Mot}(F, \mathbf{F}_p)$ et $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}}$: pour $M, N \in \text{Mot}(F, \mathbf{F}_p)$, introduisons

$$\mathcal{N}(M, N) = \{f : M \rightarrow N \mid \forall g : N \rightarrow M, \text{tr}(gf) = 0\}$$

où tr est la trace, donnée par la structure rigide de $\text{Mot}(F, \mathbf{F}_p)$: c'est un idéal monoïdal de cette catégorie (cf. [2, lemme 7.1.1]) et on définit $\text{Mot}_{\text{num}}(F, \mathbf{F}_p)$ comme l'enveloppe pseudo-abélienne de $\text{Mot}(F, \mathbf{F}_p)/\mathcal{N}$, et de même pour $\text{Mot}_{\text{num}}(F, \mathbf{F}_p)_{\text{tate}}$.

Donnons-nous un nombre premier p . Si F est séparablement clos, le foncteur canonique $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}} \rightarrow \text{Mot}_{\text{num}}(F, \mathbf{F}_p)_{\text{tate}}$ est une équivalence de catégories ($\mathcal{N} = 0$). On prendra garde au fait que le foncteur d'extension des scalaires

$$H : \text{Mot}(F, \mathbf{F}_p)_{\text{tate}} \rightarrow \text{Mot}(\bar{F}, \mathbf{F}_p)_{\text{tate}}$$

n'induit pas un foncteur de $\text{Mot}_{\text{num}}(F, \mathbf{F}_p)_{\text{tate}}$ vers $\text{Mot}_{\text{num}}(\bar{F}, \mathbf{F}_p)_{\text{tate}}$ (voir ci-dessous). L'analogie avec la situation classique est tentante : définissons l'équivalence homologique sur $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}}$ comme le noyau de H , et $\text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{tate}}$ comme l'enveloppe

pseudo-abélienne de $\text{Mot}(F, \mathbf{F}_p)_{\text{tate}} / \text{Ker}(H)$. On est donc dans la situation suivante :

$$(10.6) \quad \begin{array}{ccc} \text{Mot}(F, \mathbf{F}_p)_{\text{tate}} & \xrightarrow{H} & \text{Mot}(\bar{F}, \mathbf{F}_p)_{\text{tate}} \\ \Pi \downarrow & & \downarrow \iota \\ \text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{tate}} & \xrightarrow{\bar{H}} & \text{Mot}_{\text{num}}(\bar{F}, \mathbf{F}_p)_{\text{tate}} \\ \downarrow & & \\ \text{Mot}_{\text{num}}(F, \mathbf{F}_p)_{\text{tate}} & & \end{array}$$

où le foncteur \bar{H} est (par définition) fidèle. L’argument de Jannsen [27] (cf. aussi [1, prop. 2.6] ou [3, th. 1]) donne :

PROPOSITION 10.3. — *La catégorie $\text{Mot}_{\text{num}}(F, \mathbf{F}_p)_{\text{tate}}$ est abélienne semi-simple.* \square

Malheureusement, pour $M \in \text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{tate}}$, il n’est pas vrai en général que $\mathcal{N}(M, M)$ soit égal au radical de $\text{End}(M)$; c’est même loin d’être le cas :

PROPOSITION 10.4. — *Soit X une quadrique anisotrope. Alors l’image de $h(X)$ dans $\text{Mot}_{\text{num}}(F, \mathbf{F}_2)_{\text{tate}}$ est égale à 0.*

Démonstration. — Il suffit de voir que tout endomorphisme de $h(X)$ dans $\text{Mot}(F, \mathbf{F}_2)_{\text{tate}}$ est de trace nulle. Cela découle immédiatement du fait que tout 0-cycle sur $X \times X$ est de degré pair, ce qui résulte du théorème 3.2. \square

Ceci limite fortement l’intérêt de la proposition 10.3. Néanmoins, la proposition 10.4 a une conséquence intéressante :

LEMME 10.5. — *Soit N un facteur direct de $h(X)$, où X est une quadrique anisotrope. Alors $\delta(N)$ est pair (voir notation 10.1).*

En effet, $\delta(N) = \text{tr}(\pi_N)$, où $\pi_N \in \text{End}(h(X))$ est le projecteur définissant N ; le lemme résulte donc immédiatement de la proposition 10.4.

D’après (10.3), cela veut dire que *le nombre de motifs de Tate intervenant dans $H(N)$ est pair.*

10.4. Variétés projectives homogènes

Soit X une variété projective homogène sous un groupe réductif connexe G . Si G est déployé, X admet une décomposition cellulaire [12] et on est dans la situation du §10.2. C’est par exemple le cas si X est une *quadrique hyperbolique*. En général on est dans la situation du §10.3.

Si G n’est pas déployé mais que X a un point rationnel, on peut décomposer partiellement le motif de X . Par exemple, si X est une quadrique isotrope définie par la forme quadratique $q \simeq \mathbb{H} \perp q'$, on a

$$(10.7) \quad h(X) \simeq \mathbf{1} \oplus h(X')(1) \oplus L^d$$

où $d = \dim X$ et X' est la quadrique d'équation $q' = 0$. Ce résultat est dû à Rost [53]. Il montre que $h(X)$ « contient » l'indice de Witt $i = i(q)$: en itérant, on obtient une décomposition

$$(10.8) \quad h(X) \simeq \mathbf{1} \oplus L \oplus \cdots \oplus L^{i-1} \oplus h(Y)(i) \oplus L^{d-i+1} \oplus \cdots \oplus L^d$$

où Y est une quadrique dont l'équation est donnée par la partie anisotrope de q . De plus,

LEMME 10.6. — *$h(X)$ ne contient pas L^i en facteur direct.*

En effet

$$\mathrm{Hom}(L^i, h(X)) = \mathrm{Hom}(\mathbf{1}, h(Y)) = CH_0(Y)$$

et

$$\mathrm{Hom}(h(X), L^i) = \mathrm{Hom}(h(Y), \mathbf{1}) = CH^0(Y).$$

La composition des correspondances

$$\mathrm{Hom}(h(X), L^i) \times \mathrm{Hom}(L^i, h(X)) \rightarrow \mathrm{Hom}(L^i, L^i) = \mathrm{Hom}(\mathbf{1}, \mathbf{1})$$

correspond au produit d'intersection $CH^0(Y) \times CH_0(Y) \rightarrow \mathbf{Z}$. Or le théorème 3.2 montre que ce produit est d'image $2\mathbf{Z}$.

La décomposition (10.7) a été généralisée par Karpenko, puis par Chernousov-Gille-Merkurjev au cas d'une variété projective homogène quelconque ayant un point rationnel :

PROPOSITION 10.7 ([33, th. 6.5 et cor. 6.11]; [13, th. 7.1]). — *Soit X une variété projective lisse admettant une filtration*

$$\emptyset = X_{-1} \subset X_0 \subset \cdots \subset X_n = X$$

où les X_i sont fermés et où, pour tout $i \in [0, n]$, $X_i - X_{i-1}$ est fibré sur une variété projective lisse Y_i à fibres des espaces affines de dimension a_i . Alors on a un isomorphisme canonique dans $\mathrm{Mot}(F)$

$$h(X) \simeq \bigoplus_{i=0}^n h(Y_i)(a_i).$$

La manière la plus agréable (mais certainement pas la moins chère) de comprendre la démonstration de cette proposition est de se placer dans la catégorie triangulée des motifs géométriques $DM_{\mathrm{gm}}^{\mathrm{eff}}(F)$ de Voevodsky ([63], voir aussi l'exposé Bourbaki de Friedlander [18, §3]) : rappelons que dans cette catégorie, le motif $M(U)$ d'une variété lisse U est défini et que la catégorie $\mathrm{Mot}^{\mathrm{eff}}(F)$ s'y plonge de manière pleinement fidèle d'après [64]. Traitons le cas $n = 1$: c'est de toute façon le cas essentiel. On a le triangle exact de Gysin

$$M(X - X_0) \rightarrow M(X) \rightarrow M(X_0)(c)[2c] \xrightarrow{+}.$$

Soit $p : X - X_0 \rightarrow Y_1$ la projection de l'énoncé. Par invariance homotopique, le morphisme $M(p)$ est un isomorphisme. Le point est alors que l'adhérence dans $X \times Y_1$ du graphe de p fournit une correspondance de Chow qui scinde ce triangle exact. On obtient ensuite la proposition en dualisant.

Il en résulte :

THÉORÈME 10.8 ([13, th. 7.4]). — *Soit X une variété projective homogène ayant un point rationnel. Écrivons $X = G/P$, où G est semi-simple adjoint et P est un sous-groupe parabolique de G défini sur F (c'est toujours possible). Alors on a un isomorphisme canonique dans $\text{Mot}(F)$, de la forme*

$$h(X) = \bigoplus_{\delta \in \Delta} h(Z_\delta)(l(\delta))$$

où Δ est l'ensemble (fini) des coïnvariants de l'action de Galois sur $(P \backslash X)(\bar{F})$ et où $l(\delta), Z_\delta$ sont un certain entier ≥ 0 et une certaine variété projective homogène associés à δ .

Karpenko avait obtenu auparavant ce théorème dans le cas où G est classique [33]. En itérant, il en résulte que le motif de Chow d'une variété projective homogène est somme directe canonique de tordus à la Tate de motifs de Chow de variétés projectives homogènes anisotropes.

10.5. Le théorème de nilpotence de Rost

NOTATION 10.9. — *On note $\text{Mot}(F)_{\text{hmg}}$ la sous-catégorie épaisse (c'est-à-dire pleine, additive et stable par facteurs directs) de $\text{Mot}(F)$ engendrée par les $h(X)(n)$, où $n \in \mathbf{Z}$ et X est une variété projective homogène. On note de même $\text{Mot}(F, \mathbf{F}_p)_{\text{hmg}}$ la catégorie correspondante à coefficients \mathbf{F}_p .*

La catégorie $\text{Mot}(F)_{\text{hmg}}$ est visiblement stable par produit tensoriel et par dual ; d'après les remarques du début du §10.4, c'est une sous-catégorie pleine de $\text{Mot}(F)_{\text{tate}}$. Mêmes remarques à coefficients finis.

THÉORÈME 10.10. — *Pour $M, N \in \text{Mot}(F)_{\text{hmg}}$, notons $I(M, N)$ l'ensemble des homomorphismes f de M vers N tels que $f_{\bar{F}} = 0$. Alors, pour tout $M \in \text{Mot}(F)_{\text{hmg}}$, $I(M, M)$ est un idéal nilpotent de $\text{End}(M)$.*

Ce théorème est dû à Chernousov–Gille–Merkurjev [13, th. 8.2] ; dans le cas où M est le motif d'une quadrique, c'est le célèbre théorème de nilpotence de Rost. Pour le démontrer, on peut visiblement supposer que M est somme de motifs de la forme $h(X)(n)$ pour X projective homogène (cas que considèrent les auteurs). De plus, d'après un lemme de Nagata et Higman (cf. [2, lemme 7.2.8]), il suffit de montrer qu'il existe un entier d tel que $f^d = 0$ pour tout $f \in I(h(X), h(X))$. En utilisant le théorème 10.8, on se ramène facilement au théorème suivant :

THÉORÈME 10.11. — Soient X, Y deux variétés projectives lisses et $f \in \text{End } h(X)$. Supposons que f agisse trivialement sur $CH_*(X_{F(y)})$ pour tout point $y \in Y$. Alors $f^{\dim Y+1}$ agit trivialement sur $CH_*(Y \times X)$.

Ce théorème est dû à Brosnan [11, th. 3.1]; le cas $CH_{\dim Y}(Y \times X)$ avait été démontré par Rost. Soit $(F_n CH_*(Y \times X))_{n \geq 0}$ la filtration de $CH_*(Y \times X)$ par la dimension du support de Y . Si $(gr_n)_{n \geq 0}$ est le gradué associé, il suffit de montrer que f agit trivialement sur gr_n pour tout n . Brosnan le démontre en généralisant légèrement la composition classique des correspondances, autorisant trois variétés X_1, X_2, X_3 quelconques à condition que la variété intermédiaire X_2 soit projective lisse. Pour ce faire, il utilise la formule

$$f \circ g = (p_{13})_* \varphi^!(g \otimes f)$$

où φ est l'immersion régulière $X_1 \times X_2 \times X_3 \rightarrow X_1 \times X_2 \times X_2 \times X_3$ donnée par la diagonale de X_2 et $\varphi^!$ est le morphisme de Gysin correspondant [19, ch. 6]. Essentiellement, cela lui permet de faire opérer la correspondance f sur la suite spectrale de niveau qui aboutit à la filtration (F_n) ⁽⁸⁾. Rost, quant à lui, utilisait une suite spectrale obtenue à l'aide de sa théorie des modules de cycles.

Remarques 10.12. — a) Comme le groupe $\text{End}(M_{\bar{F}})$ est libre de type fini, $I(M)$ est aussi le sous-groupe de torsion de $\text{End}(M)$ comme le montre un argument de transfert bien connu.

b) Les théorèmes 10.8 et 10.10 s'étendent aux motifs à coefficients dans \mathbf{F}_p , avec les mêmes démonstrations.

c) J'ignore si le théorème 10.10 s'étend à tous les objets de $\text{Mot}(F)_{\text{tate}}$. Même perplexité à coefficients finis.

Puisque les Hom dans $\text{Mot}(\bar{F}, \mathbf{F}_p)_{\text{tate}}$ sont de dimension finie, la remarque 10.12 b) entraîne :

COROLLAIRE 10.13. — Soit p un nombre premier. Alors la catégorie $\text{Mot}(F, \mathbf{F}_p)_{\text{hmg}}$ est de Wedderburn au sens de [2, I.2] : pour tout objet M , l'anneau $\text{End}(M)$ est extension d'une \mathbf{F}_p -algèbre semi-simple par un idéal nilpotent. En particulier le théorème de Krull-Schmidt est vrai : tout objet est somme directe d'un nombre fini d'objets indécomposables, uniques à isomorphisme près. \square

La preuve du corollaire suivant est standard :

COROLLAIRE 10.14 (cf. [13, cor. 8.3]). — Soit $M \in \text{Mot}(F)_{\text{hmg}}$, soit K/F une extension, et soit $(p_i) \in \text{Im}(\text{End}(M) \rightarrow \text{End}(M_K))$ une famille de projecteurs orthogonaux de somme 1. Alors les p_i se relèvent en une famille de projecteurs orthogonaux de somme 1 de $\text{End}(M)$, de manière unique à conjugaison près. Cet énoncé reste vrai à coefficients finis.

⁽⁸⁾Sa formulation est plus élémentaire.

On aimerait pouvoir également démontrer que le foncteur d'extension des scalaires $\text{Mot}(F) \rightarrow \text{Mot}(\bar{F})$ est conservatif. Il faut faire un peu attention car ce foncteur n'est pas plein. Le problème est de montrer que, si f est un morphisme tel que f_K soit un isomorphisme, l'inverse de f_K est défini sur F : on s'en tire essentiellement quand on sait ceci a priori ou quand la source et le but de f sont égaux. Ce problème est de même nature que pour la conjecture standard B. Cela donne l'énoncé un peu désagréable suivant.

COROLLAIRE 10.15. — a) Soient $M, N \in \text{Mot}(F)_{\text{hmg}}$, K/F une extension, et $f : M \rightarrow N$, $g : N \rightarrow M$ deux morphismes tels que $(gf)_K$ soit un isomorphisme. Alors gf est un isomorphisme. Si $(fg)_K$ est également un isomorphisme, alors f et g sont des isomorphismes.

b) Si N est indécomposable, il suffit de demander que $(gf)_K$ soit un isomorphisme pour que f et g soient des isomorphismes.

c) [13, cor. 8.4] Soient X, Y deux variétés projectives homogènes, K/F une extension et $f \in \text{Hom}(h(X), h(Y))$. Si f_K est un isomorphisme, alors f est un isomorphisme.

Ces énoncés restent vrais à coefficients finis.

Démonstration. — a) On se ramène immédiatement au cas $M = N, g = 1$. Quitte à agrandir K , on peut supposer que M_K est un motif de Tate mixte. Alors $\text{End}(M_K)$ est produit d'algèbres de matrices M_i sur \mathbf{Z} ; en observant que le terme constant du polynôme caractéristique de l'image de f_K dans M_i est égal à ± 1 (c'est le déterminant), on voit que l'inverse de f_K est donné par un polynôme $Q(f_K)$. En considérant $fQ(f)$, on se ramène au cas où $f_K = 1$; alors f est unipotent, donc inversible.

b) En utilisant a), on se ramène au cas où $(gf)_K = 1$. Alors $(fg)_K$ est idempotent, donc égal à 0 ou 1 par hypothèse. Mais $(fg)_K = 0$ est impossible : cela entraînerait que $g_K = (gfg)_K = 0$, contredisant $(gf)_K = 1$.

c) On peut encore supposer que X et Y sont déployées sur K . En considérant le motif de Tate de poids maximal intervenant dans $h(X_K)$ et $h(Y_K)$, on remarque que nécessairement $\dim X = \dim Y$. Alors la transposée de f définit un morphisme $g : h(Y) \rightarrow h(X)$ et g_K est évidemment un isomorphisme. On est donc ramené au cas a). \square

Dans la veine de (10.6), la définition suivante clarifie bien les choses et sera utile plus loin :

DÉFINITION 10.16. — a) On note $\text{Mot}_{\text{hom}}(F)_{\text{hmg}}$ le quotient de la catégorie $\text{Mot}(F)_{\text{hmg}}$ par l'idéal I du théorème 10.10.

b) Soit p un nombre premier. On note $\text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{hmg}}$ l'image essentielle de $\text{Mot}(F, \mathbf{F}_p)_{\text{hmg}}$ dans $\text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{tate}}$ (cf. (10.6) et notation 10.9).

Par définition, les morphismes dans la catégorie $\text{Mot}_{\text{hom}}(F)_{\text{hmg}}$ (resp. $\text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{hmg}}$), entre des motifs de variétés $h(X)$ et $h(Y)$, disons, sont formés de l'image de $CH_{\dim X}(X \times Y)$ dans $CH_{\dim X}(\bar{X} \times \bar{Y})$ (resp. dans $CH_{\dim X}(\bar{X} \times \bar{Y})/p$). Le

théorème 10.10 et sa version modulo p montrent, comme ci-dessus, que les foncteurs $\text{Mot}(F)_{\text{hmg}} \rightarrow \text{Mot}_{\text{hom}}(F)_{\text{hmg}}$ et $\text{Mot}(F, \mathbf{F}_p)_{\text{hmg}} \rightarrow \text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{hmg}}$ sont conservatifs et que l'image d'un motif indécomposable est indécomposable. Le corollaire 10.15 et les remarques le précédant concernent la conservativité partielle (?) des foncteurs $\text{Mot}_{\text{hom}}(F)_{\text{hmg}} \rightarrow \text{Mot}_{\text{hom}}(\bar{F})_{\text{hmg}}$ et $\text{Mot}_{\text{hom}}(F, \mathbf{F}_p)_{\text{hmg}} \rightarrow \text{Mot}_{\text{hom}}(\bar{F}, \mathbf{F}_p)_{\text{hmg}}$.

10.6. La multiplicité d'une correspondance

Soit $\alpha : h(X) \rightarrow h(Y)$ une correspondance entre variétés projectives lisses. Il existe un unique entier $\mu(\alpha)$ tel que $(p_Y)_* \circ \alpha = \mu(\alpha)(p_X)_*$, où p_X et p_Y sont les morphismes structuraux : c'est la *multiplicité de α* . Cette fonction a les propriétés évidentes suivantes :

1. $\mu(\alpha + \beta) = \mu(\alpha) + \mu(\beta)$.
2. $\mu(\alpha \circ \beta) = \mu(\alpha)\mu(\beta)$.
3. $\mu(\alpha \otimes \beta) = \mu(\alpha)\mu(\beta)$.
4. Si α est le graphe d'une application rationnelle, $\mu(\alpha) = 1$.

En particulier, si $X = Y$ et que π est une correspondance idempotente, on a $\mu(\pi) = 0$ ou 1 ; le second cas se produit si et seulement si la composition

$$N \rightarrow h(X) \rightarrow \mathbf{1},$$

où N est le facteur direct défini par π , est non triviale.

On peut aussi définir la multiplicité d'une correspondance $\alpha \in CH_{\dim X}(X \times Y)$, où X et Y sont des variétés quelconques avec Y propre : c'est l'entier $\mu(\alpha)$ tel que $p_*\alpha = \mu(\alpha)[X] \in CH_{\dim X}(X)$, où p est la projection $X \times Y \rightarrow Y$. Elle coïncide avec la précédente dans le cas projectif lisse.

10.7. Opérations de Steenrod sur les groupes de Chow

Dans [65], Voevodsky définit des opérations de Steenrod en cohomologie motivique modulo p : pour $p = 2$, ces opérations jouent un rôle essentiel dans sa preuve de la conjecture de Milnor ([66], voir aussi [30, §6]). En comptant les degrés, on observe que les plus importantes d'entre elles préservent les groupes de Chow modulo p

$$CH^i(X)/p = H^{2i}(X, \mathbf{Z}/p(i))$$

pour X une F -variété lisse. Il est tentant d'essayer de les définir directement dans ce cadre, en évitant la théorie homotopique des schémas : cela correspond d'ailleurs à une question de Fulton [19, ex. 19.2.8]. C'est ce qu'a fait Brosnan dans sa thèse [10]⁽⁹⁾. Sa construction et les propriétés principales de ces opérations (pour $p = 2$) sont reformulées lucidement et succinctement par Karpenko dans [36, §2], suivant des remarques de Merkurjev :

⁽⁹⁾Brosnan ne démontre pas que ses opérations coïncident avec celles de Voevodsky.

Au moins pour les variétés quasi projectives lisses X , Brosnan suit exactement la construction de Steenrod, en utilisant les groupes de Chow équivariants définis par Edidin et Graham [16] (voir aussi Totaro [59]). Il obtient ainsi des opérations

$$S^i : CH^n(X)/2 \rightarrow CH^{n+i}(X)/2$$

(en cohomologie modulo 2, S^i correspondrait à l'opération de Steenrod Sq^{2i}). Ces opérations ont les propriétés suivantes :

1. L'opération de Steenrod totale $S = \sum S^i$ est un endomorphisme de l'anneau $CH^*(X)/2$.
2. S est contravariant pour les morphismes quelconques. Compte tenu de (1), cela fournit la *formule de Cartan* pour les cross-produits de cycles.
3. Sur $CH^n(X)/2$, S^i est égal à
 - 0 pour $n < i$;
 - l'élévation au carré pour $n = i$.
4. $S^i = 0$ pour $i < 0$; S^0 est l'identité.
5. *Formule de Riemann-Roch* : si $f : Y \rightarrow X$ est un morphisme propre et $\alpha \in CH^*(Y)/2$, on a

$$f_*(S(\alpha) \cdot c(-T_Y)) = S(f_*\alpha) \cdot c(-T_X)$$

où T_X et T_Y sont respectivement les fibrés tangents de X et Y et c désigne la classe de Chern totale (les expressions $-T_X$ et $-T_Y$ ayant un sens dans $K_0(X)$ et $K_0(Y)$).

Dans le cas où f est une immersion fermée et où α est la classe de Y , la formule de Riemann-Roch se réduit à la *formule de Wu* :

$$(10.9) \quad S([Y]) = f_*c(N)$$

où N est le fibré normal de l'immersion f .

10.8. Le motif d'une quadrique déployée

Soit X une quadrique déployée de dimension d et d'équation $q = 0$. L'anneau de Chow de X admet une description très simple : on a

$$(10.10) \quad CH^i(X) = \begin{cases} \mathbf{Z}h^i & \text{si } i < d/2 \\ \mathbf{Z}l_{d-i} & \text{si } i > d/2 \\ \mathbf{Z}l^1 \oplus \mathbf{Z}l^2 & \text{si } i = d/2 \end{cases}$$

où h est la classe d'une section hyperplane de X et, si $j < d/2$, l_j désigne la classe d'un sous-espace projectif de X de dimension j . Pour $j = d/2$ (donc d pair), on a deux telles familles de sous-espaces l^1 et l^2 , qui sont permutées par l'action de $O(q)$. Plus précisément :

LEMME 10.17. — a) Soit $u \in O(q)$. Alors u opère trivialement sur $CH^i(X)$ si $i \neq d/2$. Si $i = d/2$, u opère trivialement sur $CH^i(X)$ si $u \in SO(q)$ et échange l^1 et l^2 si $u \notin SO(q)$.

b) L'application naturelle $O(q) \rightarrow \text{End}(h(X))$ est triviale si d est impair et se factorise (non trivialement) à travers le déterminant si d est pair.

(Il suffit de tester a) sur les réflexions puisque celles-ci engendrent $O(q)$, ce qui est facile; b) résulte de a) puisque $\text{End}(h(X))$ opère fidèlement sur les $CH^i(X)$, cf. (10.5) ci-dessous.)

On a de plus les relations (cf. par exemple [32])

$$(10.11) \quad \begin{aligned} h^i &= 2l_{d-i} & (i > d/2) \\ h^{d/2} &= l^1 + l^2 & \text{si } d \text{ est pair} \\ \langle h^i, l_{d-i} \rangle &= 1 & (i < d/2) \\ \langle l^1, l^2 \rangle &= \begin{cases} 0 & \text{si } d \equiv 0 \pmod{4} \\ 1 & \text{si } d \equiv 2 \pmod{4} \end{cases} \\ \langle l^a, l^a \rangle &= \begin{cases} 1 & \text{si } d \equiv 0 \pmod{4} \\ 0 & \text{si } d \equiv 2 \pmod{4} \end{cases} & \text{pour } a = 1, 2. \end{aligned}$$

Ceci permet de donner une formule explicite pour les « projecteurs de Künneth » π_i de X selon la décomposition (10.4) (π_i projette sur $CH_i(X)$) :

$$(10.12) \quad \pi_i = \begin{cases} l_i \times h^i & \text{si } i < d/2 \\ h^{d-i} \times l_{d-i} & \text{si } i > d/2 \\ l^1 \times l^1 + l^2 \times l^2 & \text{si } i = d/2 \text{ et } d \equiv 0 \pmod{4} \\ l^1 \times l^2 + l^2 \times l^1 & \text{si } i = d/2 \text{ et } d \equiv 2 \pmod{4}. \end{cases}$$

Bien entendu, tous ces calculs restent valables modulo 2.

Remarque 10.18. — (10.10) et (10.11) sont des cas particuliers de [51, prop. 1], qui permettrait de généraliser la formule (10.12) à une variété projective homogène déployée quelconque.

Les opérations de Steenrod opèrent de la manière suivante sur les images de h et des l_i dans les groupes de Chow modulo 2 :

$$(10.13) \quad \begin{aligned} S(h^i) &= h^i(1+h)^i & (i \geq 0) \\ S(l_{d-i}) &= l_{d-i}(1+h)^{i+1} & (i \geq d/2). \end{aligned}$$

(La première formule se réduit au cas $i = 1$ par la propriété (1) du §10.7; elle est alors évidente compte tenu de (3) et (4). Quant à la deuxième formule, elle est démontrée par exemple par Karpenko dans [36, cor. 3.3] au moyen de la formule de Wu (10.9). Noter également que $h^i = 0$ pour $i > d/2$ d'après les relations (10.11), de sorte que la contradiction entre les deux formules pour $i = d/2$ quand d est pair n'est qu'apparente!)

11. FORMES QUADRATIQUES ET MOTIFS : THÉORIES DE ROST ET DE VISHIK

Dans cette section, nous exposons les résultats de Rost et Vishik sur la structure du motif d'une quadrique. Ils reposent sur le théorème de nilpotence de Rost (théorème 10.10), que Rost a démontré pour calculer le motif d'une quadrique de Pfister (théorème 11.14). Ce calcul a ensuite été vastement généralisé par Vishik dans sa thèse [60]. Vishik y travaille apparemment dans la catégorie $DM_{-}^{\text{eff}}(F)$ des motifs triangulés de Voevodsky, mais l'article [62], sur lequel nous nous reposons, clarifie les choses et montre que tout se passe en fait dans la catégorie des motifs purs et résulte de calculs remarquablement élémentaires (à l'exception d'un résultat, le théorème 11.31 qui utilise les opérations de Steenrod motiviques).

11.1. Facteurs directs du motif d'une quadrique, d'après Vishik

Soit X une quadrique anisotrope de dimension d . Notons

- $h(X)$ son motif dans $\text{Mot}(F)_{\text{hmg}}$;
- $h(X, \mathbf{F}_2)$ son motif dans $\text{Mot}(F, \mathbf{F}_2)_{\text{hmg}}$;
- $h_{\text{hom}}(X)$ son motif dans $\text{Mot}_{\text{hom}}(F)_{\text{hmg}}$;
- $h_{\text{hom}}(X, \mathbf{F}_2)$ son motif dans $\text{Mot}_{\text{hom}}(F, \mathbf{F}_2)_{\text{hmg}}$.

On notera aussi $\bar{X} = X \times_F \bar{F}$ et on ne distinguera pas $h(\bar{X})$ et $h_{\text{hom}}(\bar{X})$, ni $h(\bar{X}, \mathbf{F}_2)$ et $h_{\text{hom}}(\bar{X}, \mathbf{F}_2)$.

D'après le corollaire 10.14, les facteurs directs de $h(X)$ (*resp.* de $h(X, \mathbf{F}_2)$) sont en correspondance bijective, à isomorphisme près, avec ceux de $h_{\text{hom}}(X)$ (*resp.* de $h_{\text{hom}}(X, \mathbf{F}_2)$). Vishik démontre de plus que les facteurs directs de $h_{\text{hom}}(X)$ et de $h_{\text{hom}}(X, \mathbf{F}_2)$ sont aussi en correspondance bijective. Son raisonnement pour relever des projecteurs modulo 2 en des projecteurs entiers est assez délicat dans le cas où d est pair : nous en donnons l'essentiel en 11.1.2. En réalité, toutes les applications aux formes quadratiques utilisent les groupes de Chow modulo 2, et même les groupes Hom dans $\text{Mot}_{\text{hom}}(F, \mathbf{F}_2)_{\text{hmg}}$: la distinction entre le cas entier et le cas modulo 2 n'a donc pas une grande importance en pratique.

11.1.1. Le cas de dimension impaire. — Si d est impair, la situation est relativement simple : pour tout $i \in [0, d]$, la correspondance $2\pi_i$ (*cf.* (10.12)) est rationnelle sur F , représentée par le cycle $h^{d-i} \times h^i$. En particulier, $\text{End}(h_{\text{hom}}(X)) = \text{Im}(\text{End}(h(X)) \rightarrow \text{End}(h(\bar{X})))$ contient $2 \text{End}(h(\bar{X}))$. Pour comprendre cette image on peut donc réduire modulo 2 ; d'après (10.5), on a un isomorphisme d'anneaux

$$\text{End}(h(\bar{X}, \mathbf{F}_2)) \simeq \prod_{i=0}^d \mathbf{F}_2.$$

L'image $\text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ de $\text{End}(h_{\text{hom}}(X))$ dans cet anneau correspond donc à une *partition* P de $\{0, \dots, d\}$, et tout idempotent de $\text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ se relève en un idempotent de $\text{End}(h_{\text{hom}}(X))$, unique puisque cette algèbre est commutative.

Explicitement, soit I une partie de $\{0, \dots, d\}$ telle que la correspondance $\pi_I := \sum_{i \in I} \pi_i$ soit définie sur F : alors les I minimaux forment la partition en question. En appliquant le corollaire 10.14, on obtient que les $(\pi_I)_{I \in P}$ se relèvent en des projecteurs orthogonaux $\tilde{\pi}_I$ de somme 1 dans $\text{End}(h(X))$, de manière unique à conjugaison près. En particulier, les $\tilde{\pi}_I$ définissent une décomposition

$$h(X) \simeq \bigoplus_{I \in P} N(I)$$

en somme directe de motifs indécomposables, unique à isomorphisme près. On a

$$N(I)_{\bar{F}} \simeq \bigoplus_{i \in I} L^{\otimes i}.$$

Vishik note $N \mapsto \Lambda(N)$ la bijection inverse de $I \mapsto N(I)$: il appelle $\Lambda(N)$ le *support* du facteur direct indécomposable N .

Énonçons une partie de ce qui précède :

PROPOSITION 11.1. — *Supposons d impair. Alors l'algèbre $\text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ est commutative et semi-simple. En particulier,*

- (i) *Tout facteur direct indécomposable de $h_{\text{hom}}(X, \mathbf{F}_2)$ apparaît avec multiplicité 1.*
- (ii) *Si N, N' sont deux facteurs directs indécomposables non isomorphes de $h_{\text{hom}}(X, \mathbf{F}_2)$, on a $\text{Hom}(N, N') = 0$. □*

11.1.2. *Le cas de dimension paire.* — Le premier résultat difficile de Vishik est que la description précédente s'étend (essentiellement) au cas pair.

THÉORÈME 11.2. — *Soit X une quadrique non hyperbolique de dimension paire d . Alors*

- a) *Le radical de $\text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ est engendré par $\theta := 1 - \tau$, où τ est (le graphe d') une réflexion quelconque.*
- b) *L'algèbre quotient est commutative.*
- c) *Tout système d'idempotents orthogonaux de somme 1 de $\text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ se relève dans $\text{End}(h_{\text{hom}}(X))$, de manière unique à conjugaison près.*

En particulier, tout facteur direct indécomposable de $h(X)$ apparaît avec multiplicité 1.

Ce théorème est (essentiellement) le contenu de [62, Sublemma 5.11]. Voici une démonstration de a) et b) : d'après (10.5), $\text{End}(h(X))$ est une sous-algèbre de

$$\prod_{i=0}^d \text{End}(CH^i(\bar{X})/2) = \prod_{i \neq d/2} \mathbf{F}_2 \times M_2(\mathbf{F}_2).$$

L'image de θ dans cette algèbre est nulle dans chaque facteur \mathbf{F}_2 , et vaut $\bar{\theta} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ dans $M_2(\mathbf{F}_2)$ (lemme 10.17).

Comme X n'est pas hyperbolique, tous les éléments de $CH^{d/2}(X)/2$ sont de degré pair (cf. remarques suivant (10.7)). En particulier, pour tout $\psi \in \text{End}(h(X))$, $\psi(h^{d/2})$ est de degré pair ; autrement dit, si $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est l'image de ψ dans $M_2(\mathbf{F}_2)$, on a $a+b+c+d = 0$. Notons A la sous-algèbre de $M_2(\mathbf{F}_2)$ définie par cette condition.

LEMME 11.3 (Vishik). — *Pour $x \in A$, soit x , soit $x + \bar{\theta}$ est idempotent.* □

Ce lemme entraîne que soit ψ , soit $\psi + \theta$ est idempotent. Pour conclure la preuve de a) et b), il suffit de montrer que θ est dans le radical de $\text{End}(h(X))$: mais il est clair que $\bar{\theta}$ est de carré nul et que son produit avec toute matrice de A est un multiple de θ .

Quant à c), il n'apparaît pas sous cette forme dans [62], mais voici une manière de le déduire des calculs qui y sont faits. Tout d'abord, l'énoncé de c) est vrai pour l'homomorphisme $\text{End}(h(\bar{X})) \rightarrow \text{End}(h(\bar{X}, \mathbf{F}_2))$ (vérification facile) ; ceci implique c) au cas où l'on a

$$(11.1) \quad 2 \text{End}(h(\bar{X})) \subset \text{End}(h_{\text{hom}}(X)).$$

LEMME 11.4. — (11.1) est vrai dans les cas suivants :

- (i) $\text{End}(h_{\text{hom}}(X))$ contient un élément ψ tel que $\text{tr}(\psi | CH^{d/2}(\bar{X}))$ soit impair.
- (ii) Le groupe de Galois absolu G_F opère non trivialement sur $CH^*(\bar{X})$ (c'est-à-dire sur $CH^{d/2}(\bar{X})$).

Démonstration. — (i) est [62, sublemma 5.7] : nous renvoyons à *loc. cit.* pour la démonstration, très technique. (ii) Si l'action de G_F n'est pas triviale, G_F permute l^1 et l^2 , donc opère comme $O(q)$; alors $\text{End}_{G_F}(CH^{d/2}(\bar{X})) = \mathbf{Z} \oplus \mathbf{Z}\theta$ et (11.1) est vérifié. □

Déduisons-en c) : le cas intéressant est celui où la condition du lemme 11.4 n'est pas vérifiée. Soit $\varepsilon \in \text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ un idempotent. L'hypothèse implique que la trace de ε opérant sur $CH^{d/2}(\bar{X})/2$ est égale à 0. Donc ε opère sur ce groupe comme 0 ou l'identité. Quitte à le remplacer par $1 - \varepsilon$, on peut supposer qu'il opère trivialement. Or X est déployée par une extension multiquadratique de F : il existe donc un entier s tel que $2^s \text{End}(h(\bar{X}))^{G_F} = 2^s \text{End}(h(\bar{X})) \subset \text{End}(h_{\text{hom}}(X))$.

Soit A_s l'image de $\text{End}(h_{\text{hom}}(X))$ dans $\text{End}(h(\bar{X}))/2^s$. Comme le noyau de $A_s \rightarrow \text{End}(h_{\text{hom}}(X, \mathbf{F}_2))$ est nilpotent, ε se relève en un idempotent $\varepsilon_s \in A_s$, dont l'action sur $CH^{d/2}(\bar{X})/2^s$ est nécessairement triviale. Par conséquent, l'image de ε_s dans $\text{End}(h(\bar{X}))/2^s$ se relève en un idempotent ε de $\text{End}(h(\bar{X}))$, et on a automatiquement $\varepsilon \in \text{End}(h_{\text{hom}}(X))$. □

Pour compléter cette description, il faut encore étudier les homomorphismes entre facteurs directs indécomposables. Notons $\pi_{d/2}^1$ et $\pi_{d/2}^2$ les deux idempotents de A de

somme 1 (voir la définition de A juste avant le lemme 11.3)

$$\pi_{d/2}^1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \pi_{d/2}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

On a

$$\pi_{d/2}^1 \bar{\theta} = 0, \quad \pi_{d/2}^2 \bar{\theta} = \bar{\theta}, \quad \bar{\theta} \pi_{d/2}^1 = \bar{\theta}, \quad \bar{\theta} \pi_{d/2}^2 = 0.$$

Deux cas se présentent donc :

PROPOSITION 11.5. — a) *Supposons que, dans $\text{End}(h(X))/\langle\theta\rangle$, un idempotent indécomposable ε contienne $\pi_{d/2}^1 + \pi_{d/2}^2$. Alors, si N_ε est le facteur direct correspondant de $h(X)$, on a $\text{End}(N_\varepsilon) = \mathbf{F}_2 \oplus \mathbf{F}_2 \varepsilon \theta \varepsilon$. Pour tout autre facteur direct indécomposable N , on a $\text{End}(N) = \mathbf{F}_2$, et $\text{Hom}(N, N') = 0$ si N et N' sont deux facteurs directs indécomposables non isomorphes.*

b) *Supposons que, dans $\text{End}(h(X))/\langle\theta\rangle$, un idempotent indécomposable ε_1 contienne $\pi_{d/2}^1$ et qu'un autre idempotent indécomposable ε_2 contienne $\pi_{d/2}^2$. Soient N_1 et N_2 les facteurs directs de $h(X)$ correspondants. Alors, pour tout facteur direct indécomposable N de $h(X)$, on a $\text{End}(N) = \mathbf{F}_2$. Si N, N' sont deux facteurs directs indécomposables non isomorphes, on a $\text{Hom}(N, N') = 0$, sauf si $(N, N') = (N_1, N_2)$ auquel cas $\text{Hom}(N_1, N_2) = \pi_{d/2}^2 \theta \pi_{d/2}^1 \mathbf{F}_2 \neq 0$. \square*

Remarque 11.6. — Comme me l'a fait observer Vishik, il est facile de donner un exemple de deux facteurs directs indécomposables M, N de motifs de quadriques (différentes) tels que $\dim \text{Hom}(M, N) \geq 3$ dans $\text{Mot}_{\text{hom}}(F, \mathbf{F}_2)_{\text{hmg}}$: prenons une forme quadratique q de dimension $d + 2$ avec $d \geq 5$, de quadrique associée X , telle que $N = h_{\text{hom}}(X, \mathbf{F}_2)$ soit indécomposable (par exemple q générique). On voit facilement que $\text{Im}(CH^2(X \times X) \rightarrow CH^2(\bar{X} \times \bar{X})/2)$ a pour base $(h^2 \times 1, h \times h, 1 \times h^2)$ où h est une section hyperplane. Mais

$$\text{Hom}(N(d-2), N) = \text{Hom}(N(d-2), N^\vee(d)) = \text{Hom}(N \otimes N, L^2)$$

est égal à ce groupe. (Noter que $N(d-2)$ est facteur direct de $h(Y)$, où Y est la quadrique définie par $q \perp (d-2)\mathbb{H}$, cf. (10.8).)

11.1.3. *Les diagrammes de Vishik.* — On a vu en 11.1.1 que, si la dimension d de la quadrique X est impaire, les facteurs directs indécomposables de $h(X)$ sont en bijection avec une certaine partition de l'ensemble des motifs de Tate intervenant dans $H(h(X))$, qui peut être identifié à $\{0, \dots, d\}$. Lorsque d est pair, le même énoncé est vrai d'après 11.1.2, mais on ne peut plus identifier cet ensemble à un ensemble d'entiers puisque $L^{d/2}$ intervient avec multiplicité 2. En fait, comme le remarque Vishik, ces deux facteurs sont dissymétriques. Plus précisément, il fait le choix suivant :

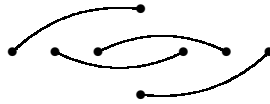
$$\begin{aligned} \pi^{up} &= l^1 \times (l^1 + l^2) \\ \pi_{lo} &= \pi_{d/2} - \pi^{up} \quad (\text{cf. (10.12)}). \end{aligned}$$

(Ainsi, π^{up} opère comme $\pi_{d/2}^2$ sur $CH^2(\bar{X})/2$ si $d \equiv 2 \pmod{4}$ et comme $\pi_{d/2}^2 + \bar{\theta}$ si $d \equiv 0 \pmod{4}$.) Il note L^{up} et L_{lo} les deux facteurs directs correspondants : la restriction

de $\overline{\text{deg}}_X$ à $CH_{d/2}(L^{up})$ (cf. 11.1.4) est nulle tandis que sa restriction à $CH_{d/2}(L_{lo})$ est non nulle. Il note $\Lambda(X)$ l'ensemble de ces motifs de Tate : ainsi, pour tout facteur direct M de $h(X)$, on peut identifier l'ensemble $\Lambda(M)$ des motifs de Tate intervenant dans $H(X)$ à un sous-ensemble de $\Lambda(X)$, et M est déterminé par $\Lambda(M)$ à isomorphisme près. Énonçons ceci explicitement, pour référence ultérieure :

PROPOSITION 11.7. — *Soient N, N' deux facteurs directs du motif d'une même quadrique. Si $\Lambda(N) \cap \Lambda(N') \neq \emptyset$, alors $N \simeq N'$.*

Pour représenter la décomposition de $h(X)$ en facteurs directs indécomposables, Vishik dessine des diagrammes fondés sur cette description. Ainsi, voici un diagramme représentant la décomposition motivique d'une quadrique définie par une 3-forme de Pfister :



11.1.4. *Le caractère de Vishik.* — Vishik a introduit un critère très pratique pour démontrer un isomorphisme entre motifs indécomposables :

DÉFINITION 11.8. — *Soit X une quadrique. On note $\overline{\text{deg}}_X$ l'homomorphisme*

$$\overline{\text{deg}}_X : CH_*(\bar{X})/2 \rightarrow \mathbf{F}_2$$

prenant la valeur 1 sur tous les générateurs apparaissant dans (10.10). C'est le caractère de Vishik.

On a le lemme trivial suivant :

LEMME 11.9. — *Si $\dim X$ est paire, $\overline{\text{deg}}_X \circ \theta = 0$, où θ est comme dans le théorème 11.2.*

On en déduit :

THÉORÈME 11.10 ([62, th. 3.8]). — *Soient X_1, X_2 deux quadriques et a_1, a_2 deux entiers ≥ 0 . Soient $\alpha : h(X_1)(a_1) \rightarrow h(X_2)(a_2)$ et $\beta : h(X_2)(a_2) \rightarrow h(X_1)(a_1)$ deux morphismes de $\text{Mot}(F, \mathbf{F}_2)$, et soit $r \geq 0$ tels que $(\overline{\text{deg}}_{X_1} \circ \beta \circ \alpha)_{|CH_r(\bar{X}_1)/2} \neq 0$. Alors il existe des facteurs directs indécomposables N_1 et N_2 de $h(X_1)(a_1)$ et $h(X_2)(a_2)$, isomorphes, tels que $L^r \in \Lambda(N_1)$ et $L^r \in \Lambda(N_2)$.*

Démonstration. — Tout d'abord, on peut se ramener à $a_1 = a_2 = 0$ en remplaçant X_1 et X_2 par X'_1 et X'_2 , où X'_i est d'équation $a_i \mathbb{H} \perp q_i$ si X_i est d'équation q_i (cf. (10.8)). Choisissons maintenant des décompositions en somme de facteurs directs indécomposables

$$h(X_1) \simeq \bigoplus_{s \in S} N_s^1, \quad h(X_2) \simeq \bigoplus_{t \in T} N_t^2.$$

Travaillons dans $\text{Mot}_{\text{hom}}(F, \mathbf{F}_2)_{\text{hmg}}$. Sur les décompositions ci-dessus, les morphismes $\Pi(\alpha)$ et $\Pi(\beta)$ ont des écritures matricielles (α_{st}) et (β_{ts}) (voir (10.6) pour se rappeler la définition des foncteurs H et Π). De même, $\Pi(\beta \circ \alpha)$ a une écriture matricielle $((\beta \circ \alpha)_{ss'})$.

De plus, la proposition 11.5 implique que, modulo θ si $\dim X$ est paire, cette matrice est diagonale. L’hypothèse et le lemme 11.9 impliquent donc qu’il existe s_0 tel que

$$(\beta \circ \alpha)_{s_0 s_0} = \sum_{t \in T} \beta_{ts_0} \alpha_{s_0 t} \equiv 1 \pmod{\langle \theta \rangle}.$$

Par conséquent, il existe aussi un t_0 tel que $\beta_{t_0 s_0} \alpha_{s_0 t_0} \equiv 1 \pmod{\langle \theta \rangle}$. Soient $N_1 = N_{s_0}^1$ et $N_2 = N_{t_0}^2$: le corollaire 10.15 b) implique que $N_1 \simeq N_2$. De plus, l’hypothèse implique immédiatement que $L^r \in \Lambda(N_1)$ et $L^r \in \Lambda(N_2)$. \square

11.2. Premières applications

DÉFINITION 11.11. — *Pour toute quadrique anisotrope X , on note $N_0(X)$ l’unique facteur direct indécomposable de $h(X)$ tel que $\mathbf{1} \in \Lambda(N_0(X))$: c’est le motif initial de X .*

THÉORÈME 11.12. — *Soient X, Y deux quadriques anisotropes. Alors $X \approx Y \iff N_0(X) \simeq N_0(Y)$.*

L’implication \Rightarrow est le corollaire 3.9 de [62]. Voici la démonstration de Vishik : choisissons un point rationnel $p \in Y(F(X))$ et un point rationnel $q \in X(F(Y))$. Alors p et q correspondent à des applications rationnelles $p : X \rightsquigarrow Y$ et $q : Y \rightsquigarrow X$. Soient $\alpha : h(X) \rightarrow h(Y)$ et $\beta : h(Y) \rightarrow h(X)$ les correspondances données respectivement par (l’adhérence du) graphe de p et de q . Il est immédiat que la condition du théorème 11.10 est vérifiée avec $r = 0$.

Je n’ai pas trouvé l’implication \Leftarrow dans [62], mais sa démonstration est facile : soit $\varphi : N_0(X) \rightarrow N_0(Y)$ un isomorphisme. Il induit un morphisme

$$h(X) \rightarrow N_0(X) \xrightarrow{\varphi} N_0(Y) \rightarrow h(Y)$$

qui induit évidemment un isomorphisme $CH_0(\bar{X}) \xrightarrow{\sim} CH_0(\bar{Y})$. Par le théorème 3.2, tout point rationnel de X (sur une extension de F) fournit un point rationnel de Y , donc $X \preccurlyeq Y$, et de même $Y \preccurlyeq X$. \square

THÉORÈME 11.13. — *Soit X une quadrique anisotrope. Alors $h(X)$ contient*

$$\bigoplus_{i=0}^{i_1(X)-1} N_0(X)(i)$$

en facteur direct (cf. la définition 6.1 pour se rappeler la définition de $i_1(X)$).

C’est le corollaire 3.10 de [62]. Vishik le démontre ainsi : comme les $N_0(X)(i)$ sont visiblement deux à deux non isomorphes, il suffit de montrer que chacun est facteur direct de $h(X)$. Fixons $i < i_1(X)$. Dans $X_{F(X)}$, choisissons un sous-espace projectif de dimension i . Son adhérence dans $X \times X$ définit une correspondance $\alpha : h(X)(i) \rightarrow h(X)$. D’autre part, considérons une section plane de codimension i de X , et plongeons-la diagonalement dans $X \times X$: cela définit une correspondance $\beta : h(X) \rightarrow h(X)(i)$. Il est alors facile de voir que la condition du théorème 11.10 est vérifiée avec $r = i$.

11.3. Le motif de Rost

THÉORÈME 11.14. — Soit φ une n -forme de Pfister anisotrope, et soit $X = X_\varphi$ la quadrique associée. Alors il existe un unique motif $M = M_\varphi$ tel que

- (i) $M_{\bar{F}} \simeq \mathbf{1} \oplus L^{\otimes(2^{n-1}-1)}$;
- (ii) $h(X) \simeq \bigoplus_{i=0}^{2^{n-1}-1} M(i)$.

Ce théorème est dû à Rost [53]. Le motif M_φ est appelé le *motif de Rost* associé à φ : il joue un rôle clé dans la démonstration par Voevodsky de la conjecture de Milnor (cf. [30, §8.1]).

Voici comment Vishik le déduit du théorème 11.13 : nous allons voir que $M_\varphi = N_0(X_\varphi)$. Partons de ce dernier motif (noté N pour simplifier). Comme $i_1(X_\varphi) = 2^{n-1}$, $h(X_\varphi)$ contient

$$M = \bigoplus_{i=0}^{2^{n-1}-1} N(i)$$

en facteur direct. D'autre part, le lemme 10.5 implique que $H(N)$ contient au moins deux motifs de Tate. En comptant, on voit successivement que

- le nombre de motifs de Tate intervenant dans M est \geq le nombre de motifs de Tate intervenant dans $h(X)$;
- $M = h(X)$;
- $H(N)$ contient exactement deux motifs de Tate ;
- le motif de Tate $\neq \mathbf{1}$ intervenant dans $H(N)$ est nécessairement $L^{2^{n-1}-1}$. \square

On a le complément suivant :

LEMME 11.15. — Posons $d = 2^n - 2 = \dim X$. Alors le facteur $L^{d/2}$ contenu dans $\Lambda(M_\varphi)$ est L^{up} .

C'est le point clé du contenu de [62, §5.7] (démonstration de la proposition 4.8). Le raisonnement de Vishik est le suivant : $\Lambda(M_\varphi(d/2))$ contient L^d , donc $CH_d(H(M_\varphi(d/2))) = CH_d(\bar{X}) = CH^0(\bar{X})$, dont le générateur est évidemment défini sur le corps de base. Il en est donc de même pour le générateur de $CH_{d/2}(H(M_\varphi)) = CH_d(H(M_\varphi(d/2)))$. Comme X n'est pas hyperbolique, ce générateur est nécessairement $h^{d/2}$, et donc $\overline{\deg}_X(CH_{d/2}(H(M_\varphi))) = 0$.

Avant d'énoncer un corollaire, donnons une définition :

DÉFINITION 11.16. — Soit N un facteur direct de $h(X)$. On note :

- $a(N)$ le plus petit entier a tel que $L^a \in \Lambda(N)$.
- $b(N)$ le plus grand entier b tel que $L^b \in \Lambda(N)$.
- $t(N) = b(N) - a(N)$ (c'est la taille de N).

COROLLAIRE 11.17. — *Soit X une quadrique anisotrope de dimension d . Soit $a < i_1(X)$, et soit N un facteur direct indécomposable de $h(X)$ tel que $L^a \in \Lambda(N)$. Si $a < d/2$, on a $N \simeq N_0(X)(a)$. Si $a = d/2$, on a $N \simeq N_0(X)(a)$ ou $N \simeq N_0(X)$ selon que $L^a = L_{lo}$ ou que $L^a = L^{up}$.*

Cet énoncé recouvre le contenu de [62, §5.7]. Pour le démontrer, rappelons que $N_0(X)(a)$ est facteur direct de $h(X)$ d'après le théorème 11.13. Si $a < d/2$, $\Lambda(N) \cap \Lambda(N_0(X)(a)) \neq \emptyset$, d'où l'assertion. Supposons maintenant $a = d/2$. Alors $d/2 < i_1(X)$, donc d'après la proposition 6.6, X est définie par une forme de Pfister. D'après le théorème 11.14, $h(X)$ contient exactement deux motifs indécomposables N' tels que $L^{d/2} \in \Lambda(N')$, à savoir $N' = N_0(X)$ et $N' = N_0(X)(d/2)$. L'énoncé résulte donc du lemme 11.15.

11.4. La taille du motif initial

THÉORÈME 11.18. — *Soit X une quadrique anisotrope. Alors :*

- a) $t(N_0(X)) = \dim_{\text{es}}(X)$.
- b) $N_0(X) \simeq N_0(X)^\vee(\dim_{\text{es}}(X))$.

La première partie de ce théorème est le corollaire 4.7 de [62]; la deuxième partie est un cas particulier du théorème 4.19 de *loc. cit.* La démonstration se fait en deux étapes :

- 1) Le cas $i_1(X) = 1$.
- 2) Réduction au cas 1).

Pour l'étape 1), nous proposons la démonstration suivante, différente de celle de Vishik : posons $N = N_0(X)$ et $K = F(X)$. D'après (10.7), on a

$$h(X)_K \simeq \mathbf{1} \oplus h(X_1)(1) \oplus L^d$$

où X_1 est la quadrique anisotrope définie par q_1 . On a aussi

$$N_K \simeq \mathbf{1} \oplus N'.$$

D'autre part,

$$\delta(N) = \delta(N_K) = 1 + \delta(N')$$

est pair (lemme 10.5), donc $\delta(N')$ est impair. En réappliquant le lemme 10.5, il en résulte que N' n'est pas facteur direct de $h(X_1)(1)$. Mais alors on a forcément $L^d \in \Lambda(N)$, ce qui démontre a).

Pour b), on observe que $N^\vee(d)$ est facteur direct de $h(X)^\vee(d) \simeq h(X)$ et que, d'après a), $\Lambda(N) \cap \Lambda(N^\vee(d)) \neq \emptyset$ et on conclut par la proposition 11.7.

L'étape 2) est une méthode devenue classique dans le sujet. Il y a plusieurs manières de faire cette réduction : dans [26, dém. du lemme 7.9], Izhiboldin utilise une technique générique. Vishik, pour sa part, démontre que X contient une sous-quadrique Y telle que $X \approx Y$ et $i_1(Y) = 1$ [62, sublemma 5.25]. Si $i_1(X) = 1$, il n'y a rien à démontrer.

Supposons que $i_1(X) > 1$. Alors toute sous-quadrique Z de codimension 1 est stablement birationnellement équivalente à X (lemme 7.1 c)) et on s'en tire par récurrence sur $d = \dim X$.

Le théorème 11.12 montre maintenant que $N_0(X) \simeq N_0(Y)$. D'après le théorème 11.13, $N_0(X)(i_1(X) - 1)$ est facteur direct de $h(X)$, ce qui implique a priori que $b = b(N_0(X)) \leq \dim_{\text{es}}(X)$. Pour montrer l'inégalité inverse, Vishik observe que le lemme 10.6 implique d'abord que $b > d/2$, puis que $b > d - i_1(X)$ parce que $N_{F(X)}$, et donc $h(X)_{F(X)}$, contient un facteur direct isomorphe à L^b .

Le corollaire 8.4 découle immédiatement des théorèmes 11.12 et 11.18. On déduit aussi du théorème 11.18 le corollaire suivant, dû à Karpenko [34, th. 6.4] :

COROLLAIRE 11.19. — *Soit α une correspondance de X vers elle-même, où X est une quadrique anisotrope telle que $i_1(X) = 1$. Alors $\mu(\alpha) \equiv \mu(\alpha^t) \pmod{2}$, où α^t est la correspondance transposée et μ est la multiplicité (cf. 10.6).*

Démonstration. — Soit π le projecteur définissant $N_0(X)$: on a $\mu(\pi) = 1$ (cf. fin de 10.6). On a donc par 10.6 (2)

$$\mu(\alpha) = \mu(\pi\alpha\pi).$$

Le théorème 11.18 implique en particulier que $\pi = \pi^t$. On a donc

$$(\pi\alpha\pi)^t = \pi\alpha^t\pi.$$

Rappelons que $\text{End}(N_0(X)) = \mathbf{F}_2$ ou $\mathbf{F}_2 \oplus \mathbf{F}_2\theta$, où $\theta = 1 - \tau \in \text{End}(h(X))$ est l'élément du théorème 11.2 (proposition 11.5). Soit $A = \text{End}(N_0(X))$ dans le premier cas et $A = \text{End}(N_0(X))/\langle\theta\rangle$ dans le second. Remarquons que $\theta^t = \theta$, donc la transposition induit une involution de $A \simeq \mathbf{F}_2$, et cette involution est nécessairement l'identité. Ainsi

$$(\pi\alpha\pi)^t = \pi\alpha\pi \quad \text{dans } A.$$

D'autre part, $\mu(\theta) = \mu(1) - \mu(\tau_*) = 0$ par 10.6 (4). On obtient donc finalement, modulo 2 :

$$\mu(\alpha) = \mu(\pi\alpha\pi) = \mu((\pi\alpha\pi)^t) = \mu(\pi\alpha^t\pi) = \mu(\alpha^t)$$

comme demandé. □

11.5. Motifs supérieurs

Soit X une quadrique anisotrope de dimension d , d'équation $q = 0$, et soit (i_1, \dots, i_h) sa suite de déploiement (définition 6.1). Pour $0 \leq r \leq h$, posons également

$$I_r = \sum_{t=1}^r i_t.$$

Ainsi I_r est l'indice de Witt de q_{F_r} (voir encore définition 6.1).

THÉORÈME 11.20. — Soit $r \in [0, h[$. Supposons qu'il existe un facteur direct indécomposable N de $h(X)$ tel que $a(N) \in [I_r, I_{r+1}[$. Alors :

- a) Pour tout $j \in [I_r, I_{r+1}[$, $N(j - a(N))$ est facteur direct de $h(X)$.
- b) On a $t(N) = \dim_{\text{es}}(X_r)$.
- c) On a $N \simeq N^\vee(2a(N) - \dim_{\text{es}}(X_r))$.

Ce théorème est équivalent à [62, th. 4.13 et cor. 4.14]. Pour démontrer a), Vishik distingue deux cas : $j \geq a(N)$ et $j \leq a(N)$. Dans le premier, la démonstration est analogue à celle du théorème 11.13 ([62, §5.8] ; il utilise les motifs des grassmanniennes quadratiques du §6.2). Ensuite, il ramène le deuxième au premier par dualité.

Nous proposons la démonstration suivante de b), un peu différente de celle de Vishik : dans le cas $r = 0$, c'est déjà connu (théorème 11.13, corollaire 11.17, théorème 11.18). À partir de là, on procède par récurrence sur r de la manière suivante.

LEMME 11.21 (cf. [62, sublemma 5.29]). — $b(N) \leq d - I_r$.

En effet, supposons le contraire. Alors $a(N^\vee(d)) = d - b(N) < i_r$. Observons que $N^\vee(d)$ est facteur direct (indécomposable) de $h(X)^\vee(d) \simeq h(X)$. Par récurrence, on a

$$t(N) = t(N^\vee(d)) = \dim_{\text{es}}(X_s) = d - I_s + 1$$

pour un $s < r$. Mais alors

$$b(N) = a(N) + d - I_s + 1 \geq d + I_r - I_s + 1 > d + 1$$

ce qui est impossible. □

Le lemme 11.21 implique que N_{F_r} est un facteur direct de $h(X_r)(I_r)$. Par conséquent, $N' = N_{F_r}(-I_r)$ est facteur direct de $h(X_r)$, et $a(N') < i_1(X_r)$. Il résulte du corollaire 11.17 que N' contient $N_0(X_r)(a(N'))$ en facteur direct, et d'après le théorème 11.18 que

$$t(N) = t(N') \geq \dim_{\text{es}}(X_r).$$

Grâce à la partie a) du théorème 11.20, on peut supposer que $a(N) = I_{r+1} - 1$. Alors $t(N) \geq \dim_{\text{es}}(X_r) \Rightarrow b(N) \geq I_{r+1} + \dim(X_r) - i_{r+1} = d - I_r$, donc $b(N) = d - I_r$ grâce au lemme 11.21, d'où b). Enfin, pour c), on remarque que $d - b(N) \in [I_r, I_{r+1}[$ et on applique la proposition 11.7.

COROLLAIRE 11.22. — Pour N comme dans le théorème 11.20, on a $b(N) \geq d/2$.

En effet, on a

$$b(N) = a(N) + t(N) \geq I_r + \dim_{\text{es}} X_r = I_r + d - 2I_r - i_{r+1} + 1 = d - I_{r+1} + 1$$

et cet entier est toujours $\geq d/2$. □

Le corollaire 11.22 implique par dualité que $a(N) \leq d/2$ pour tout facteur direct indécomposable N de $h(X)$: ainsi le théorème 11.20 décrit tous ces facteurs directs. En particulier :

COROLLAIRE 11.23 ([62, th. 4.19]). — Tout facteur direct indécomposable N du motif d'une quadrique est polarisable : il existe un entier r tel que $N^\vee \simeq N(r)$. □

On en déduit :

COROLLAIRE 11.24. — *Soient N, N' deux facteurs directs indécomposables de motifs de quadriques et soit $f : N \rightarrow N'$ un morphisme. Supposons que f_K soit un isomorphisme, où K/F est une extension. Alors f est un isomorphisme.*

Démonstration. — En tenant compte du corollaire 11.23, c'est la même que celle du corollaire 10.15 c) (on observe d'abord que, nécessairement, $a(N) = a(N')$ et $b(N) = b(N')$). \square

Remarque 11.25. — D'après le corollaire 11.17, pour N comme dans le théorème 11.20, N_{F_r} contient un twist $N_0(X_r)(a)$ de $N_0(X_r)$ en facteur direct. Vishik m'a donné un exemple où $N_{F_r} \neq N_0(X_r)(a)$ bien que $t(N) = t(N_0(X_r)(a))$: il prend une sous-forme q de codimension 1 d'une forme p de dimension 12 telle que $p \in I^3 F$. Si X est la quadrique correspondante (de dimension 9), on a $h(X) = N_0(X) \oplus N$ où N est indécomposable (avec $\Lambda(N) = \{L, L^3, L^6, L^8\}$). Mais q_1 est une voisine de Pfister (ce résultat est dû à Izhboldin), donc $N_0(X_1)$ est un motif binaire d'après le théorème 11.14, ce qui montre que N_{F_1} est décomposable. *cf.* [62, p. 77].

11.6. Équivalence motivique

THÉORÈME 11.26. — *Soient q, p deux formes quadratiques, $m, n \geq 0$ et X, Y les quadriques associées. Supposons que, pour toute extension K/F , les conditions $i(p_K) > n$ et $i(q_K) > m$ soient équivalentes. Supposons que $h(Y)$ admette un facteur direct indécomposable N tel que $a(N) = n$. Alors $h(X)$ admet $N(m - n)$ comme facteur direct.*

C'est le théorème 4.17 de [62]. Sa preuve est dans [62, §5.9] : c'est une généralisation de celle de l'implication \Rightarrow dans le théorème 11.12, qui repose également sur le théorème 11.10 et utilise les grassmanniennes quadratiques. En voici deux corollaires :

COROLLAIRE 11.27. — *Soit X une quadrique anisotrope, et soit $r \in [0, h[$ où h est la hauteur de X . On prend les notations de la définition 6.1. Supposons qu'il existe une quadrique anisotrope Y telle que $Y \approx X^{(r+1)}$, où $X^{(r+1)}$ est une grassmannienne quadratique de corps des fonctions équivalent à F_{r+1} (*cf.* §6.2). Alors l'hypothèse du théorème 11.20 est vérifiée.*

C'est le cas particulier $n = 0$ du théorème 11.26. \square

COROLLAIRE 11.28 (Équivalence motivique). — *Soient X, Y deux quadriques anisotropes de la même dimension. Alors $h(X) \simeq h(Y)$ si et seulement si $i(X_K) = i(Y_K)$ pour toute extension K/F .*

C'est l'un des plus beaux résultats de Vishik, le théorème 4.18 de [62]. La nécessité résulte facilement du théorème 11.26 ; pour la suffisance, Vishik remarque simplement que le motif de X (ou celui de Y) « code » les indices de Witt supérieurs (*cf.* lemme 10.6).

Izhboldin a remarqué :

PROPOSITION 11.29 ([24, cor. 2.9]). — *Si dans le corollaire 11.28 la dimension commune de X et Y est impaire, alors ses deux conditions équivalentes équivalent encore à $X \simeq Y$.*

En effet, choisissons deux équations q, q' de X et Y . Quitte à multiplier q' par un scalaire, on peut supposer que $q \perp -q' \in I^2F$ (lemme 4.4). Par hypothèse, le corps $F_1 = F(q)$ est un corps d'isotropie générique commun à q et q' . Par récurrence sur la hauteur commune de q et q' , on peut supposer que $q_1 \simeq q'_1$; autrement dit, $(q \perp -q')_{F_1} \sim 0$. En appliquant le théorème 6.12 de Fitzgerald (ou la version plus faible obtenue plus élémentairement par Izhboldin dans [24, cor. 1.2]), on en déduit que $q \perp -q'$ est soit hyperbolique, soit semblable à une forme de Pfister. Mais le deuxième cas est impossible puisque $\dim q = \dim q'$ est impaire. \square

11.7. La taille d'un motif binaire

DÉFINITION 11.30. — *Un motif $N \in \text{Mot}(F)_{\text{tate}}$ est binaire si $H(N)$ est de la forme $L^a \oplus L^b$.*

Le théorème suivant est l'un des résultats les plus profonds de Vishik :

THÉORÈME 11.31 ([62, th. 4.20]). — *Soit N un motif binaire, facteur direct du motif d'une quadrique anisotrope X . Alors $t(N)$ est de la forme $2^n - 1$.*

Sa démonstration originelle [60, dém. de Statement 6.1] utilise les opérations de Steenrod motiviques de Voevodsky : elle est reproduite dans [25, th. 6.1]. Plus récemment, Karpenko et Merkurjev ont donné dans [39] une démonstration n'utilisant que les opérations construites par Brosnan (cf. 10.7).

Remarque 11.32. — Vishik prouve plus : si $\dim_{\text{es}} X = 2^n - 1$ (cas auquel on peut toujours se ramener), alors $\text{Ker}(H^{n+1}(F, \mathbf{Z}/2) \rightarrow H^{n+1}(F(X), \mathbf{Z}/2)) \neq 0$. Karpenko et Merkurjev ne retrouvent pas ce complément. Il est directement lié à la conjecture ci-dessous (cf. [31, « conjecture » énoncée après le corollaire 2 de l'introduction] : cette conjecture prédit que le noyau précédent est engendré par un symbole (a_1, \dots, a_{n+1})). Pour des compléments là-dessus, nous renvoyons à l'article [25] d'Izhboldin et Vishik.

CONJECTURE 11.33. — *Soit N un facteur direct indécomposable binaire du motif d'une quadrique, de taille $2^n - 1$. Alors il existe une $(n + 1)$ -forme de Pfister φ telle que N soit de la forme $N_0(X_\varphi)(a)$.*

C'est la conjecture 4.21 de [62], cf. aussi [35, conj. 1.6] : elle implique que, si tous les facteurs indécomposables de $h(X)$ sont binaires, la quadrique anisotrope X est définie par une forme excellente. Elle est connue pour $n \leq 2$ (facilement) et pour $n = 3$ (Karpenko [35]).

11.8. Compléments sur le motif initial

Soit X une quadrique anisotrope. La proposition 11.7 montre que les termes de $\Lambda(N_0(X))$ contrôlent dans une certaine mesure quels motifs supérieurs interviennent dans le théorème 11.20, et présentent donc une interaction avec la suite de déploiement de X . Vishik a explicité cette observation dans un certain nombre de théorèmes.

NOTATION 11.34. — Gardons les notations du théorème 11.20. Soit $r \in [0, h[$. On note

$$\Lambda_r(X) = \{n \in [I_r, I_{r+1}[\mid L^n \in \Lambda(N_0(X))\}.$$

(Vishik emploie le terme *r-th shell* pour désigner l'ensemble $[I_r, I_{r+1}[$.)

THÉORÈME 11.35. — a) Soit $r \in]1, h]$. Si $i_r < i_1$, alors $\Lambda_r(X) = \emptyset$.
b) Si i_2 n'est pas divisible par i_1 , alors $\Lambda_2(X) = \emptyset$.

THÉORÈME 11.36. — Supposons que $\Lambda_r(X) = \emptyset$ pour tout $r > 0$. Alors $N_0(X)$ est binaire.

THÉORÈME 11.37. — Supposons que $h(X_1) \simeq \bigoplus_{l=0}^{i_2-1} N_0(X_1)(l)$ (on pourrait dire que $h(X_1)$ est engendré par $N_0(X_1)$). Alors soit $h(X) \simeq \bigoplus_{l=0}^{i_1-1} N_0(X)(l)$, soit $N_0(X)$ est binaire (les deux cas étant possibles simultanément).

Ce sont respectivement les théorèmes 7.7, 7.8 et 7.9 de [62] : ils se démontrent par des arguments de comptage. Nous renvoyons à [62, §7] pour les démonstrations, ou laissons au lecteur le plaisir de les reconstituer à partir des résultats exposés précédemment. Nous leur ajoutons :

THÉORÈME 11.38. — Supposons que, pour tout $r \in [0, h[$, il existe un facteur direct indécomposable N de $h(X)$ tel que $a(N) \in [I_r, I_{r+1}[$. Alors tous les facteurs directs indécomposables de $h(X)$ sont binaires, et donc aussi tous les $N_0(X_r)$ (cf. remarque 11.25).

La démonstration est du même tonneau, cf. celle du théorème 11.14.

À tous ces résultats on peut bien sûr ajouter ceux sur la taille des motifs supérieurs (théorème 11.18) et d'un motif binaire (théorème 11.31). Dans [62, §7], Vishik utilise ceci en conjonction avec des théorèmes de structure d'autres auteurs pour déterminer toutes les suites de déploiement possibles pour les formes de dimension paire ≤ 12 ou impaire ≤ 21 . (Hoffmann [22] avait auparavant traité le cas de toutes les formes de dimension ≤ 10 .)

12. QUELQUES DÉMONSTRATIONS

12.1. Démonstration du théorème 8.2

La démonstration de Karpenko et Merkurjev n'utilise pas les opérations de Steenrod ni la théorie de Vishik ; nous allons la simplifier très légèrement en utilisant cette théorie, ce qui nous permet de considérer le corollaire 8.4 comme connu, *cf.* remarque juste avant le corollaire 11.19, ainsi que ce dernier énoncé. Le corollaire 8.4 montre que l'énoncé du théorème 8.2 est stablement birationnel en X , et la réciproque du lemme 6.5 nous réduit alors au cas où $i_1(X) = 1$, ce que font Karpenko et Merkurjev plus élémentairement. Ils démontrent alors (1) et (2) simultanément par double récurrence sur $n = \dim X + \dim Y$, le cas $n = 0$ étant trivial.

Soit $\alpha \in CH_d(X \times Y)$ la correspondance donnée par le graphe Z de l'application rationnelle $X \rightsquigarrow Y$ déduite d'un point fermé de $Y_{F(X)}$ de degré impair : alors $\mu(\alpha)$ est impair.

Supposons d'abord $\dim X = \dim Y =: d$ et démontrons (2) : nous allons en fait montrer que $\mu(\alpha^t)$ est impair. Supposons le contraire ; soit $x \in X$ un point de degré 2 : quitte à modifier α par un multiple de $x \times Y$, on peut alors supposer que $\mu(\alpha^t) = 0$. Comme $\deg : CH_0(X_{F(Y)}) \rightarrow \mathbf{Z}$ est injectif (théorème 3.2), cela implique que α provient d'une correspondance $\alpha' \in CH_d(X \times Y')$ avec Y' un fermé propre de Y , et $\mu(\alpha') = \mu(\alpha)$ est impair. Choisisant un cycle représentant α' , on peut supposer ce cycle irréductible et donc aussi Y' . Mais $\dim Y' < \dim Y = \dim X$, ce qui contredit (1) par récurrence sur n .

Cette partie de la démonstration utilise seulement le fait que X est une quadrique, mais pas encore l'hypothèse $i_1(X) = 1$; elle va intervenir maintenant.

Démontrons maintenant (1). Supposons au contraire que $\dim Y < \dim X$. Karpenko et Merkurjev se ramènent d'abord au cas où $Z \rightarrow Y$ est surjectif en remplaçant Y par l'image de Z . Leur stratégie est alors la suivante :

- (i) Produire une correspondance $\gamma : Y \rightarrow X$ de multiplicité impaire.
- (ii) À l'aide de α et γ , fabriquer une correspondance $\delta : X \rightarrow X$ telle que $\mu(\delta)$ soit impaire mais $\mu(\delta^t) = 0$, ce qui contredira le corollaire 11.19.

Pour (i), ils font intervenir une astuce : quitte à faire une extension transcendante pure, ce qui ne change pas la situation, on peut supposer que X contient une sous-quadrique X' de même dimension que Y et vérifiant encore $i_1(X) = 1$ (cette construction est due à Izhboldin ; nous y avons fait allusion au §11.4). Si $\iota : X' \rightarrow X$ est l'immersion fermée correspondante, ils prennent

$$\gamma = \iota_* \iota^*(\alpha^t)$$

qui est bien de multiplicité impaire grâce à (2), par récurrence.

Pour (ii), comme dans la démonstration de (2) on peut « extraire » de γ une correspondance γ' de support T irréductible, tel que la projection $T \rightarrow Y$ soit surjective de

degré générique impair. Choisissons maintenant une variété propre S munie de morphismes $S \rightarrow T, S \rightarrow Z$ génériquement finis, le second étant de degré générique impair (prendre pour S une composante irréductible convenable de $T \times_Y Z$). Les deux projections (passant par Z et par T) $S \rightarrow X$ fournissent la correspondance δ cherchée (on a $\mu(\delta^t) = 0$ parce que γ n'est pas dominante).

(Comme ils le remarquent, si Y est lisse il suffit de prendre la composée $\gamma \circ \alpha$; on s'en tire dans le cas général en approximant cette construction, parce que le problème est birationnel.)

12.2. Démonstration du théorème 8.5

Cette démonstration utilise les opérations de Steenrod (plus précisément une opération) et, implicitement, la théorie de Vishik, mais Karpenko exprime sa démonstration en termes de cycles algébriques sans parler de motifs. Nous allons la résumer. Cela donnera une idée de sa démonstration du théorème 8.6, qui est de la même eau mais en plus compliqué.

Karpenko raisonne par l'absurde en supposant qu'il existe une quadrique anisotrope X de dimension d telle que $i_1 := i_1(X) > 2^r$, où $r = v_2(d - i_1 + 2)$. Quitte à prendre une section hyperplane itérée, on peut supposer $i_1 = 2^r + 1$ sans changer $d - i_1$ (réciproque du lemme 6.5). En particulier, on remarque que d est *impair*.

Posons

$$e^i = \begin{cases} h^i & \text{si } i < d/2 \\ l_{d-i} & \text{si } i > d/2 \end{cases}$$

cf. (10.10). Dans $CH^{d-2^r}(\bar{X} \times \bar{X})/2$, tout cycle α s'écrit

$$\alpha = \sum_{i=0}^d \alpha_i e^i \times e^{d-2^r-i}, \quad \alpha_i \in \mathbf{F}_2.$$

Karpenko exhibe un cycle α , défini sur F , tel que $\alpha_{d-2^r} = 1$: pour cela il prend le même cycle que Vishik dans la démonstration du théorème 11.13, à savoir l'adhérence dans $X \times X$ d'un sous-espace projectif de dimension 2^r de $X_{F(X)}$. Observant que $e^i \times e^{d-2^r-i} = h^i \times h^{d-2^r-i}$ est défini sur F pour $d/2 - 2^r < i < d/2$, il modifie α de façon à assurer $\alpha_i = 0$ pour ces valeurs de i .

Comme $2^r < i_1 \leq d/2$, e^{2^r} est défini sur F . Le cycle $\beta = \alpha \cdot (e^0 \times e^{2^r})$ s'écrit $\sum \beta_i e^i \times e^{d-i}$, avec $\beta_i = \alpha_i$ pour $i \in [0, d - 2^r]$: ceci résulte de la normalisation ci-dessus et des formules (10.11). Considérant maintenant β comme une correspondance de X vers lui-même, les théorèmes 11.13 et 11.18 impliquent

$$(12.1) \quad \beta_i = \beta_{d-i_1+1+i} = \beta_{d-2^r+i} \text{ pour } i \in [0, 2^r].$$

Comme on a évidemment $\beta_i = 0$ pour $i \in [d - 2^r + 1, d]$, (12.1) implique que $\beta_i = 0$ pour $i \in [1, 2^r]$, et donc que

$$(12.2) \quad \alpha_1 = \dots = \alpha_{2^r} = 0$$

(remarquer que $2^r \leq d - 2^r$).

Karpenko considère maintenant la correspondance de degré 0 $\gamma = S^{2^r}(\alpha)$. Il va obtenir une contradiction en montrant par un calcul que (avec des notations analogues aux précédentes) $\gamma_d = 0$, et par un autre calcul que $\gamma_d = 1$.

Pour le premier calcul, en réappliquant (12.1), il suffit de montrer que $\gamma_{2^r} = 0$: mais cela découle trivialement de (12.2) en appliquant simplement les propriétés (2) (formule de Cartan), (3) (pour $n = 0$) et (4) des opérations de Steenrod (§10.7). Pour le second calcul, l'égalité $\alpha_{d-2^r} = 1$ (qui n'a pas encore été utilisée!) implique que $S^{2^r}(e^{d-2^r} \times e^0) = \gamma_d(e^d \times e^0)$, et donc que $S^{2^r}(e^{d-2^r}) = \gamma_d e^d$ ou encore $S^{2^r}(l_{2^r}) = \gamma_d l_0$. Mais en appliquant (10.13), on obtient

$$S^{2^d}(l_{2^r}) = \binom{d - 2^r + 1}{2^r} l_0$$

et $\binom{d - 2^r + 1}{2^r}$ est impair. □

RÉFÉRENCES

- [1] Y. André, *Motifs de dimension finie (d'après Kimura, O'Sullivan...)*, Séminaire Bourbaki (2003/04), Exp. **929**.
- [2] Y. André, B. Kahn, *Nilpotence, radicaux et structures monoïdales* (avec un appendice de P. O'Sullivan), Rend. Sem. Mat. Univ. Padova **108** (2002), 107–291.
- [3] Y. André, B. Kahn, *Erratum : Nilpotence, radicaux et structures monoïdales*, à paraître aux Rend. Sem. Mat. Univ. Padova **113**, (2005).
- [4] J.Kr. Arason, *Cohomologische Invarianten quadratischer Formen*, J. Alg. **36** (1975), 448–491.
- [5] J.Kr. Arason, M. Knebusch, *Über die Grade quadratischer Formen*, Math. Ann. **234** (1978), 167–192.
- [6] J.Kr. Arason, A. Pfister, *Beweis des Krullschen Durchschnittsatzes für den Witttring*, Invent. Math. **12** (1971), 173–176.
- [7] P. Balmer, *Witt cohomology, Mayer-Vietoris, Homotopy invariance and the Gersten Conjecture*, K-Theory **23** (2001), 15–30.
- [8] P. Balmer, C. Walter, *A Gersten-Witt spectral sequence for regular schemes*, Ann. Sci. Éc. Norm. Sup. **35** (2002), 127–152.
- [9] A. Borel, *Linear algebraic groups* (2ème édition), Springer, 1991.
- [10] P. Brosnan, *Steenrod operations in Chow theory*, Trans. AMS **355** (2003), 1869–1903.
- [11] P. Brosnan, *A short proof of Rost nilpotence via refined correspondences*, Doc. Math. **8** (2003) 69–78.

- [12] C. Chevalley, *Sur les décompositions cellulaires des espaces G/B* , in Algebraic groups and their generalizations : classical methods, Proc. Sympos. pure Math. **56** (I), 1–23, AMS, 1994.
- [13] V. Chernousov, S. Gille, A. Merkurjev, *Motivic decomposition of isotropic projective homogeneous varieties*, à paraître au Duke Math. J.
- [14] M. Demazure, *Motifs de variétés algébriques*, Séminaire Bourbaki (1969/70), Exp. **365**, Lect. Notes in Math. **180**, Springer (1971), 19–38.
- [15] L.E. Dickson *On quadratic forms in a general field*, Bull. AMS **14** (1907), 108–115 = Math. papers IV, Chelsea, 1975, 512–519.
- [16] D. Edidin, W. Graham, *Equivariant intersection theory*, Invent. Math. **131** (1998), 595–634.
- [17] R.W. Fitzgerald, *Function fields of quadratic forms*, Math. Z. **178** (1981), 63–76.
- [18] E. Friedlander, *Motivic complexes of Suslin-Voevodsky*, Séminaire Bourbaki (1996/97), Exp. **833** (juin 1997) Astérisque **245** (1997), 355–378.
- [19] W. Fulton, *Intersection theory*, Springer, 1984 (2ème édition, 1998).
- [20] D.K. Harrison, *A Grothendieck ring of higher degree forms*, J. Algebra **35** (1975), 123–138.
- [21] D.W. Hoffmann, *Isotropy of quadratic forms over the function field of a quadric*, Math. Z. **220** (1995), 461–476.
- [22] D.W. Hoffmann, *Splitting patterns and invariants of quadratic forms*, Math. Nachr. **190** (1998), 149–168.
- [23] J. Hurrelbrink, U. Rehmann, *Splitting patterns of quadratic forms*, Math. Nachr. **176** (1995), 111–127.
- [24] O. Izhboldin, *Motivic equivalence of quadratic forms*, Doc. Math. **3** (1998), 341–351.
- [25] O. Izhboldin, A. Vishik, *Quadratic forms with absolutely maximal splitting*, in Quadratic forms and their applications (Dublin, 1999), Contemp. Math. **272**, AMS, 2000, 103–125.
- [26] O. Izhboldin, *Fields of u -invariant 9*, Ann. of Math. **154** (2001), 529–587.
- [27] U. Jannsen, *Motives, numerical equivalence and semi-simplicity*, Invent. Math. **107** (1992), 447–452.
- [28] B. Kahn, *The total Stiefel-Whitney class of a regular representation*, J. of Alg. **144** (1991), 214–247.
- [29] B. Kahn, *Formes quadratiques de hauteur et de degré 2*, Indag. Mathem. **7** (1996), 47–66.
- [30] B. Kahn, *La conjecture de Milnor, d’après V. Voevodsky*, Séminaire Bourbaki (1996/97), Exp. **834** (juin 1997), Astérisque **245** (1997), 379–418.

- [31] B. Kahn, M. Rost, R. Sujatha, *Unramified cohomology of quadrics, I*, Amer. J. Math. **120** (1998), 841–891.
- [32] N. Karpenko, *Invariants algébro-géométriques de formes quadratiques* (en russe), Algebra i Analiz **2** (1990), 141–162. Traduction anglaise : Leningrad Math. J. **2** (1991), 119–138.
- [33] N. Karpenko, *Isotropie d’espaces cellulaires relatifs et de variétés de drapeaux isotropes* (en russe), Algebra i Analiz **12** (2000), 3–69. Traduction anglaise : St. Petersburg Math. J. **12** (2001), 1–50.
- [34] N. Karpenko, *On the anisotropy of orthogonal involutions*, J. Ramanujan Math. Soc. **15** (2000), 1–22.
- [35] N. Karpenko, *Characterization of minimal Pfister neighbors by Rost projectors*, J. pure appl. Algebra **160** (2001), 195–227.
- [36] N. Karpenko, *On the first Witt index of quadratic forms*, Invent. Math. **153** (2003), 455–462.
- [37] N. Karpenko, *Holes in I^n* , Ann. Sci. Éc. Norm. Sup. **37** (2004), 973–1002.
- [38] N. Karpenko, *A relation between higher Witt indices*, à paraître aux Transl. AMS.
- [39] N. Karpenko, A. Merkurjev, *Rost projectors and Steenrod operations*, Doc. Math. **7**(2002), 481–493.
- [40] N. Karpenko, A. Merkurjev, *Essential dimension of quadrics*, Invent. Math. **153** (2003), 361–372.
- [41] I. Kersten, U. Rehmann, *Generic splitting of reductive groups*, Tôhoku Math. J. **46** (1994), 35–70.
- [42] M. Knebusch, *Specialization of quadratic and symmetric bilinear forms, and a norm theorem*, Acta Arith. **24** (1973), 279–299.
- [43] M. Knebusch, *Generic splitting of quadratic forms, I*, Proc. London Math. Soc. **33** (1976) 65–93.
- [44] M. Knebusch, *Generic splitting of quadratic forms, II*, Proc. London Math. Soc. **34** (1977), 1–31.
- [45] T.Y. Lam, *The algebraic theory of quadratic forms* (2ème édition), Benjamin, New York, 1980.
- [46] F. Loeser, *Cobordisme de variétés algébriques (d’après M. Levine et F. Morel)*, Séminaire Bourbaki (2001/02), Exp. **901** (mars 2002), Astérisque **290** (2003), 167–192.
- [47] Yu.I. Manin, *Correspondances, motifs et transformations monoïdales* (en russe), Mat. Sb. **77** (**119**) (1968), 475–507. Traduction anglaise : Math. USSR Sbornik.
- [48] F. Morel, *An introduction to \mathbf{A}^1 -homotopy theory*, In Contemporary Developments in Algebraic K-theory, ICTP Lecture notes **15** (2003), 357–441, M. Karoubi, A.O. Kuku, C. Pedrini (ed.).

- [49] F. Morel, V. Voevodsky, \mathbf{A}^1 -homotopy theory of schemes, Publ. Math. IHÉS **90** (1999), 45–143 (2001).
- [50] D. Orlov, A. Vishik, V. Voevodsky, *An exact sequence for $K_*^M/2$ with applications to quadratic forms*, prépublication, 2000, www.math.uiuc.edu/K-theory/454.
- [51] E. Peyre, *Corps de fonctions de variétés homogènes et cohomologie galoisienne*, C. R. Acad. Sci. Paris **321** (1995), 891–896.
- [52] A. Pfister, *Multiplikative quadratische Formen*, Arch. Math. **16** (1965), 363–370.
- [53] M. Rost, *Some new results on the Chow groups of quadrics*, prépublication, Regensburg, 1990.
- [54] W. Scharlau *On the history of the algebraic theory of quadratic forms*, in Quadratic forms and their applications (Dublin, 1999), Contemp. Math. **272**, AMS, 2000, 229–259.
- [55] J.E. Schneider, *Orthogonal groups of nonsingular forms of higher degree*, J. Algebra **27** (1973), 112–116.
- [56] A. Scholl, *Classical motives*, in Motives, Proc. Symp. pure Math. **55** (I), 163–187, AMS, 1994.
- [57] T.A. Springer, *Sur les formes quadratiques d'indice zéro*, C.R. Acad. Sci. Paris **234** (1952), 1517–1519.
- [58] R.G. Swan, *Zero cycles on quadric hypersurfaces*, Proc. Amer. Math. Soc. **107** (1989), 43–46.
- [59] B. Totaro, *The Chow ring of a classifying space*, in Algebraic K-theory (Seattle, 1997), Proc. Symp. pure Math. **67** (1999), 249–281.
- [60] A. Vishik, *Integral motives of quadrics*, prépublication du Max Planck Institut für Mathematik (Bonn) 1998 (**13**), 1–82.
- [61] A. Vishik, *On dimensions of anisotropic forms in I^n* , prépublication du Max Planck Institut für Mathematik (Bonn) 2000 (**11**), 1–41.
- [62] A. Vishik, *Motives of quadrics with applications to the theory of quadratic forms*, in Geometric methods in the algebraic theory of quadratic forms, Lect. Notes in Math. **1835**, Springer, 2004, 25–101.
- [63] V. Voevodsky, *Triangulated categories of motives over a field*, in Cycles, transfers, and motivic homology theories, Ann. of Math. Stud. **143**, Princeton Univ. Press, 2000, 188–238.
- [64] V. Voevodsky, *Cancellation theorem*, prépublication, 2000.
- [65] V. Voevodsky, *Reduced power operations in motivic cohomology*, Publ. Math. IHÉS **98** (2003), 1–57.
- [66] V. Voevodsky, *Motivic cohomology with $\mathbf{Z}/2$ -coefficients*, Publ. Math. IHÉS **98** (2003), 59–104.

- [67] A.R. Wadsworth, Noetherian pairs and function fields of quadratic forms, Thèse, Université de Chicago, 1972.
- [68] A.R. Wadsworth, *Similarity of quadratic forms and isomorphism of their function fields*, Trans. Amer. Math. Soc. **208** (1975), 352–358.
- [69] E. Witt, *Theorie der quadratischen Formen in beliebigen Körpern*, J. reine angew. Math. **176** (1937), 31–44.

Bruno KAHN

Institut de Mathématiques de Jussieu
175–179 rue du Chevaleret
F–75013 Paris
E-mail : kahn@math.jussieu.fr

**CORRESPONDANCES DE HECKE, ACTION DE GALOIS
ET LA CONJECTURE D'ANDRÉ-OORT**
[d'après Edixhoven et Yafaev]

par **Rutger NOOT**

1. INTRODUCTION

Il y a une analogie très forte entre les conjectures de Manin–Mumford et d'André–Oort, dont la première est en effet un théorème dû à Raynaud. Dans les deux cas, on considère une certaine classe de variétés algébriques: les variétés abéliennes dans la première et les variétés de Shimura dans la deuxième conjecture. Dans chaque cas, on définit ensuite la notion de sous-variété (irréductible) *spéciale*. Dans le cas de variétés abéliennes, on parlera de *sous-variétés de torsion*, dans le cas des variétés de Shimura de *sous-variétés de type Hodge*. Une sous-variété spéciale de dimension nulle est un *point spécial*. Les définitions, dans le cas des variétés de Shimura, peuvent être consultées dans les paragraphes 2.1 et 2.2, l'analogie avec la conjecture de Manin–Mumford sera expliquée en détail dans 3.1. Les deux conjectures s'énoncent alors de la manière suivante.

CONJECTURE 1.1 (cf. Conjecture 2.3). — *Soient S une variété de l'espèce considérée et $\Sigma \subset S$ un ensemble de points spéciaux. Alors les composantes irréductibles de l'adhérence de Zariski de Σ sont des sous-variétés spéciales de S .*

On reviendra plus tard, dans les paragraphes 3.1 et 5.1, sur l'analogie entre ces deux conjectures. La suite de cette introduction sera consacrée au cas le plus simple où la conjecture d'André–Oort n'est pas triviale. Cet exemple sera développé, dans le langage plus savant des variétés de Shimura, dans le paragraphe 5.2.

Soit $\mathbb{H} = \{\tau \in \mathbb{C} \mid \text{Im}(\tau) > 0\}$ le *demi-plan de Poincaré*. Le groupe $\text{SL}_2(\mathbb{R})$ opère sur \mathbb{H} par les transformations de Moebius. Cette action est transitive et induit une action fidèle de groupe quotient $\text{SL}_2(\mathbb{R})/\{\pm 1\}$. On obtient une action de $\Gamma = \text{SL}_2(\mathbb{Z})$ sur \mathbb{H} et on peut montrer que le quotient $\Gamma \backslash \mathbb{H}$ est une variété analytique. C'est le premier exemple d'une variété de Shimura. La fonction modulaire $j: \mathbb{H} \rightarrow \mathbb{C}$ (dont on peut trouver la définition dans [36, Chap. VII] par exemple) est Γ -invariante et elle induit

L'auteur a bénéficié du soutien du programme MRTN de l'Union Européenne, dans le cadre du réseau *Arithmetic Algebraic Geometry* (contrat MRTN-CT2003-504917).

un isomorphisme $\Gamma \backslash \mathbb{H} \cong \mathbb{C}$. Un élément z du quotient $\Gamma \backslash \mathbb{H}$ est dit *spécial* si c'est la classe d'un élément $\tau \in \mathbb{H}$ qui est algébrique sur \mathbb{Q} de degré 2, donc nécessairement quadratique imaginaire. On note que cette condition ne dépend pas du représentant τ choisi. De même, un point $z \in \mathbb{C}$ est *spécial* si c'est l'image d'un élément spécial de $\Gamma \backslash \mathbb{H}$ sous l'isomorphisme ci-dessus. Dans ce cas, on dit aussi que z est un invariant modulaire *singulier*. Les points spéciaux de \mathbb{C} sont donc les nombres $j(\tau)$ avec $\tau \in \mathbb{H}$ quadratique imaginaire. On peut montrer, et cela traduit un principe général dans la théorie de variétés de Shimura, que tout point spécial de \mathbb{C} est un nombre algébrique.

La théorie présentée ci-dessus possède une interprétation naturelle en termes de courbes elliptiques. Pour tout $\tau \in \mathbb{H}$, le quotient $E_\tau = \mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ est une courbe elliptique complexe et toute courbe elliptique sur \mathbb{C} est isomorphe à une courbe de cette forme. Deux courbes E_τ et $E_{\tau'}$ sont isomorphes si et seulement si $j(\tau) = j(\tau')$, de sorte que l'invariant modulaire $j(\tau)$ définit une bijection entre l'ensemble des classes d'isomorphisme de courbes elliptiques complexes et \mathbb{C} . En plus, pour $F \subset \mathbb{C}$ algébriquement clos, une courbe elliptique peut être définie sur F si et seulement si son invariant modulaire appartient à F . Une courbe elliptique est de *type CM* si son anneau d'endomorphismes $\text{End}(E_\tau)$ n'est pas réduit à \mathbb{Z} . Pour la courbe E_τ (avec $\tau \in \mathbb{H}$) cela est le cas si et seulement si τ est quadratique imaginaire et $\text{End}(E_\tau)$ est alors un ordre dans le corps $\mathbb{Q}(\tau)$. Les points spéciaux de \mathbb{C} sont donc les points correspondant aux courbes elliptiques de type CM. Comme une courbe de type CM peut être définie sur $\overline{\mathbb{Q}}$, cela implique que les points spéciaux de \mathbb{C} sont algébriques.

La conjecture d'André–Oort étant triviale pour la variété de Shimura que l'on vient d'introduire, on va considérer dans la suite le produit de cette variété avec elle-même, c'est-à-dire qu'on regarde $\mathbb{C} \times \mathbb{C}$ comme quotient de $\mathbb{H} \times \mathbb{H}$ sous l'action de $\Gamma \times \Gamma$. Les points spéciaux sont alors les $(z, z') \in \mathbb{C}^2$ avec z et z' spéciaux. La seule sous-variété de \mathbb{C}^2 de dimension 2 est \mathbb{C}^2 et celle-ci est de type Hodge.

On distingue deux types de courbes (irréductibles) dans \mathbb{C}^2 de type Hodge. La courbe $\{z\} \times \mathbb{C}$ resp. $\mathbb{C} \times \{z'\}$ est de type Hodge si et seulement si z , resp. z' , est spécial. Pour tout entier N , il existe une courbe de type Hodge $\tilde{Y}_0(N) \subset \mathbb{C}^2$ du deuxième type : $\tilde{Y}_0(N)$ est l'image de \mathbb{H} dans \mathbb{C} sous l'application $\tau \mapsto (j(N\tau), j(\tau))$. Pour montrer que cette image est bien une courbe algébrique on utilise le fait que l'application $\tau \mapsto (j(N\tau), j(\tau))$ se factorise par le quotient de \mathbb{H} pour un sous-groupe arithmétique de Γ . Le fait que cette définition *ad hoc* coïncide avec la définition donnée dans 2.2 est (mal) justifié dans le paragraphe 5.2 (alinéa suivant le théorème 5.3) et aussi (beaucoup mieux) dans Edixhoven [19, §2]. Notons aussi que $\tilde{Y}_0(N)$ est seulement birationnellement équivalente à la courbe modulaire $Y_0(N)$.

Maintenant qu'on sait ce que veut dire la conjecture d'André–Oort pour \mathbb{C}^2 , passons aux résultats connus dans ce cas. Pour la variété de Shimura \mathbb{C}^2 , la conjecture 1.1 a été prouvée par André [4] sous une condition supplémentaire sur l'ensemble Σ et indépendamment par Edixhoven [17] sous l'hypothèse de Riemann généralisée. André [5] a ensuite réussi à enlever la condition sur Σ , donnant une démonstration inconditionnelle

dans ce cas. La démonstration d’Edixhoven est reprise dans [19], qui donne aussi une variante (corollaire du théorème 7.1) où l’hypothèse de Riemann a été remplacée par une condition sur Σ . Dans le paragraphe 5.2 on reviendra sur les idées de la démonstration d’André. Quant à la démonstration d’Edixhoven, celle-ci a été généralisée par Edixhoven et Yafaev et s’applique maintenant, toujours sous des conditions supplémentaires, à beaucoup d’autres variétés de Shimura. Elle fait l’objet de la section 7.

On termine l’introduction en donnant les idées principales de cette méthode, appliquée au cas de \mathbb{C}^2 . Il est suffisant de montrer que toute courbe algébrique irréductible $Z \subset \mathbb{C}^2$ contenant un ensemble infini Σ de points spéciaux est de type Hodge. Soit Z une telle courbe. Comme les points spéciaux sont algébriques, la courbe Z est définie sur un corps de nombres F , c’est-à-dire que c’est une courbe $Z \subset \mathbb{A}_F^2$. Si une des deux projections de Z est réduite à un point, alors ce point est spécial et l’énoncé est trivialement vérifié. On peut donc se borner au cas contraire où il faut montrer que $Z = \tilde{Y}_0(N)$ pour un certain entier N . Supposons que cela ne soit pas le cas et essayons d’en déduire une contradiction.

Pour tout entier m , la courbe $\tilde{Y}_0(m) \subset \mathbb{C}^2$ est une correspondance $T_m: \mathbb{C} \rightarrow \mathbb{C}$, agissant sur les sous-ensembles $X \subset \mathbb{C}$ par $T_m X = \pi_2(\pi_1^{-1}(X))$, où les $\pi_i: \tilde{Y}_0(m) \rightarrow \mathbb{C}$ sont les restrictions des projections $\mathbb{C}^2 \rightarrow \mathbb{C}$. Le produit $T_{m,m} = T_m \times T_m: \mathbb{C}^2 \rightarrow \mathbb{C}^2$ est alors aussi une correspondance. Il suffit de trouver un nombre premier p tel que $T_{p,p}Z = Z$. En effet, si $T_{p,p}Z = Z$, alors la $T_{p,p}$ -orbite de tout élément de Z est contenue dans Z . Comme toutes les $T_{p,p}$ -orbites sont denses dans \mathbb{C}^2 , cela implique que $Z = \mathbb{C}^2$, contredisant le fait que Z est une courbe. La conjecture est alors prouvée pour \mathbb{C}^2 .

Il reste à trouver le nombre premier p avec cette propriété miraculeuse. L’argument se déroule en trois étapes.

1. On montre que si l’intersection $Z \cap T_{p,p}Z$ est finie, alors son ordre est majoré par $c(p+1)^2$ (pour une constante $c > 0$).
2. Le fait que $Z \neq \tilde{Y}_0(N)$ pour tout N implique, via un théorème d’André sur le groupe de monodromie algébrique, que $T_{p,p}Z$ est irréductible pour $p > M$, avec M assez grand. On utilise ici un cas particulier du théorème 7.5.
3. On se sert d’une description explicite de l’action du groupe de Galois absolu Γ_F de F sur les points spéciaux pour montrer qu’il existe un $z \in \Sigma$ et un nombre premier $p > M$ tels que $T_{p,p}z$ contient un conjugué galoisien de z et tels que l’ordre de la Γ_F -orbite de z est supérieur à $c(p+1)^2$.

La dernière étape nécessite une version effective du théorème de Chebotarev qui n’a été prouvé que sous l’hypothèse de Riemann généralisée, ce qui explique que le résultat dépend de GRH. Alternativement, une condition assez forte sur les points spéciaux dans Σ implique aussi l’existence de z et p , voir le théorème 7.1 et sa démonstration. Il faut aussi souligner que $T_{p,p}\tilde{Y}_0(N)$ n’est irréductible pour aucun nombre premier p et entier $N > 0$, donc l’hypothèse que Z ne soit pas une courbe modulaire est essentielle dans la deuxième étape.

Comme Z et $T_{p,p}Z$ sont définis sur F , la dernière étape implique que $Z \cap T_{p,p}Z$ contient toute l'orbite $\Gamma_F \cdot z$. En combinant la minoration de l'ordre de cette orbite avec la majoration de l'intersection établie dans la première étape on déduit que $Z \cap T_{p,p}Z$ est infini. Comme Z et $T_{p,p}Z$ sont irréductibles on conclut que $Z = T_{p,p}Z$.

Remerciements. Je remercie tous ceux qui m'ont aidé dans la préparation de cet exposé et en particulier Bas Edixhoven pour sa relecture rapide et minutieuse du manuscrit.

2. LA CONJECTURE

2.1. Variétés de Shimura

Une variété de Shimura connexe est un quotient d'un domaine hermitien symétrique par l'action d'un groupe arithmétique. Une variété de Shimura générale est une réunion disjointe de variétés de Shimura connexes. Même si la conjecture d'André–Oort peut s'énoncer dans toute sa généralité pour les variétés de Shimura connexes, on utilisera le langage adélique de Deligne [15] et [16] parce que c'est le cadre naturel pour introduire les opérateurs de Hecke et les lois de réciprocité. Pour les détails du résumé succinct suivant, le lecteur est renvoyé aux deux articles de Deligne. Les notations introduites resteront en vigueur dans tout ce texte.

Notons $\mathbb{C}^\times = \text{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m$ le tore sur \mathbb{R} obtenu par restriction de scalaires. Ce tore est caractérisé par la propriété que $\mathbb{C}^\times(A) = (\mathbb{C} \otimes A)^\times$ pour toute \mathbb{R} -algèbre A . Une *donnée de Shimura* est un couple (G, X) , où G est un groupe linéaire algébrique réductif sur \mathbb{Q} et $X \subset \text{Hom}_{\mathbb{R}\text{-grp}}(\mathbb{C}^\times, G_{\mathbb{R}})$ une $G(\mathbb{R})$ -classe de conjugaison telle que les conditions habituelles [16, 2.1.1. {1,2,3}] soient vérifiées. Pour la suite, on fixe une donnée de Shimura (G, X) . Les composantes connexes de X sont alors des (produits de) domaines hermitiens symétriques, en particulier X possède une structure complexe naturelle. Il est clair que les composantes de X sont toutes isomorphes entre elles et on en fixe une, notée X^+ .

Soit \mathbb{A} (resp. \mathbb{A}_f) l'anneau des adèles (finis) de \mathbb{Q} , de sorte que $\mathbb{A} = \mathbb{R} \times \mathbb{A}_f$ et $\mathbb{A}_f = \widehat{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{Q}$ où $\widehat{\mathbb{Z}}$ est le complété profini de \mathbb{Z} . Pour tout sous-groupe compact ouvert $K \subset G(\mathbb{A}_f)$ on considère le quotient

$$\text{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C}) = G(\mathbb{Q}) \backslash (X \times G(\mathbb{A}_f) / K),$$

où $G(\mathbb{Q})$ opère sur X par conjugaison (c'est-à-dire par composition avec des automorphismes intérieurs) et sur $G(\mathbb{A}_f) / K$ par translation à gauche. Chaque composante connexe de $\text{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ est isomorphe à un quotient de X^+ par l'action d'un groupe arithmétique. Plus précisément, soient $G(\mathbb{R})^+$ la composante connexe de $G(\mathbb{R})$ pour la topologie euclidienne et $G(\mathbb{Q})^+ = G(\mathbb{Q}) \cap G(\mathbb{R})^+$, alors $\text{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ est une réunion disjointe finie de quotients $\Gamma_g \backslash X^+$ avec $g \in G(\mathbb{A}_f)$ et $\Gamma_g = gKg^{-1} \cap G(\mathbb{Q})^+$. Ce quotient $\text{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ est un espace analytique et il résulte d'un théorème de Baily et Borel que c'est la variété des points complexes d'une variété algébrique complexe quasi

projective $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. Cette variété est lisse pour K , et donc les Γ_g , assez petits. Cela est le cas en particulier si les Γ_g sont sans torsion.

Les variétés de Shimura $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ forment un système projectif indexé par les sous-groupes compacts ouverts $K \subset G(\mathbb{A}^f)$ et leur limite projective est un \mathbb{C} -schéma $\mathrm{Sh}(G, X)_{\mathbb{C}}$ avec une action continue du groupe $G(\mathbb{A}^f)$. L'action de $g \in G(\mathbb{A}^f)$ sera notée

$$(1) \quad \mathrm{Sh}(G, X)_{\mathbb{C}} \xrightarrow{g} \mathrm{Sh}(G, X)_{\mathbb{C}}.$$

Par construction de $\mathrm{Sh}(G, X)_{\mathbb{C}}$, la variété $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est le quotient de $\mathrm{Sh}(G, X)_{\mathbb{C}}$ par l'action de K . Dans 6.2 on reviendra plus amplement sur cette action de $G(\mathbb{A}^f)$.

2.2. Sous-variétés de type Hodge et points spéciaux

On définit de façon évidente la notion de morphisme $f: (H, Y) \rightarrow (G, X)$ de données de Shimura. Un tel morphisme induit un morphisme de schémas

$$(2) \quad \mathrm{Sh}(f): \mathrm{Sh}(H, Y)_{\mathbb{C}} \rightarrow \mathrm{Sh}(G, X)_{\mathbb{C}}.$$

Une sous-variété irréductible fermée $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est appelée une *sous-variété de type Hodge* s'il existe un morphisme $f: (H, Y) \rightarrow (G, X)$ de données de Shimura et un élément $g \in G(\mathbb{A}^f)$ tels que Z soit une composante irréductible de l'image d'un morphisme composé

$$(3) \quad \mathrm{Sh}(H, Y)_{\mathbb{C}} \xrightarrow{\mathrm{Sh}(f)} \mathrm{Sh}(G, X)_{\mathbb{C}} \xrightarrow{g} \mathrm{Sh}(G, X)_{\mathbb{C}} \longrightarrow \mathrm{Sh}_K(G, X)_{\mathbb{C}}.$$

Il n'est pas difficile de montrer (voir 6.2 pour plus de détails) que l'image d'une telle application est une sous-variété fermée, pas nécessairement irréductible, de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. Chaque composante irréductible de l'image de $\mathrm{Sh}(H, Y)_{\mathbb{C}}$ est l'image d'une composante connexe de $\mathrm{Sh}(H, Y)_{\mathbb{C}}$.

Dans le cas particulier où H est un tore, la variété $\mathrm{Sh}(H, Y)_{\mathbb{C}}$ est de dimension nulle, donc les sous-variétés de type Hodge obtenues à partir de la construction précédente appliquée à (H, Y) sont des points. Les points obtenus de cette manière sont les *points spéciaux* de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. Les conditions imposées à une donnée de Shimura (H, Y) impliquent que les points spéciaux d'une variété de Shimura sont exactement les sous-variétés de type Hodge de dimension nulle. Une caractérisation équivalente des points spéciaux est obtenue en utilisant la notion de groupe de Mumford–Tate.

DÉFINITION 2.1. — *Pour $h \in X$, son groupe de Mumford–Tate $\mathrm{MT}(h)$ est le plus petit sous-groupe algébrique $H \subset G$ (défini sur \mathbb{Q}) tel que $h: \mathbb{C}^{\times} \rightarrow G_{\mathbb{R}}$ se factorise par $H_{\mathbb{R}}$.*

Pour un point $s \in \mathrm{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$, on choisit un représentant $(h, a) \in X \times G(\mathbb{A}^f)$ et on définit le groupe de Mumford–Tate $\mathrm{MT}(s)$ de s comme étant $\mathrm{MT}(h)$.

Soient les notations comme dans la définition. Si $(h', a') \in X \times G(\mathbb{A}^f)$ est un autre représentant de s , alors $\mathrm{MT}(h')$ est conjugué à $\mathrm{MT}(h)$ par un élément de $G(\mathbb{Q})$ donc $\mathrm{MT}(s)$ est défini à isomorphisme près (comme groupe algébrique sur \mathbb{Q}) et à conjugaison

près comme sous-groupe de G . Tout groupe de Mumford–Tate est réductif donc un groupe de Mumford–Tate est commutatif si et seulement si c’est un tore.

LEMME 2.2. — *Un point $s \in \mathrm{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ est spécial si et seulement si $\mathrm{MT}(s)$ est un tore.*

On peut maintenant formuler la conjecture d’André–Oort.

CONJECTURE 2.3 (André–Oort). — *Fixons une donnée de Shimura (G, X) et un sous-groupe compact ouvert $K \subset G(\mathbb{A}_f)$. Soit $\Sigma \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ un ensemble de points spéciaux. Alors chaque composante irréductible de l’adhérence de Zariski de Σ dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est une sous-variété de type Hodge.*

On obtient un énoncé équivalent en considérant des variétés de Shimura connexes.

2.3. Densité de l’ensemble des points spéciaux

Notons tout de suite que la réciproque de la conjecture est vraie et qu’on a même un énoncé bien plus fort. Soit $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une variété de Shimura. Il suit de [15, §5] qu’il existe $h \in X$ tel que $\overline{\mathrm{MT}(h)}$ soit un tore et que, pour un tel h , l’ensemble des points spéciaux de la forme (h, a) avec $a \in G(\mathbb{A}^f)$ soit dense dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$ pour la topologie euclidienne. Ce fait joue un rôle important dans l’approche de la conjecture qui fait l’objet de cet exposé. On en déduit aussi, trivialement, que l’ensemble de tous les points spéciaux d’une variété de type Hodge est dense pour la topologie euclidienne.

3. HISTORIQUE

3.1. Analogie avec le théorème de Raynaud

La conjecture implique que si Z est une courbe irréductible fermée dans une variété de Shimura contenant un nombre infini de points spéciaux, alors Z est une sous-variété de type Hodge. Ce cas particulier figure dans [2, X.4.5] en tant que «Problem 3». Dans ce livre, André souligne l’analogie avec le théorème de Raynaud [34]. Ce théorème affirme qu’une courbe fermée dans une variété abélienne complexe A contenant un nombre infini de points de torsion de $A(\mathbb{C})$ est une translatée d’une courbe elliptique $E \subset A$ par un point de torsion de $A(\mathbb{C})$.

Raynaud [35] a généralisé ce théorème à des sous-variétés quelconques de A contenant un sous-ensemble Zariski-dense de points de torsion, prouvant qu’une telle sous-variété est une *sous-variété de torsion*, c’est-à-dire un translaté d’une sous-variété abélienne par un point de torsion. Il a ainsi résolu la conjecture de Manin–Mumford. Dans l’introduction, on a souligné l’analogie entre la conjecture 2.3 et le théorème de Raynaud via le dictionnaire traduisant «variété de Shimura» en «variété abélienne», «point spécial» en «point de torsion» et «sous-variété de type Hodge» en «sous-variété de torsion». Dans 5.1 on verra comment cette comparaison peut être poussée plus loin.

Hindry [21] et [22] montre des variantes quantitatives du théorème de Raynaud, pour des variétés abéliennes et des groupes algébriques commutatifs définis sur un corps de nombres F . Sa stratégie repose sur le fait que l'image de l'action du groupe de Galois absolu Γ_F sur l'ensemble des points de torsion contient beaucoup d'homothéties et sur un argument d'intersection. Ainsi, la stratégie de Hindry présente une analogie remarquable avec l'approche de la conjecture d'André–Oort par Edixhoven et Yafaev.

La conjecture de Manin–Mumford a été généralisée par Bogomolov qui considère une variété abélienne A sur un corps de nombres et demande si toute sous-variété $X \subset A$ munie d'un ensemble Zariski-dense de points de $X(\overline{\mathbb{Q}})$ de hauteurs «assez petites» est de torsion. La conjecture de Bogomolov a été démontrée par Szpiro, Ullmo et Zhang, par des méthodes d'équidistribution, voir [1] pour un compte rendu des méthodes. Des travaux récents de Clozel et Ullmo mettent en œuvre des idées analogues dans le contexte des variétés de Shimura. On reviendra là-dessus dans le paragraphe 5.3.

3.2. Espaces de modules de variétés abéliennes

Dans de nombreux cas, une variété de Shimura peut être interprétée comme un espace de modules de variétés abéliennes polarisées de dimension donnée, munies d'une structure de niveau et éventuellement de certaines structures supplémentaires. Du point de vue de la conjecture 2.3, il est suffisant de considérer le cas où on n'impose pas de structure supplémentaire, voir 6.3.

Soient $g \geq 1$ un entier, V un \mathbb{Z} -module libre de rang $2g$ muni d'une forme symplectique parfaite $\psi = \langle \cdot, \cdot \rangle$ et $\mathrm{GSp}_{2g} = \mathrm{GSp}(V_{\mathbb{Q}}, \psi)$ le groupe des similitudes de la forme induite sur $V_{\mathbb{Q}} = V \otimes_{\mathbb{Z}} \mathbb{Q}$. Soit $h: \mathbb{C}^{\times} \rightarrow \mathrm{GSp}_{2g/\mathbb{R}}$ un morphisme définissant sur $V_{\mathbb{R}}$ une structure de \mathbb{C} -espace vectoriel tel que $\langle x, h(i)x \rangle > 0$ pour tout $0 \neq x \in V_{\mathbb{R}}$ et soit X_g la $\mathrm{GSp}_{2g}(\mathbb{R})$ -classe de conjugaison de h . Pour $n \geq 1$, on définit $K_n \subset \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \subset \mathrm{GSp}_{2g}(\mathbb{A}^f)$ comme le noyau de la projection $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/n\mathbb{Z})$. La variété de Shimura $\mathrm{Sh}_{K_n}(\mathrm{GSp}_{2g}, X_g)_{\mathbb{C}}$ s'identifie alors à l'espace de modules $\mathcal{A}_{g,n/\mathbb{C}}$ des variétés abéliennes complexes principalement polarisées de dimension g , munies d'une structure de niveau n , cf. [15]. En ce qui concerne la structure de niveau, il faut l'interpréter comme une structure de niveau symplectique «de Jacobi» comme définie dans [27, 1.4].

Comme il est expliqué dans [26, §2], les sous-variétés de $\mathcal{A}_{g,n/\mathbb{C}}$ de type Hodge sont les sous-variétés irréductibles fermées maximales où certaines classes dans la cohomologie du schéma abélien universel sont des classes de Hodge. Les sous-variétés maximales où la variété abélienne correspondante admet des endomorphismes supplémentaires en sont des cas particuliers.

Les points spéciaux de $\mathcal{A}_{g,n/\mathbb{C}}$ sont les points correspondant aux variétés abéliennes de type CM. Rappelons qu'une variété abélienne simple A/\mathbb{C} est *de type CM* si $\dim \mathrm{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q} = 2 \dim A$ et qu'en général, une variété abélienne est *de type CM* si elle est isogène à un produit de variétés abéliennes simples de type CM. On peut définir le groupe de Mumford–Tate d'une variété abélienne A/\mathbb{C} (à isomorphisme près) comme le groupe de Mumford–Tate du point de $\mathcal{A}_{g,n/\mathbb{C}}$ correspondant à A , voir la définition 2.1.

On peut montrer qu’une variété abélienne est de type CM si et seulement si son groupe de Mumford–Tate est un tore.

Dans le cas particulier d’une variété de modules, la conjecture 2.3 affirme que toute sous-variété irréductible fermée $Z \subset \mathcal{A}_{g,n/\mathbb{C}}$ telle que $Z(\mathbb{C})$ contienne un ensemble Zariski-dense de points correspondant à des variétés abéliennes de type CM est une sous-variété de type Hodge. Cette version de la conjecture est émise par Oort dans [31] et [32], voir aussi la thèse de Moonen [25] et les deux publications [26] et [27] issues de cette thèse. Le contenu des deux derniers articles, lié à la conjecture d’André–Oort, fait l’objet de 5.1.

3.3. Généralisations

Dans [4, 3.2] et [6], André évoque la question suivante, qui généralise simultanément la conjecture 2.3 et le théorème de Raynaud cité en 3.1. On considère un groupe algébrique linéaire connexe G sur \mathbb{Q} , un sous-groupe maximal compact $K_\infty \subset G(\mathbb{R})$ contenant un \mathbb{R} -tore maximal et un sous-groupe de congruence $\Gamma \subset G(\mathbb{Q})$. On suppose cette fois-ci que le quotient $X = G(\mathbb{R})/K_\infty$ possède une structure complexe $G(\mathbb{R})^+$ -invariante et que $S(\mathbb{C}) = \Gamma \backslash X$ est l’ensemble de points d’une variété algébrique complexe S .

Une variété spéciale $Z \subset S$ est une sous-variété fermée telle que $Z(\mathbb{C})$ soit l’image de $gH(\mathbb{R}) \subset G(\mathbb{R})$ dans $S(\mathbb{C})$, pour $g \in G(\mathbb{Q})$ et $H \subset G$ un sous-groupe algébrique tel que $H(\mathbb{R}) \cap K_\infty$ soit un sous-groupe compact maximal de $H(\mathbb{R})$. Un point spécial est encore une sous-variété spéciale de dimension nulle.

L’énoncé de la conjecture d’André–Oort se généralise par la question suivante.

QUESTION 3.1 (André). — *Une sous-variété fermée $Z \subset S$ est-elle spéciale si et seulement si $Z(\mathbb{C})$ contient un sous-ensemble Zariski-dense de points spéciaux?*

Cette question généralise la conjecture aux variétés de Shimura mixtes et elle généralise aussi l’énoncé du théorème de Raynaud. En effet, si n est assez grand, la variété de modules $\mathcal{A}_{g,n/\mathbb{C}}$ porte un schéma abélien universel $\mathcal{X}_g \rightarrow \mathcal{A}_{g,n/\mathbb{C}}$. L’espace total \mathcal{X}_g peut être obtenu par la construction esquissée ci-dessus. Dans cet exemple, les points spéciaux de $\mathcal{X}_g(\mathbb{C})$ sont les points de torsion dans les fibres de \mathcal{X}_g au-dessus des points spéciaux de $\mathcal{A}_{g,n/\mathbb{C}}$. Les sous-variétés spéciales irréductibles sont les familles $Z \rightarrow \overline{Z}$ avec $\overline{Z} \subset \mathcal{A}_{g,n/\mathbb{C}}$ une sous-variété irréductible de type Hodge et Z l’adhérence dans \mathcal{X}_g d’une sous-variété de torsion de la fibre générique de $\mathcal{X}_g \times \overline{Z}$. Dans [4], André traite le cas particulier d’un pinceau de courbes elliptiques, voir aussi [6].

Pink [33] propose une variante généralisant à la fois la conjecture 2.3 aux variétés de Shimura mixtes, avec l’hypothèse supplémentaire que tous les points spéciaux appartiennent à une seule orbite de Hecke généralisée, et impliquant la conjecture de Mordell–Lang. Très récemment, Pink a proposé une autre question qui implique la conjecture de Mordell–Lang et la conjecture d’André–Oort pour les variétés de Shimura mixtes, en toute généralité.

Une variante de la conjecture 2.3 pour le produit de deux courbes modulaires de Drinfeld a été prouvée par Breuer [7]. La méthode de Breuer est basée sur celle d’Edixhoven et Yafaev. Dans le cas des courbes modulaires de Drinfeld, cette méthode donne un résultat inconditionnel parce que l’hypothèse de Riemann généralisée est vraie pour les corps de fonctions sur un corps fini.

4. QUELQUES APPLICATIONS

4.1. Jacobiennes de type CM

Les deux auteurs André [2] et Oort [31], [32] de la conjecture évoquent la relation entre leur variante de la conjecture et une conjecture de Coleman [12] concernant les jacobiniennes de type CM. Il s’agit du problème suivant.

CONJECTURE 4.1 (Coleman). — *Soit $g \geq 4$ un entier. Il n’existe alors, à isomorphisme près, qu’un nombre fini de courbes projectives et lisses C/\mathbb{C} de genre g telles que $\text{Jac}(C)$ soit de type CM.*

En associant à une courbe sa variété jacobienne, on définit un morphisme de l’espace de modules des courbes algébriques de genre g vers $\mathcal{A}_{g,1/\mathbb{C}}$. Soient $\mathcal{T}_{g/\mathbb{C}}$ l’image de ce morphisme et $\overline{\mathcal{T}}_{g/\mathbb{C}}$ l’adhérence de Zariski de $\mathcal{T}_{g/\mathbb{C}}$. En admettant la conjecture d’André–Oort, la conjecture de Coleman est équivalente à ce que $\overline{\mathcal{T}}_{g/\mathbb{C}}$ ne contient aucune sous-variété de type Hodge de dimension strictement positive qui rencontre $\mathcal{T}_{g/\mathbb{C}}$. Or, les exemples donnés dans [14] montrent que pour $g = 4, 6$ la variété $\overline{\mathcal{T}}_{g/\mathbb{C}}$ contient bien de telles sous-variétés. Cela est aussi valable pour $g = 7$ car Oort a remarqué que les méthodes de [14] s’appliquent aussi à la famille de courbes lisses définie par $y^9 = x(x-1)(x-\lambda)$. Il s’ensuit que la conjecture de Coleman est fautive pour $g = 4, 6$ et 7. La conjecture de Coleman et la question suivante restent toutefois ouvertes et intéressantes pour $g \geq 8$.

QUESTION 4.2 (Oort [31], [32]). — *Quelles sont les variétés de type Hodge de dimension strictement positive qui sont contenues dans $\overline{\mathcal{T}}_{g/\mathbb{C}}$ et qui rencontrent $\mathcal{T}_{g/\mathbb{C}}$?*

Pour répondre à la question, il faut comparer les dimensions des espaces de déformations de courbes munies d’une collection de classes de Hodge aux dimensions des espaces de déformations de variétés abéliennes avec les classes de Hodge correspondantes. Ce problème pourrait être un peu plus accessible que la conjecture de Coleman. Pour prouver la conjecture de Coleman par cette méthode, on aurait bien entendu besoin de la conjecture d’André–Oort.

4.2. Transcendance et valeurs spéciales de fonctions hypergéométriques

Pour $a, b, c \in \mathbb{Q}$ avec $-c \notin \mathbb{N}$, Wolfart [38] considère la fonction hypergéométrique multivaluée sur $\mathbb{P}^1 \setminus \{0, 1, \infty\}$ définie pour $|z| < 1$ par

$$F(a, b, c; z) = 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{2!c(c+1)}z^2 + \dots$$

La fonction $F(a, b, c; z)$ est solution de l'équation différentielle hypergéométrique. Wolfart s'intéresse alors au théorème suivant.

THÉORÈME 4.3. — *Soient $E_{a,b,c} = \{\xi \in \overline{\mathbb{Q}} \setminus \{0, 1\} \mid \exists F(a, b, c; \xi) \in \overline{\mathbb{Q}}\}$ et Δ le groupe de monodromie de l'équation différentielle hypergéométrique (ce groupe est muni d'une représentation fidèle naturelle). Alors*

- $E_{a,b,c} = \overline{\mathbb{Q}} \setminus \{0, 1\}$ si et seulement si Δ est fini,
- $E_{a,b,c} \subsetneq \overline{\mathbb{Q}} \setminus \{0, 1\}$ est infini si et seulement si Δ est un groupe arithmétique et
- $E_{a,b,c}$ est fini dans tous les autres cas.

La démonstration de Wolfart comportait une lacune, découverte par Gubler, qui a été comblée par Cohen et Wüstholz [11] en utilisant le cas particulier de la conjecture 2.3 prouvé par Edixhoven et Yafaev ([20] et le théorème 7.1 ci-après).

Indiquons brièvement comment on se ramène à la conjecture d'André–Oort. Si Δ est fini, alors $F(a, b, c; z)$ est algébrique sur $\overline{\mathbb{Q}}(z)$. On supposera dans la suite que Δ est infini. La stratégie est d'exprimer $F(a, b, c; z)$ comme le quotient d'une période d'une courbe algébrique $C_{a,b,c,z}$ et une période d'une courbe algébrique $D_{a,b,c}$. La courbe $D_{a,b,c}$ est une courbe de Fermat qui ne dépend pas de z et la période associée vient d'une variété abélienne de type CM.

La période de $C_{a,b,c,z}$ est associée à une partie convenable A_z de sa jacobienne, dont la dimension g est indépendante de z . L'application qui envoie $z \in \mathbb{C} \setminus \{0, 1\}$ vers A_z définit un morphisme $\mathbb{P}_{\mathbb{C}}^1 \rightarrow \mathcal{A}_{g/\mathbb{C}}$. Soit $Z_{a,b,c} \subset \mathcal{A}_{g/\mathbb{C}}$ l'image de cette application. D'une part, on montre que $Z_{a,b,c}$ est une sous-variété de type Hodge si et seulement si Δ est un groupe arithmétique.

D'autre part, on déduit d'un théorème profond de Wüstholz que z et $F(a, b, c; z)$ sont algébriques si et seulement si A_z et la jacobienne de $D_{a,b,c}$ ont un facteur simple en commun. Dans ce cas, A_z est de type CM, avec une algèbre d'endomorphismes qui ne dépend pas de z . Si Z est de type Hodge, alors A_z est de ce type pour un nombre infini de z . Réciproquement, si $E_{a,b,c}$ est infini, alors $Z(\mathbb{C})$ contient un nombre infini de points spéciaux et la conjecture 2.3 affirme que cela est le cas (si et) seulement si Z est de type Hodge. Comme les fibres de type CM ont la même algèbre d'endomorphismes, le théorème 7.1 fait déjà l'affaire et la preuve du théorème est achevée.

4.3. Une conjecture de Mazur

Vatsal et Cornut ont obtenu des résultats importants liés à une conjecture de Mazur concernant les sous-groupes d'une courbe elliptique E engendrés par les images des points de Heegner, via une paramétrisation modulaire de E . Dans [13], Cornut montre qu'une partie de son résultat peut être déduit assez facilement du cas particulier de la conjecture d'André–Oort prouvé par Moonen (théorème 5.2).

5. AUTRES APPROCHES

5.1. Les variétés de type Hodge sont formellement linéaires

On considère à nouveau l'espace de modules des variétés abéliennes principalement polarisées de dimension g . Il existe un tel espace de modules (grossier) sur \mathbb{Z} ; on le note $\mathcal{A}_{g,1} = \mathcal{A}_{g,1/\mathbb{Z}}$. Pour tout corps algébriquement clos k , l'ensemble $\mathcal{A}_{g,1}(k)$ est en bijection avec l'ensemble des classes d'isomorphisme des variétés abéliennes principalement polarisées sur k . En particulier, l'espace de modules $\mathcal{A}_{g,1/\mathbb{C}}$ sur \mathbb{C} considéré en 3.2 s'identifie à la fibre complexe $\mathcal{A}_{g,1} \otimes_{\mathbb{Z}} \mathbb{C}$.

Dans le reste de ce paragraphe, on fixe un entier n suffisamment grand pour que l'espace de modules des variétés abéliennes principalement polarisées de dimension g munies d'une structure de niveau n soit un espace de modules fin. On note $\mathcal{A}_{g,n}$ cet espace de modules sur $\mathbb{Z}[1/n]$. En fixant un nombre premier p ne divisant pas n , le corps $k = \overline{\mathbb{F}}_p$ et un point $x \in \mathcal{A}_{g,n}(k)$, on forme le complété formel $\widehat{\mathcal{A}}_{g,n}$ de $\mathcal{A}_{g,n/W(k)}$ en x , où $W(k)$ désigne l'anneau des vecteurs de Witt à coefficients dans k . Pour tout anneau local complet R avec corps résiduel k , l'ensemble $\widehat{\mathcal{A}}_{g,n}(R)$ s'identifie à l'ensemble des points de $\mathcal{A}_{g,n}(R)$ dont la réduction modulo l'idéal maximal \mathfrak{m}_R de R est égale à x .

Supposons que le point x correspond à une variété abélienne *ordinaire* A/k , ce qui veut dire que le groupe de p -torsion de $A(k)$ est de dimension maximale, isomorphe à $(\mathbb{Z}/p\mathbb{Z})^g$. La théorie de Serre–Tate montre que, dans ce cas, $\widehat{\mathcal{A}}_{g,n}$ est un groupe formel, isomorphe à $\widehat{\mathbb{G}}_m^d$, avec $d = \dim_{\mathbb{Z}} \mathcal{A}_{g,n} = g(g-1)/2$. Ici $\widehat{\mathbb{G}}_m$ est le complété du groupe multiplicatif en l'élément neutre, de sorte que, pour tout R comme ci-dessus, on a $\widehat{\mathbb{G}}_m(R) = (1 + \mathfrak{m}_R)^\times$. Les points de $\widehat{\mathcal{A}}_{g,n}(R)$ correspondant aux variétés abéliennes de type CM sont exactement les points de torsion pour cette structure de groupe formel.

Si $Z \subset \mathcal{A}_{g,n/\mathbb{C}}$ est une sous-variété, on note \overline{Z} son adhérence dans $\mathcal{A}_{g,n}$ et si $x \in \overline{Z}(k)$, alors le complété de $\overline{Z}_{W(k)}$ en x est un sous-schéma formel $\widehat{Z} \subset \widehat{\mathcal{A}}_{g,n}$.

THÉORÈME 5.1 ([29] et Moonen [27]). — *Soient Z , \overline{Z} et \widehat{Z} comme ci-dessus; supposons que $x \in \overline{Z}(k)$ correspond à une variété abélienne ordinaire. Alors $Z \subset \mathcal{A}_{g,n/\mathbb{C}}$ est de type Hodge si et seulement si toute composante de $\widehat{Z} \subset \widehat{\mathcal{A}}_{g,n} \cong \widehat{\mathbb{G}}_m^d$ est le translaté d'un sous-tore par un point de torsion.*

Tout cela renforce encore l’analogie avec le théorème de Raynaud donné par le dictionnaire du 3.1. L’analogie du théorème de Raynaud pour les tores formels est vrai et, en utilisant Moonen [27, 3.7] et la proposition 6.7, cela implique le résultat suivant.

THÉORÈME 5.2 (Moonen). — *Soit $\Sigma \subset \mathcal{A}_{g,1}(\overline{\mathbb{Q}})$ un ensemble de points spéciaux. Soit p un nombre premier tel que pour tout $s \in \Sigma$ il existe une place p -adique de $\overline{\mathbb{Q}}$ où la variété abélienne correspondant à s a bonne réduction ordinaire. Pour $s \in \Sigma$ donné, soit s_0 le point de $\mathcal{A}_{g,1}$ correspondant à cette réduction et supposons que chaque $s \in \Sigma$ est l’élément neutre du complété de $\mathcal{A}_{g,1}$ en s_0 . Alors les composantes irréductibles de l’adhérence de Zariski de Σ sont de type Hodge.*

Pour $s \in \Sigma$, soit End_s l’anneau des endomorphismes de la variété abélienne correspondante. Les conditions du théorème sont alors vérifiées si $\text{End}_s \otimes_{\mathbb{Z}} \mathbb{F}_p \cong \oplus_i M_{n_i}(\mathbb{F}_p)$ pour tout $s \in \Sigma$.

5.2. Le cas de \mathbb{C}^2

On fixe $H = \text{GL}_2$, on prend pour Y la $G(\mathbb{R})$ -classe de conjugaison d’un morphisme déduit de l’action naturelle de \mathbb{C}^\times sur $\mathbb{C} \cong \mathbb{R}^2$ et on définit $L \subset \text{GL}_2(\mathbb{A}_f)$ comme le stabilisateur d’un réseau $\widehat{\mathbb{Z}}^2 \subset \mathbb{A}_f^2$. La variété de Shimura $\text{Sh}_L(H, Y)_{\mathbb{C}}$ s’identifie alors à l’espace de modules des courbes elliptiques, donc $\text{Sh}_L(H, Y)_{\mathbb{C}} \cong \mathbb{A}_{\mathbb{C}}^1$. Comme dans l’introduction, on s’intéresse à la conjecture 2.3 pour la variété de Shimura

$$\mathbb{A}_{\mathbb{C}}^2 = \text{Sh}_L(H, Y)_{\mathbb{C}} \times \text{Sh}_L(H, Y)_{\mathbb{C}} = \text{Sh}_K(G, X)_{\mathbb{C}},$$

où $G = H \times H$, $X = Y \times Y$ et $K = L \times L$. Elle est interprétée comme espace de modules de produits $E \times E'$ avec E et E' des courbes elliptiques complexes. Pour toute courbe elliptique E/\mathbb{C} , on note $j(E)$ son invariant modulaire et pour $j \in \mathbb{A}^1(\mathbb{C})$, on notera E_j une courbe elliptique complexe d’invariant modulaire j . Rappelons de l’introduction et du paragraphe 3.2 que j est spécial si et seulement si E_j est de type CM.

La courbe modulaire $Y_0(N)$ est l’espace de modules de triplets (E, E', ι) où E et E' sont des courbes elliptiques complexes et $\iota: E \rightarrow E'$ est une isogénie dont le noyau est cyclique d’ordre N . L’application $(E, E', \iota) \mapsto (j(E), j(E'))$ définit un morphisme $Y_0(N) \rightarrow \mathbb{A}_{\mathbb{C}}^2$ dont l’image est une sous-variété de type Hodge. Cette image est la courbe $\widetilde{Y}_0(N)$ définie dans l’introduction. Avec les notations du paragraphe 2.2, c’est l’image de l’application (3) associée à l’inclusion diagonale $f: (H, Y) \rightarrow (G, X)$ et $g = ((\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}), (\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}))$.

THÉORÈME 5.3 (André [5]). — *Soit $Z \subset \mathbb{A}_{\mathbb{C}}^2$ une courbe algébrique irréductible, qui ne soit ni une droite horizontale ni une droite verticale. Alors Z est l’image d’une courbe modulaire $Y_0(N)$ si et seulement si Z contient une infinité de points (j, j') tels que j et j' soient spéciaux.*

Comme on a vu dans l’introduction, ceci prouve la conjecture 2.3 pour la variété de Shimura $\mathbb{A}_{\mathbb{C}}^2$. Cela implique aussi le fait que les sous-variétés de type Hodge de $\mathbb{A}_{\mathbb{C}}^2$ différentes des droites horizontales et verticales sont les courbes modulaires, plongées

dans $\mathbb{A}_{\mathbb{C}}^2$ comme ci-dessus. Notons qu'il n'est pas nécessaire de recourir à la conjecture d'André–Oort pour prouver ce fait, voir Edixhoven [19]. Des variantes du théorème 5.3 prouvées avec la méthode d'Edixhoven et Yafaev se trouvent dans Edixhoven [17] et [19].

Dans sa démonstration, André utilise une toute autre méthode, tellement élégante qu'il serait dommage de ne pas en donner un résumé. Soit Z comme dans l'énoncé et soit Σ un ensemble infini de points spéciaux dans Z . D'après le lemme 6.1, Z est alors définie sur un corps de nombres. En la remplaçant par la réunion de ses conjugués sous l'action du groupe de Galois $\Gamma_{\mathbb{Q}} = \text{Aut}_{\mathbb{Q}}(\overline{\mathbb{Q}})$, on se ramène au cas où Z est définie et irréductible sur \mathbb{Q} (mais pas absolument irréductible). André montre ensuite que l'on peut extraire de Σ une suite $((j_n, j'_n))$ telle que E_{j_n} et $E_{j'_n}$ admettent le même corps de multiplications complexes F_n et que la «distance» entre $\mathcal{O}_n = \text{End}(E_{j_n})$ et $\text{End}(E_{j'_n}) \subset F_n$ soit bornée. Comme Z est définie sur \mathbb{Q} , on peut remplacer chaque $((j_n, j'_n))$ par un conjugué sous l'action du groupe de Galois. On se ramène ainsi au cas où $E_{j_n} = \mathbb{C}/\mathcal{O}_n$. On montre ensuite qu'en passant encore à une suite extraite, on peut supposer que $(j_n, j'_n) \rightarrow (\infty, \infty)$ et que tous les points (j_n, j'_n) appartiennent à la même branche de Z passant par (∞, ∞) . Les arguments précis utilisés ici sont loin d'être triviaux, le raisonnement repose entre autres sur un résultat de Masser en théorie de transcendance. C'est cette partie de la démonstration qui manquait dans le travail antérieur d'André [4].

On montre maintenant assez facilement que, après avoir remplacé la suite par une suite extraite, on peut choisir les périodes τ_n de E_{j_n} et τ'_n de $E_{j'_n}$ sous la forme

$$\tau_n = \frac{c_n + f_n \sqrt{d_n}}{2}, \quad \tau'_n = \frac{b_n + f_n \sqrt{d_n}}{2a_n}$$

avec f, a_n, b_n, f_n entiers $a_n \geq |b_n|$ et $c_n = 0$ ou 1 . Comme les points (j_n, j'_n) appartiennent à la même branche d'une courbe algébrique, on obtient via le développement de Puiseux une comparaison des ordres de croissance $\log |j_n| \approx \rho \log |j'_n|$. Le développement de Fourier de l'invariant modulaire en fonction de la période τ donne des estimations $\log |j_n| \approx \pi f_n \sqrt{|d_n|}$ et $\log |j'_n| \approx (f/a_n) \pi f_n \sqrt{|d_n|}$. En combinant les deux estimations on montre d'abord que (a_n) est stationnaire et on se ramène ensuite trivialement au cas où les suites (b_n) et (c_n) le sont aussi. Cela implique que pour $n \gg 0$, on a $\tau'_n = (k + l\tau_n)/m$ pour certains entiers k, l, m et on en déduit que Z contient l'image de $Y_0(lm)$. Cette méthode se généralise sans trop de difficulté au cas d'une courbe $Z \subset \mathbb{A}_{\mathbb{C}}^n$.

5.3. Méthodes d'équidistribution

Dans une série de publications qui a commencée avec [10], Clozel et Ullmo proposent une approche de la conjecture d'André–Oort basée sur des méthodes ergodiques dans l'esprit de Ratner et Margulis. Ce travail se rapproche des résultats liés aux conjectures de Manin–Mumford et (surtout) de Bogomolov signalés dans le paragraphe 3.1. Ullmo [37] en donne un résumé qui souligne cette analogie.

Un problème important est que les résultats de Clozel et Ullmo sont pour l’instant valables pour les orbites de points spéciaux sous les correspondances de Hecke provenant du groupe de Mumford–Tate du point spécial en question. Pour avoir une application naturelle à la conjecture 2.3, on aurait besoin des énoncés analogues pour les orbites sous l’action du groupe de Galois.

Dans la suite, on se restreint à deux résultats de Clozel et Ullmo. On considère d’abord, à nouveau, les sous-variétés de type Hodge de $\mathbb{A}_{\mathbb{C}}^2$ et on conserve les notations de la section précédente. On note $d\mu$ la *mesure de Poincaré* sur $\mathbb{C} = \mathbb{A}^1(\mathbb{C})$. C’est la mesure déduite de la mesure hyperbolique H -invariante sur Y . Pour une courbe irréductible fermée $Z \subset \mathbb{A}_{\mathbb{C}}^2$ qui n’est ni une droite horizontale, ni une droite verticale, on note $Z_*\mu$ la mesure sur \mathbb{C} image directe de $d\mu$ par la correspondance définie par Z , concrètement

$$\int_{\mathbb{C}} f dZ_*\mu = \frac{1}{d_1} \int_{\mathbb{C}} \sum_{\substack{y \text{ tel que} \\ \pi_1(y)=x}} f(\pi_2(y)) d\mu,$$

où $\pi_i: Z \rightarrow \mathbb{C}$ est la i -ième projection et d_1 est le degré de π_1 .

THÉORÈME 5.4 (Clozel–Ullmo [8]). — *On a $Z_*\mu = \mu$ si et seulement si Z est l’image d’une courbe $Y_0(N)$.*

Supposons maintenant que $Z \subset \mathbb{A}_{\mathbb{C}}^2$ est une courbe contenant une suite dense (s_n) de points spéciaux. Comme dans la démonstration du théorème 5.3, on peut supposer que Z soit définie sur \mathbb{Q} . Dans [8, 2.4], Clozel et Ullmo proposent la piste suivante pour tenter de montrer que Z vérifie la condition du théorème 5.4 et de résoudre ainsi la conjecture 2.3 pour Z . Dans la suite on désigne, pour tout sous-ensemble fini E de Z , par $\Delta_E = \frac{1}{|E|} \sum_{s \in E} \delta_s$ la mesure de Dirac normalisée associée. Chaque s_n correspond à un produit de deux courbes elliptiques et, comme dans la démonstration du théorème 5.3, on se ramène au cas où ces deux courbes ont multiplication complexe par le même corps F_n . En utilisant des résultats de Duke et de Clozel–Ullmo, on peut montrer que, en passant à une suite extraite, la suite de mesures $(\Delta_{\Gamma_{F_n} \cdot s_n})$ converge faiblement vers une mesure $d\nu$ telle que $(\pi_1)_*d\nu = d\mu$ et $(\pi_2)_*d\nu = d\mu$. Pour en déduire la conjecture d’André–Oort pour $\mathbb{A}_{\mathbb{C}}^2$, via le théorème 5.4, on devrait montrer que ce fait implique que Z préserve la mesure $d\mu$. Ce problème est encore ouvert.

La conjecture d’André–Oort a la conséquence suivante pour les sous-variétés d’une variété de Shimura. Soient $\text{Sh}_K(G, X)_{\mathbb{C}}$ une variété de Shimura et $Y \subset \text{Sh}_K(G, X)_{\mathbb{C}}$ une sous-variété. Soit Σ l’ensemble des points spéciaux de $Y(\mathbb{C})$. La conjecture 2.3 implique alors que l’adhérence de Σ est une réunion finie $\cup_{i=1}^r S_i$ de sous-variétés de type Hodge $S_i \subset Y$, maximales par construction. Si $S' \subset Y$ est une sous-variété de type Hodge, alors la densité des points spéciaux de $S'(\mathbb{C})$ résultant du paragraphe 2.3 implique que S' est contenue dans une des S_i . Réciproquement, si dans toute sous-variété $Y \subset \text{Sh}_K(G, X)_{\mathbb{C}}$ il existe une réunion finie $\cup_{i=1}^r S_i$ de sous-variétés de type Hodge qui contient toute sous-variété de type Hodge $S' \subset Y$, alors la conjecture 2.3 est vraie pour Y . À noter que

cette condition inclut le cas où S' ou certaines S_i sont de dimension nulle, c'est-à-dire des points spéciaux.

En utilisant des méthodes ergodiques, Clozel et Ullmo montrent la variante suivante de cette propriété conjecturale. Dans ce théorème, où G est supposé adjoint, une *sous-variété fortement spéciale* $S \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est une sous-variété de type Hodge qui est l'image d'une application composée (3) de 2.2 pour un groupe semi-simple H qui n'est pas contenu dans un \mathbb{Q} -sous-groupe parabolique propre de G . En particulier, un point spécial n'est pas une sous-variété fortement spéciale.

THÉORÈME 5.5 (Clozel–Ullmo [9]). — *Soit $Y \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ comme avant; supposons de plus que G est un groupe adjoint. Il existe alors un ensemble fini $\{S_1, \dots, S_n\}$ de sous-variétés fortement spéciales de dimensions strictement positives tel que toute sous-variété fortement spéciale $S' \subset Y$ (avec $\dim S' > 0$) soit contenue dans une des S_i .*

Le rapport d'Ullmo [37] donne une excellente introduction au sujet de cette section, même si la publication [9] a évolué depuis la parution du rapport en question. Pour terminer ce paragraphe, il est signalé que les questions d'équidistribution sont également étudiées dans Zhang [42] et Jiang, Li et Zhang [23].

6. COMPLÉMENTS SUR LES VARIÉTÉS DE SHIMURA

6.1. Modèles canoniques et lois de réciprocité

On associe à la donnée de Shimura (G, X) son *corps dual*¹, un corps de nombres $E = E(G, X) \subset \mathbb{C}$, voir [16, 2.2] pour la définition. D'après des travaux de Shimura, Deligne, Borovoi et Milne, la variété de Shimura $\mathrm{Sh}(G, X)_{\mathbb{C}}$ admet un *modèle canonique* $\mathrm{Sh}(G, X)$ sur E , voir [28, §2] pour un compte rendu des arguments. Ce modèle canonique est un E -schéma tel que $\mathrm{Sh}(G, X)_{\mathbb{C}} = \mathrm{Sh}(G, X) \otimes_E \mathbb{C}$ et tel que l'action de $G(\mathbb{A}_f)$ sur $\mathrm{Sh}(G, X)_{\mathbb{C}}$, définie par les applications (1) dans 2.1, se descende en une action sur $\mathrm{Sh}(G, X)$. Tout cela implique que, pour tout sous-groupe compact ouvert $K \subset G(\mathbb{A}_f)$, la variété $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ admet également un modèle (canonique) sur E et que chaque composante connexe de cette variété admet un modèle sur un corps de nombres. Par abus de langage, on parlera également de modèle canonique s'agissant d'une variété obtenue par extension de scalaires à partir d'un modèle canonique et d'une composante connexe d'une telle variété.

Le modèle canonique doit vérifier les conditions de [16, 2.2.5] exigeant que les points spéciaux soient algébriques et décrivant l'action du groupe de Galois sur ces points. De ces conditions on déduit sans peine que les morphismes de variétés de Shimura $\mathrm{Sh}(f)$ définis par le diagramme (2) dans 2.2 sont définis sur le composé $E(H, Y)E(G, X) \subset \mathbb{C}$ des corps duaux des données de Shimura en jeu. La même chose est vraie pour la suite des applications (3), donc pour toute sous-variété de type Hodge $S_{\mathbb{C}} \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ il

1. *Reflex field* en anglais.

existe une extension finie $F \supset E(G, X)$ contenue dans \mathbb{C} telle que $S_{\mathbb{C}}$ provienne, par extension de scalaires, d'une sous-variété $S \subset \mathrm{Sh}_K(G, X)_F$.

Comme les points spéciaux sont algébriques, on a également le lemme suivant.

LEMME 6.1. — *Soient (G, X) une donnée de Shimura, $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert et $Z_{\mathbb{C}} \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une sous-variété irréductible et fermée contenant un ensemble Zariski-dense de points spéciaux. Alors il existe un corps de nombres F avec $E(G, X) \subset F \subset \mathbb{C}$ tel que $Z_{\mathbb{C}} \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ provienne par extension de scalaires d'une sous-variété $Z \subset \mathrm{Sh}_K(G, X)_F$.*

Remarque 6.2. — Considérons à nouveau le groupe de similitudes symplectiques GSp_{2g} , la donnée de Shimura (GSp_{2g}, X_g) et les sous-groupes $K_n \subset \mathrm{GSp}(\widehat{\mathbb{Z}})$ introduits dans 3.2. On a vu que la variété de Shimura $\mathrm{Sh}_{K_n}(\mathrm{GSp}_{2g}, X_g)_{\mathbb{C}}$ peut être identifiée à l'espace de modules $\mathcal{A}_{g,n/\mathbb{C}}$ de variétés abéliennes complexes munies d'une structure de niveau n . Comme le problème de modules admet une solution (grossière) sur \mathbb{Q} , on obtient un modèle $\mathcal{A}_{g,n/\mathbb{Q}}$ de $\mathcal{A}_{g,n/\mathbb{C}} = \mathrm{Sh}_{K_n}(\mathrm{GSp}_{2g}, X_g)_{\mathbb{C}}$. La théorie de la multiplication complexe des variétés abéliennes permet de montrer qu'il s'agit du modèle canonique. Bien entendu, la définition d'un modèle canonique est conçue pour que cela soit le cas.

Dans la suite on se servira de la description de l'action du groupe de Galois sur l'ensemble des points spéciaux d'une variété de Shimura. Celle-ci est déduite de la *loi de réciprocité* décrivant l'action du groupe de Galois sur la variété de Shimura associée à un tore. Pour décrire cette loi, soient T un \mathbb{Q} -tore et $h: \mathbb{C}^{\times} \rightarrow T_{\mathbb{R}}$ un morphisme tels que $(T, \{h\})$ constitue une donnée de Shimura. Le modèle canonique $\mathrm{Sh}(T, \{h\})$ est un $E = E(T, \{h\})$ -schéma profini et sa donnée est donc équivalente à celle de l'ensemble $\mathrm{Sh}(T, \{h\})(\overline{E})$ muni de l'action continue de $\Gamma_E = \mathrm{Aut}_E(\overline{E})$.

L'inclusion $\mathbb{R} \subset \mathbb{C}$ induit un morphisme de groupes $\mathbb{C}^{\times}(\mathbb{R}) \rightarrow \mathbb{C}^{\times}(\mathbb{C})$ et les caractères z et $\bar{z}: \mathbb{C}_{\mathbb{C}}^{\times} \rightarrow \mathbb{G}_{m/\mathbb{C}}$ sont définis par la condition que les composés

$$\mathbb{C}^{\times} = \mathbb{C}^{\times}(\mathbb{R}) \longrightarrow \mathbb{C}^{\times}(\mathbb{C}) \xrightarrow{z, \bar{z}} \mathbb{G}_m(\mathbb{C}) = \mathbb{C}^{\times}$$

soient respectivement l'identité et la conjugaison complexe². On définit le cocaractère $r: \mathbb{G}_{m/\mathbb{C}} \rightarrow \mathbb{C}_{\mathbb{C}}^{\times}$ comme le dual du caractère z . On pose

$$(4) \quad \mu_{\mathbb{C}} = h \circ r: \mathbb{G}_{m/\mathbb{C}} \rightarrow T_{\mathbb{C}}$$

et on définit le corps de nombres $E = E(T, \{h\}) \subset \mathbb{C}$ comme le corps de définition de $\mu_{\mathbb{C}}$. Le cocaractère $\mu_{\mathbb{C}}$ provient par extension de scalaires d'un morphisme $\mu: \mathbb{G}_{m/E} \rightarrow T_E$. Le morphisme $r(T, h)$ est défini par

$$(5) \quad r(T, h): E^{\times} = \mathrm{Res}_{E/\mathbb{Q}} \mathbb{G}_{m/E} \xrightarrow{\mathrm{Res}_{E/\mathbb{Q}} \mu} \mathrm{Res}_{E/\mathbb{Q}} T_E \xrightarrow{N_{E/\mathbb{Q}}} T,$$

2. La confusion occasionnée ici par le fait que \mathbb{C}^{\times} désigne à la fois un groupe algébrique réel et le groupe multiplicatif du corps \mathbb{C} ne semble pas insurmontable.

où $\text{Res}_{E/\mathbb{Q}}$ est la restriction de scalaires. Cette définition suit [24] où il est signalé que la définition donnée par Deligne [15] et [16] comporte une erreur de signe³.

La théorie des corps de classes fournit une identification $\Gamma_E^{\text{ab}} = \pi_0(E^\times(\mathbb{Q}) \backslash E^\times(\mathbb{A}))$, qu'on normalise de sorte qu'une uniformisante en une place v corresponde au frobenius géométrique en v . L'action de $\sigma \in \Gamma_E^{\text{ab}}$ sur

$$\text{Sh}(T, \{h\})(\overline{E}) = \lim_{K \subset T(\mathbb{A}_f)} \text{Sh}_K(T, \{h\})(\overline{E}) = \lim T(\mathbb{Q}) \backslash T(\mathbb{A}_f) / K = \overline{T(\mathbb{Q})} \backslash T(\mathbb{A}_f)$$

est la translation par la projection de $r(T, h)(\sigma)$.

LEMME 6.3. — *L'image de $r(T, h): E^\times \rightarrow T$ est le groupe de Mumford–Tate de h .*

Démonstration. — L'image T' de $r(T, h)$ est le tore engendré par les conjugués de $\mu_{\mathbb{C}}$ sous l'action de $\text{Aut}(\mathbb{C})$, donc T' est le plus petit \mathbb{Q} -sous-groupe de T tel que $\mu_{\mathbb{C}}$ se factorise par T' . Il s'ensuit que c'est aussi le plus petit sous-groupe de T tel que h se factorise par T' , c'est-à-dire le groupe de Mumford–Tate de h . \square

6.2. Correspondances de Hecke

Fixons une donnée de Shimura (G, X) et des sous-groupes $K_1, K_2 \subset G(\mathbb{A}_f)$ compacts ouverts. Le groupe $G(\mathbb{A}_f)$ opère sur $\text{Sh}(G, X)$, donc pour tout $g \in G(\mathbb{A}_f)$ on a un diagramme, défini sur $E(G, X)$,

$$\text{Sh}_{K_1}(G, X) \xleftarrow{\pi_{K_1}} \text{Sh}(G, X) \xrightarrow{\cdot g} \text{Sh}(G, X) \xrightarrow{\pi_{K_2}} \text{Sh}_{K_2}(G, X).$$

Comme π_{K_1} et $\pi_{K_2} \circ \cdot g$ se factorisent par $\text{Sh}_{K_g}(G, X)$ pour le sous-groupe compact ouvert $K_g = K_1 \cap gK_2g^{-1}$, l'image T_g de $\text{Sh}(G, X)$ dans $\text{Sh}_{K_1}(G, X) \times \text{Sh}_{K_2}(G, X)$ est une correspondance (telle que les projections de T_g sur les deux facteurs soient finies). On la note $T_g: \text{Sh}_{K_1}(G, X) \rightarrow \text{Sh}_{K_2}(G, X)$ et on observe que T_g ne dépend que de la classe de g dans $K_1 \backslash G(\mathbb{A}_f) / K_2$.

Pour toute extension $F \supset E$ et toute sous-variété fermée $Z \subset \text{Sh}_{K_1}(G, X)_F$ on écrira $T_g Z = (\pi_{K_2} \circ \cdot g)(\pi_{K_1}^{-1} Z)$; c'est une sous-variété fermée de $\text{Sh}_{K_2}(G, X)_F$ (mais pas nécessairement irréductible, même si Z l'est). Cette notation sera utilisée en particulier pour des ensembles de points fermés. Pour $x \in \text{Sh}_{K_1}(G, X)(\mathbb{C})$ on choisit $h \in X$ tel que x est la classe d'un élément $(h, a) \in X \times G(\mathbb{A}_f)$ et alors

$$T_g x = \{ \overline{(h, ag)} \in \text{Sh}_{K_2}(G, X)(\mathbb{C}) \mid a \in G(\mathbb{A}_f) \text{ tel que } \overline{(h, a)} = x \in \text{Sh}_{K_1}(G, X)(\mathbb{C}) \}.$$

Edixhoven montre le théorème suivant.

THÉORÈME 6.4 (Edixhoven [18], Theorem 7.2). — *Avec les notations introduites ci-dessus, supposons que $Z_i \subset \text{Sh}_{K_i}(G, X)_{\mathbb{C}}$ (pour $i = 1, 2$) soient des sous-variétés fermées dont une (au moins) est de dimension ≤ 1 . Il existe alors un entier c tel que pour tout $g \in G(\mathbb{A}_f)$ pour lequel $T_g Z_1 \cap Z_2$ est fini on ait*

$$|T_g Z_1 \cap Z_2| \leq c \deg(\text{Sh}_{K_g}(G, X)_{\mathbb{C}} \rightarrow \text{Sh}_{K_1}(G, X)_{\mathbb{C}}).$$

³. Noter également qu'entre [15] et [16] se produisent plusieurs changements de conventions affectant cette question, voir [24].

La démonstration se trouve dans la section 7 de [18], qui peut être lue indépendamment du reste de l'article cité. On se ramène à un calcul dans $\mathrm{Sh}_{K_g}(G, X)_{\mathbb{C}}$ en prenant les images réciproques de $T_g Z_1$ et de Z_2 . L'idée est d'utiliser ensuite les compactifications de Baily–Borel des variétés de Shimura et les fibrés en droites amples construites par Baily et Borel. On se ramène ainsi au calcul des degrés des Z_i par rapport aux fibrés en question.

COROLLAIRE 6.5. — *Soient $K \subset G(\mathbb{A}_f)$ compact ouvert et $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une courbe fermée. Alors il existe un entier c tel que pour tout $g \in G(\mathbb{A}_f)$ pour lequel $T_g Z \cap Z$ est fini on ait*

$$|T_g Z \cap Z| \leq c |K \cap g K g^{-1}|.$$

De manière analogue à la construction donnée ci-dessus, on définit les correspondances de Hecke sur les variétés de Shimura connexes. Avec les notations précédentes, fixons une composante X^+ de X comme dans 2.1 et posons $\Gamma_i = K_i \cap G(\mathbb{Q})^+$; soit $S_i \cong \Gamma_i \backslash X^+$ l'image de X^+ dans $\mathrm{Sh}_{K_i}(G, X)_{\mathbb{C}}$. Alors tout élément $g \in G(\mathbb{Q})$ définit une correspondance de S_1 vers S_2 via le diagramme

$$(6) \quad S_1 \xleftarrow{\pi_{\Gamma_1}} X^+ \xrightarrow{g \cdot} X^+ \xrightarrow{\pi_{\Gamma_2}} S_2.$$

Comme c'était le cas ci-dessus, la projection π_{Γ_1} et le composé $\pi_{\Gamma_2} \circ g \cdot$ se factorisent par des morphismes entre variétés algébriques, ici $\Gamma_g \backslash X^+ \rightarrow S_i$ avec $\Gamma_g = \Gamma \cap g^{-1} \Gamma g$. On aura à se servir du lemme suivant, dont la démonstration est élémentaire.

LEMME 6.6. — *Avec les notations en vigueur ci-dessus, soit $g \in G(\mathbb{A}_f)$ et soient les $g_i \in G(\mathbb{Q})^+$ tels que*

$$G(\mathbb{Q})^+ \cap K_1 g K_2 = \coprod_i \Gamma_1 g_i^{-1} \Gamma_2.$$

Alors les composantes connexes de $T_g \cap (S_1 \times S_2)$ sont les T_{g_i} .

6.3. Modification de (G, X) et de K

Soient (G, X) une donnée de Shimura et $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert. Pour montrer (un cas particulier de) la conjecture d'André–Oort, pour la variété de Shimura $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$, il s'avère souvent utile de remplacer K et G par d'autres groupes. On donnera ici quelques arguments tirés de [20, §2] permettant de telles modifications dans les hypothèses.

Si $K' \supset K$ est un sous-groupe compact ouvert de $G(\mathbb{A}_f)$, alors

$$\mathrm{Sh}_K(G, X)_{\mathbb{C}} \rightarrow \mathrm{Sh}_{K'}(G, X)_{\mathbb{C}}$$

est un morphisme fini et une sous-variété irréductible de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est de type Hodge si et seulement si son image l'est. Cela est vrai en particulier pour un point spécial et il s'ensuit :

PROPOSITION 6.7. — *La conjecture 2.3 pour un ensemble de points spéciaux Σ dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est équivalente à la conjecture pour l'image de Σ dans $\mathrm{Sh}_{K'}(G, X)_{\mathbb{C}}$.*

La discussion suivante est basée sur Moonen [26, 2.8 et 2.9] et Edixhoven–Yafaev [20, §2]. Soit $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une sous-variété irréductible. Il existe alors un sous-groupe $H \subset G$, un morphisme de données de Shimura $(H, Y) \rightarrow (G, X)$ et un élément $g \in G(\mathbb{A}_f)$ tels que Z soit contenu dans l'image de l'application (3) de 2.2 et que H soit le groupe de Mumford–Tate générique sur Z . La composante irréductible S de l'image de (3) contenant Z est la plus petite sous-variété de type Hodge de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ contenant Z . L'application (3) se factorise par $\mathrm{Sh}_L(H, Y)_{\mathbb{C}}$ pour un sous-groupe compact ouvert $L \subset H(\mathbb{A}_f)$ et comme ci-dessus, l'application

$$\mathrm{Sh}_L(H, Y)_{\mathbb{C}} \rightarrow \mathrm{Sh}_K(G, X)_{\mathbb{C}}$$

a la propriété qu'une sous-variété irréductible (resp. un point) de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ est de type Hodge (resp. spécial) si et seulement si son image l'est.

PROPOSITION 6.8. — *Soient $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une sous-variété irréductible fermée et Σ un ensemble de points spéciaux, Zariski-dense dans Z . Pour montrer la conjecture 2.3 pour Z , on peut supposer que Z contient un point s avec $\mathrm{MT}(s) = G$ et que la plus petite sous-variété de type Hodge contenant Z est une composante connexe de $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$.*

DÉFINITION 6.9. — *Si $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ vérifie les conditions du théorème on dira que Z est Hodge générique.*

Un énoncé similaire est valable pour le passage au groupe adjoint. Pour toute donnée de Shimura (G, X) on définit X^{ad} comme la $G^{\mathrm{ad}}(\mathbb{R})$ classe de conjugaison dans $\mathrm{Hom}(\mathbb{C}^{\times}, G_{\mathbb{R}}^{\mathrm{ad}})$ contenant l'image de X et $(G^{\mathrm{ad}}, X^{\mathrm{ad}})$ est alors aussi une donnée de Shimura. Il existe des sous-groupes compacts ouverts $K \subset G(\mathbb{A}_f)$ et $K^{\mathrm{ad}} \subset G^{\mathrm{ad}}(\mathbb{A}_f)$ tels que K^{ad} contienne l'image de K . Le morphisme $\mathrm{Sh}_K(G, X)_{\mathbb{C}} \rightarrow \mathrm{Sh}_{K^{\mathrm{ad}}}(G^{\mathrm{ad}}, X^{\mathrm{ad}})_{\mathbb{C}}$ est alors fini et respecte les sous-variétés de type Hodge et les points spéciaux. Cela implique :

PROPOSITION 6.10. — *Soient les notations comme dans la proposition 6.8. Pour montrer la conjecture 2.3 pour Z , on peut remplacer G par son groupe adjoint et Z et Σ par leurs images Z^{ad} et Σ^{ad} dans $\mathrm{Sh}_{K^{\mathrm{ad}}}(G^{\mathrm{ad}}, X^{\mathrm{ad}})$. En outre, Z est Hodge générique si et seulement si Z^{ad} l'est.*

7. L'APPROCHE D'EDIXHOVEN ET YAFAEV

7.1. Énoncés

Dans les différents résultats obtenus par Edixhoven [17], [18], Edixhoven–Yafaev [20] et Yafaev [39], [40], [41] on peut distinguer deux types d'hypothèses. Dans certains cas, on supposera que l'hypothèse de Riemann généralisée (GRH) soit vérifiée pour une classe de corps de nombres qui dépend des variétés de Shimura considérées. Dans d'autres cas, on impose une condition supplémentaire sur les points spéciaux contenus

dans Σ . La stratégie des démonstrations est toutefois très proche d'un cas à l'autre. On cite d'abord les principaux résultats.

THÉORÈME 7.1 (Edixhoven–Yafaev [20], Yafaev [40]). — *Soient (G, X) une donnée de Shimura et $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert. Soit $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une courbe irréductible fermée contenant un ensemble infini Σ de points spéciaux.*

Supposons que tous les groupes de Mumford–Tate $\mathrm{MT}(s)$ (voir définition 2.1) avec $s \in \Sigma$ sont isomorphes. Alors Z est une sous-variété de type Hodge.

Sous une condition un peu plus restrictive sur Σ , le théorème 7.1 est montré dans [20, 1.2]. L'énoncé donné ici est prouvé dans [40, Theorem 1.3]. Le théorème 2.2 de [18] est maintenant le cas particulier du théorème 7.1 où la variété de Shimura est une surface modulaire de Hilbert. C'est l'analogie du théorème 7.2 avec condition sur Σ au lieu de GRH.

L'hypothèse sur les points spéciaux en vigueur dans le théorème ci-dessus peut souvent être remplacée par l'hypothèse de Riemann généralisée. Ainsi le théorème 5.3, qui traite le cas de \mathbb{C}^2 , a été démontré par Edixhoven [17, Theorem 1.1] en admettant GRH pour les corps quadratiques imaginaires. La démonstration dans ce cas a été esquissée dans l'introduction. Le théorème d'Edixhoven a été généralisé par Yafaev [39] à certains autres produits de deux courbes de Shimura. Un deuxième exemple est le théorème suivant qui concerne les surfaces modulaires de Hilbert. On introduit d'abord la notation nécessaire.

Soient F une extension quadratique réelle de \mathbb{Q} et $G = \mathrm{Res}_{F/\mathbb{Q}} \mathrm{GL}_{2/F}$. L'action de \mathbb{C}^\times sur $\mathbb{R}^2 \cong \mathbb{C}$ induit un morphisme $\mathbb{C}^\times \rightarrow \mathrm{GL}_{2/\mathbb{R}}$ dont on déduit

$$h: \mathbb{C}^\times \rightarrow G_{\mathbb{R}} = \mathrm{GL}_{2/\mathbb{R}}^2.$$

Soit X la $G(\mathbb{R})$ -classe de conjugaison de h . Alors (G, X) est une donnée de Shimura et pour $K \subset G(\mathbb{A}_f)$ compact ouvert, on définit la surface modulaire de Hilbert $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ associée à F .

THÉORÈME 7.2 (Edixhoven [18]). — *Supposons que l'hypothèse de Riemann généralisée soit vérifiée. Soient $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une surface modulaire de Hilbert et Z une courbe dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ contenant un ensemble infini de points spéciaux. Alors Z est une sous-variété de type Hodge.*

Remarque 7.3. — Dans la prépublication [40], Yafaev annonce que ce théorème est valable pour toute variété de Shimura. C'est-à-dire que sous l'hypothèse de Riemann généralisée, une courbe $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ contenant un ensemble infini de points spéciaux est une sous-variété de type Hodge, quelle que soit la variété de Shimura $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. La démonstration suit la méthode de la démonstration du théorème 7.1 en l'améliorant sur plusieurs points.

7.2. La stratégie de la démonstration

Les démonstrations des théorèmes cités dans le paragraphe précédent reposent sur la caractérisation suivante des sous-variétés de type Hodge, donnée dans [20, 7.1].

CRITÈRE 7.4. — Soient (G, X) une donnée de Shimura, $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert assez petit et $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ une sous-variété irréductible et fermée. On suppose que G est adjoint et que Z est Hodge générique (dans le sens de la définition 6.9). Comme dans 2.1, une composante X^+ de X est fixée et S désigne l'image de X^+ dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$.

Soient p un nombre premier, $g \in G(\mathbb{Q}_p)$; définissons les $g_i \in G(\mathbb{Q})^+$ comme dans le lemme 6.6, de sorte que les T_{g_i} sont les composantes connexes de T_g sur S . Supposons que

1. les $T_{g_i}Z$ et les $T_{g_i^{-1}}Z$ sont irréductibles,
2. pour tout $s \in S(\mathbb{C})$ et tout g_i , l'orbite

$$(7) \quad \bigcup_{n \geq 0} (T_{g_i} + T_{g_i^{-1}})^n(s)$$

est dense dans S pour la topologie euclidienne et

3. $Z \subset T_g Z$.

Alors $Z = S$, en particulier Z est de type Hodge.

Démonstration. — Supposons que les trois conditions du critère soient vérifiées. Comme $Z \subset S$ et les T_{g_i} sont des correspondances $S \rightarrow S$, on a $Z \subset \bigcup T_{g_i}Z$. Soit i tel que $Z \subset T_{g_i}Z$. La condition 7.4.1 implique alors que $Z = T_{g_i}Z$, que $T_{g_i^{-1}}Z = T_{g_i^{-1}}T_{g_i}Z \supset Z$ et donc que $Z = T_{g_i}Z = T_{g_i^{-1}}Z$. Il s'ensuit que pour $s \in Z(\mathbb{C})$, la $(T_{g_i} + T_{g_i^{-1}})$ -orbite de s est contenue dans Z et 7.4.2 implique alors que $Z = S$. \square

7.3. Les conditions 7.4.1 et 7.4.2

Dans cette section on verra comment on peut établir les deux premières conditions du critère 7.4. Dans les deux cas, (G, X) est une donnée de Shimura avec G semi-simple et adjoint et $K \subset G(\mathbb{A}_f)$ est un sous-groupe compact ouvert. On écrit X^+ , $G(\mathbb{R})^+$, $G(\mathbb{Q})^+$, $\Gamma = K \cap G(\mathbb{Q})^+$ comme dans 2.1 et, comme avant, $S = \Gamma \backslash X^+$ désigne l'image de X^+ dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$. On suppose que K est assez petit pour que Γ opère librement sur X^+ et que K est le produit de sous-groupes compacts ouverts $K_p \subset G(\mathbb{Q}_p)$.

THÉORÈME 7.5 ([20], Theorem 5.1). — Avec les hypothèses ci-dessus, supposons que $Z \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$ soit une sous-variété fermée, irréductible et Hodge générique (dans le sens de la définition 6.9) contenant un point spécial. Il existe alors un entier N_1 tel que, pour tout $g \in G(\mathbb{Q})^+$ dont l'image dans $G(\mathbb{Q}_\ell)$ est dans K_ℓ pour tout ℓ divisant N_1 , l'image $T_g Z$ soit irréductible. Ici T_g est la correspondance de Hecke sur S définie par g .

Esquisse de la démonstration. — On reprend les notations de 6.2 pour la construction de T_g et on considère l'équation (6) avec $S_1 = S_2 = S = \Gamma \backslash X^+$ et $S_g = \Gamma_g \backslash X^+$ où $\Gamma_g = \Gamma \cap g^{-1} \Gamma g$. La correspondance T_g étant l'image de S_g dans $S \times S$, il suffit de montrer que, sous les conditions du théorème, l'image réciproque Z_g de Z dans S_g est connexe. Pour le faire, on peut remplacer (et on remplace) Z et Z_g par leurs sous-variétés de points lisses.

On fixe un élément $z \in Z(\mathbb{C})$ dont le groupe de Mumford–Tate est égal à G . Le revêtement $Z_g \rightarrow Z$ correspond au $\pi_1(Z, z)$ -ensemble $\Gamma_g \backslash \Gamma$ et il suffit de montrer que, toujours sous les conditions du théorème, l'action de $\pi_1(Z, z)$ sur $\Gamma_g \backslash \Gamma$ est transitive. Pour le faire, on construit une variation de \mathbb{Z} -structures de Hodge sur S dont on considère la fibre Λ_z en z . Un théorème d'André [3, Proposition 2] implique que l'adhérence de Zariski de l'image de la représentation de monodromie de $\pi_1(Z, z)$ sur $\Lambda_z \otimes \mathbb{Q}$ (le *groupe de monodromie algébrique*) coïncide avec l'adhérence de Zariski de l'image de la représentation de $\pi_1(S, z)$ sur $\Lambda_z \otimes \mathbb{Q}$.

On invoque ensuite un théorème de Nori [30, 5.3] qui assure l'existence d'un entier N_1 tel que pour tout m premier à N_1 , les images de $\pi_1(Z, z)$ et de $\pi_1(S, z)$ dans $\mathrm{GL}(\Lambda_z/m\Lambda_z)$ coïncident. Tout cela permet de montrer que si g vérifie les hypothèses de 7.5, avec N_1 comme ci-dessus, alors l'action de $\pi_1(Z, z)$ est transitive, prouvant ainsi l'énoncé. La démonstration détaillée attend le lecteur curieux dans [20, §5]. \square

THÉORÈME 7.6. — *On conserve les notations fixées avant le théorème 7.5 et on décompose $G = G_1 \times \cdots \times G_r$ comme produit de facteurs simples. Il existe un entier N_2 tel que pour tout nombre premier $p \geq N_2$ l'énoncé suivant soit vérifié. Soit $g \in G(\mathbb{Q}_p)$ un élément dont la projection dans aucun $G_j(\mathbb{Q}_p)$ ne soit contenue dans un sous-groupe compact et soient les $g_i \in G(\mathbb{Q})^+$ comme dans le lemme 6.6. Alors pour tout i et tout $s \in S(\mathbb{C})$, la $(T_{g_i} + T_{g_i^{-1}})$ -orbite de s (donnée par (7)) est dense dans S pour la topologie euclidienne.*

La démonstration est donnée dans [20, §6].

7.4. Orbites galoisiennes des points spéciaux

Soient (G, X) une donnée de Shimura avec G adjoint et $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert. On fixe une représentation fidèle \mathbb{Q} -linéaire V de G et un réseau $V_{\mathbb{Z}} \subset V$ tel que K stabilise $V_{\mathbb{Z}} \otimes \widehat{\mathbb{Z}} \subset V \otimes \mathbb{A}_f$. Pour tout sous-groupe algébrique $H \subset G$, on note $H_{\mathbb{Z}}$ l'adhérence de H dans $\mathrm{GL}(V_{\mathbb{Z}})$. Si $T \subset G$ est un tore, $\mathrm{Mauvais}(T)$ est l'ensemble de nombres premiers p tels que la fibre de $T_{\mathbb{Z}}$ en caractéristique p ne soit pas un tore.

Soit $R \subset G(\mathbb{A}_f)$ un ensemble de représentants (dont 1_G) pour l'ensemble (fini) des doubles classes $G(\mathbb{Q}) \backslash G(\mathbb{A}_f) / K$. Pour tout $s \in \mathrm{Sh}_K(G, X)(\mathbb{C})$ on fixe un représentant $(h_s, a_s) \in X \times G(\mathbb{A}_f)$ avec $a_s \in R$.

THÉORÈME 7.7. — *Avec les notations introduites ci-dessus, soient $F \subset \mathbb{C}$ un corps de nombres contenant le corps dual $E(G, X)$ et $\mathrm{Sh}_K(G, X)$ le modèle canonique sur F de la variété de Shimura $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$.*

Soit $s_0 \in \text{Sh}_K(G, X)(\overline{\mathbb{Q}})$ un point spécial. Il existe alors des constantes c_1 et $c_2 > 0$ telles que pour tout $s \in \text{Sh}_K(G, X)(\overline{\mathbb{Q}})$ avec $\text{MT}(s) \cong \text{MT}(s_0)$ on ait l'inégalité

$$|\Gamma_F \cdot s| > c_1 \prod_{p \in \text{Mauvais}(\text{MT}(h_s))} c_2 p.$$

Remarque 7.8. — Noter que les tores $\text{MT}(h_s) \subset G$ dépendent du choix des h_s et que les ensembles $\text{Mauvais}(\text{MT}(h_s))$ dépendent en plus du choix de la représentation V et du réseau $V_{\mathbb{Z}}$.

Remarque 7.9. — L'amélioration du théorème 7.7 (sous l'hypothèse de Riemann généralisée) est un des ingrédients principaux du travail de Yafaev [40] menant au résultat annoncé dans la remarque 7.3.

Esquisse de la démonstration. — Remarquons pour commencer que pour montrer le théorème, on peut modifier K et F à volonté. De plus, il suffit de prouver le théorème avec chaque s remplacé par une image sous la correspondance $T_{a_s^{-1}}$ car R étant fini, le degré des $T_{a_s^{-1}}$ est uniformément borné. Cela permet de supposer que $a_s = 1_G$ pour tout s et que chaque s est dans l'image de $\text{Sh}(\text{MT}(h_s), \{h_s\}) \rightarrow \text{Sh}_K(G, X)$.

Dans la proposition suivante, Σ_{s_0} désigne l'ensemble des $s \in \text{Sh}_K(G, X)(\overline{\mathbb{Q}})$ avec $\text{MT}(s) \cong \text{MT}(s_0)$. Via la représentation $G \rightarrow \text{GL}(V)$, on considère les h_s comme des applications $\mathbb{C}^\times \rightarrow \text{GL}(V_{\mathbb{R}})$.

PROPOSITION 7.10 (Yafaev [40]). — *L'ensemble des $h_s: \mathbb{C}^\times \rightarrow \text{GL}(V_{\mathbb{R}})$ avec $s \in \Sigma_{s_0}$ est une réunion finie de $\text{GL}(V)$ -classes de conjugaison.*

Démonstration. — Comme G est adjoint, chaque morphisme $h \in X$ est déterminé par le morphisme

$$\mu_{h/\mathbb{C}} = h \circ r: \mathbb{G}_{m/\mathbb{C}} \rightarrow \mathbb{C}^\times \rightarrow G_{\mathbb{C}} \subset \text{GL}(V_{\mathbb{C}})$$

défini comme dans 6.1, formule (4). Il suffit donc de montrer que l'ensemble des $\text{GL}(V)$ -classes de conjugaison dans l'ensemble des μ_{h_s} est fini, pour s parcourant Σ_{s_0} . Tous les μ_{h_s} , pour $s \in \text{Sh}_K(G, X)(\mathbb{C})$, sont conjugués sous $G(\mathbb{C})$, donc l'ensemble des poids de μ_{h_s} sur $V_{\mathbb{C}}$ ne dépend pas de s .

Pour tout tore T sur \mathbb{Q} , notons $X^*(T)$ (resp. $X_*(T)$) le groupe de (co)caractères de T , muni de l'action naturelle à gauche du groupe de Galois absolu $\Gamma_{\mathbb{Q}}$. L'accouplement naturel $\langle \cdot, \cdot \rangle: X_*(T) \times X^*(T) \rightarrow \mathbb{Z}$ vérifie $\langle x^\sigma, y^\sigma \rangle = \langle x, y \rangle$ pour tout $\sigma \in \Gamma_{\mathbb{Q}}$. Le groupe $\Gamma_{\mathbb{Q}}$ opère sur tous les $\text{MT}(s)$ avec $s \in \Sigma_{s_0}$ via le même quotient fini $\overline{\Gamma}_{\mathbb{Q}}$.

Soit $s \in \Sigma_{s_0}$. Comme $\text{MT}(s)$ est le groupe de Mumford–Tate de s , les conjugués galoisiens de μ_s engendrent $X_*(\text{MT}(s)) \otimes \mathbb{Q}$ et on en déduit une surjection de $\Gamma_{\mathbb{Q}}$ -modules (à gauche) $\mathbb{Q}[\overline{\Gamma}_{\mathbb{Q}}] \rightarrow X_*(\text{MT}(s)) \otimes \mathbb{Q}$. Cela donne une injection $X^*(\text{MT}(s)) \subset \mathbb{Z}[\overline{\Gamma}_{\mathbb{Q}}]$ de $\Gamma_{\mathbb{Q}}$ -modules. Soit $\Pi_s \subset X^*(\text{MT}(s)) \subset \mathbb{Z}[\overline{\Gamma}_{\mathbb{Q}}]$ l'ensemble des caractères intervenant dans la représentation de $\text{MT}(s)_{\overline{\mathbb{Q}}}$ sur $V_{\overline{\mathbb{Q}}}$. Comme Π_s est $\Gamma_{\mathbb{Q}}$ -invariant, on a

$$\langle \mu_{h_s}^\sigma, \Pi_s \rangle = \langle \mu_{h_s}^\sigma, \Pi_s^\sigma \rangle = \langle \mu_{h_s}, \Pi_s \rangle$$

pour tout $\sigma \in \Gamma_{\mathbb{Q}}$. Comme $\langle \mu_{h_s}, \Pi_s \rangle$ est l'ensemble des poids de μ_{h_s} agissant sur $V_{\mathbb{C}}$, celui-ci ne dépend pas de s . Les $\mu_{h_s}^{\sigma}$ constituent une base de $X_*(\text{MT}(s)) \otimes \mathbb{Q}$, donc cela ne laisse qu'un nombre fini de possibilités pour Π_s (pour s parcourant Σ_{s_0}).

Pour montrer que l'ensemble des classes de $\text{GL}(V)$ -conjugaison des μ_{h_s} est fini, on peut maintenant supposer que Π_s ne dépend pas de s . Comme Π_s engendre $X^*(\text{MT}(s))$ en tant que sous-module de $\mathbb{Z}[\overline{\Gamma}_{\mathbb{Q}}]$, il s'ensuit que $X^*(\text{MT}(s)) \subset \mathbb{Z}[\overline{\Gamma}_{\mathbb{Q}}]$ ne dépend pas de s . On obtient une identification des tores $\text{MT}(s)$ qui, par construction, est compatible avec les cocaractères μ_{h_s} . En passant à un sous-ensemble infini, on peut supposer que la multiplicité de chaque $\chi \in \Pi_s$ dans la représentation de $\text{MT}(s)_{\overline{\mathbb{Q}}}$ sur $V_{\overline{\mathbb{Q}}}$ est indépendante de s . Les tores $\text{MT}(s)$ sont alors conjugués dans $\text{GL}(V)$ et il en est de même pour les μ_{h_s} . \square

Soit Σ'_{s_0} l'ensemble des s tels que h_s soit le conjugué de h_{s_0} par un élément $t_s \in \text{GL}(V)$. Grâce à la proposition, il suffit de prouver le théorème 7.7 pour s parcourant Σ'_{s_0} . Pour tout élément $s \in \Sigma'_{s_0}$, la conjugaison par t_s est un isomorphisme de données de Shimura $(\text{MT}(h_{s_0}), \{h_{s_0}\}) \cong (\text{MT}(h_s), \{h_s\})$ et les corps duaux des $(\text{MT}(h_s), \{h_s\})$ coïncident avec le même corps $E \subset \mathbb{C}$. Les isomorphismes des données de Shimura sont compatibles avec les lois de réciprocité (définies dans 6.1, formule (5)).

On contemple le diagramme commutatif suivant (d'applications entre ensembles)

$$\begin{array}{ccc} \text{Sh}(\text{MT}(h_{s_0}), \{h_{s_0}\})(\mathbb{C}) & \longrightarrow & \text{Sh}_K(G, X)(\mathbb{C}) \\ \text{Sh}(\text{inn}_{t_s}) \uparrow & & \downarrow \\ \text{Sh}(\text{MT}(h_s), \{h_s\})(\mathbb{C}) & \xrightarrow{f_s} & \text{GL}_n(\mathbb{Q}) \backslash \left(Y \times \text{GL}_n(\mathbb{A}_f) / \text{GL}_n(\widehat{\mathbb{Z}}) \right). \end{array}$$

La flèche $\text{Sh}(\text{inn}_{t_s})$ vient de l'isomorphisme entre modèles canoniques, les points complexes des variétés dans la colonne de gauche sont $\overline{\mathbb{Q}}$ -rationnels et $\text{Sh}(\text{inn}_{t_s})$ est Γ_E -équivariant. La première flèche horizontale vient aussi d'un morphisme entre modèles canoniques et on voit assez facilement (cf. [20, 4.2]) que, pour prouver le théorème, il suffit de donner une borne inférieure pour l'ordre de la $(\widehat{\mathbb{Z}} \otimes \mathcal{O}_E)^{\times}$ -orbite de s dans $\text{Sh}_K(G, X)(\mathbb{C})$, pour l'action de $(\widehat{\mathbb{Z}} \otimes \mathcal{O}_E)^{\times}$ donnée par $r(\text{MT}(h_s), \{h_s\})$.

En ce qui concerne l'ensemble $\text{GL}_n(\mathbb{Q}) \backslash (Y \times \text{GL}_n(\mathbb{A}_f) / \text{GL}_n(\widehat{\mathbb{Z}}))$, on a identifié $V \cong \mathbb{Q}^n$, noté par $\rho: G \rightarrow \text{GL}_n/\mathbb{Q}$ le morphisme déduit de la représentation de G sur V et défini Y comme la classe de conjugaison de morphismes $\mathbb{C}^{\times} \rightarrow \text{GL}_n/\mathbb{R}$ contenant l'image de X . Cela définit la deuxième colonne du diagramme et permet de définir f_s comme l'application rendant le carré commutatif. Cette application est déduite de la composée $\rho|_{\text{MT}(h_s)} \circ \text{inn}_{t_s}$ en passant au quotient. Dans [20, §4] l'estimation de l'orbite de s provient d'une minoration de l'ordre de l'orbite de l'image de s dans $\text{GL}_n(\mathbb{Q}) \backslash (Y \times \text{GL}_n(\mathbb{A}_f) / \text{GL}_n(\widehat{\mathbb{Z}}))$ sous l'action de $(\widehat{\mathbb{Z}} \otimes \mathcal{O}_E)^{\times}$ via la loi de réciprocité $r(\text{MT}(h_s), h_s)$. On se ramène au calcul de l'orbite de t_s dans $\text{GL}_n(\mathbb{A}_f) / \text{GL}_n(\widehat{\mathbb{Z}})$ sous l'action de $(\widehat{\mathbb{Z}} \otimes \mathcal{O}_E)^{\times}$ donnée par $r(\text{MT}(h_{s_0}), h_{s_0})^m$ pour un entier m assez grand. Ceci est achevé par un calcul long mais explicite qui occupe [20, 4.3 et 4.4]. \square

7.5. Fin de la démonstration du théorème 7.1

Les différentes propositions dans 6.3 permettent de se ramener au cas où le groupe G est adjoint, Z est Hodge générique et K est le produit de sous-groupes compacts ouverts $K_p \subset G(\mathbb{Q}_p)$. En appliquant une correspondance de Hecke, on peut en outre supposer que

$$(8) \quad Z \subset S \subset \mathrm{Sh}_K(G, X)_{\mathbb{C}}$$

avec S l'image dans $\mathrm{Sh}_K(G, X)$ d'une composante connexe fixée X^+ de X . On fixe, pour tout $s \in \Sigma$, un élément $h_s \in X^+$ tel que $s = \overline{(h_s, 1)} \in \mathrm{Sh}_K(G, X)_{\mathbb{C}}(\mathbb{C})$. Le lemme 6.1 montre que les variétés et les inclusions de l'équation (8) sont définies sur un corps de nombres $F \subset \mathbb{C}$ et on en fixe un modèle sur F . On supposera aussi que tous les tores $\mathrm{MT}(s)$, pour $s \in \Sigma$, sont décomposés sur F . Comme dans le paragraphe 7.4 on fixe une représentation fidèle \mathbb{Q} -linéaire V de G et un réseau $V_{\mathbb{Z}} \subset V$ de sorte que l'on peut prendre l'adhérence $H_{\mathbb{Z}}$ de tout sous-groupe $H \subset G$ et définir $\mathrm{Mauvais}(T)$ pour tout tore $T \subset G$ comme dans 7.4. Dans la suite de la démonstration, on note $m(s)$ l'ordre de l'ensemble $\mathrm{Mauvais}(\mathrm{MT}(h_s))$ et on distingue deux possibilités.

Le cas où $m(s)$ est borné. — Pour tout nombre premier, p soit Σ_p l'ensemble des $s \in \Sigma$ tels que $\mathrm{MT}(h_s)_{\mathbb{F}_p}$ soit un tore. Supposons que $m(s) < B$ pour tout $s \in \Sigma$. Pour tout ensemble E de nombres premiers avec $|E| = B$, on a alors $\cup_{p \in E} \Sigma_p = \Sigma$ et pour au moins un $p \in E$ l'ensemble Σ_p est donc dense dans Z . Il y a donc au plus $B - 1$ nombres premiers tels que Σ_p ne soit pas Zariski-dense. Il s'ensuit qu'il existe une infinité de nombres premiers p supérieurs à N_1 (comme dans le théorème 7.5), et à N_2 (du théorème 7.6), avec $G_{\mathbb{F}_p}$ lisse, $K_p = G_{\mathbb{Z}}(\mathbb{Z}_p)$ et $\Sigma_p \subset Z$ Zariski-dense et tels que tous les $\mathrm{MT}(h_s)_{\mathbb{Q}_p}$ (pour $s \in \Sigma$) soient scindés. Seule la dernière condition pourrait exclure un ensemble infini de nombres premiers, mais le théorème de Chebotarev implique qu'elle est vérifiée pour un ensemble de p avec une densité positive.

En utilisant [20, Proposition 7.3.1] on se ramène au cas où les tores $\mathrm{MT}(h_s)_{\mathbb{Z}_p}$ (pour $s \in \Sigma_p$) sont conjugués sous K_p . On fixe un élément $s_0 \in \Sigma_p$ et, pour tout $s \in \Sigma_p$, un transporteur $t_s \in K_p$ tel que $\mathrm{MT}(h_s) = t_s \mathrm{MT}(h_{s_0}) t_s^{-1}$. Dans [20, 7.3], il est montré qu'il existe $g \in \mathrm{MT}(h_{s_0})(\mathbb{Q}_p)$ dans l'image de Γ_F (via le morphisme de réciprocité) vérifiant les conditions du théorème 7.6. La démonstration utilise les faits que $r(\mathrm{MT}(h_{s_0}), \{h_{s_0}\})$ est une surjection (de groupes algébriques) et que $\mathrm{MT}(h_{s_0})_{\mathbb{Q}_p}$ est scindé.

Pour tout $s \in \Sigma_p$, le conjugué $\overline{t_s g t_s^{-1}} \in \mathrm{MT}(h_s)$ est dans l'image de Γ_F par le morphisme de réciprocité donc $\overline{(h_s, t_s g t_s^{-1})} \in \Gamma_F \cdot s$. Comme on a également $\overline{(h_s, t_s g t_s^{-1})} = \overline{(h_s, t_s g)} \in T_g s$, on conclut que $\Gamma_F \cdot s$ rencontre $T_g Z$. Comme $T_g Z$ est définie sur F , cela implique que $s \in T_g Z$.

On a montré que $\Sigma_p \subset T_g Z(\mathbb{C})$ et cela implique que $Z \subset T_g Z$. Les conditions du critère 7.4 étant vérifiées, le théorème est démontré dans ce premier cas. \square

Remarque 7.11. — Les résultats de la section 7.4 concernant les orbites galoisiennes n'ont pas été utilisés dans le cas qui vient d'être traité. On ne s'est pas non plus servi de l'hypothèse que Z soit une courbe.

Le cas où $m(s)$ n'est pas borné. — Pour les détails le lecteur est renvoyé à [20, 7.4]. En utilisant le corollaire 6.5 et une étude de l'action des tores $\text{MT}(s)_{\mathbb{Z}_p}$ sur $V_{\mathbb{Z}_p}$ on montre le lemme suivant, voir [20, Corollary 7.4.4].

LEMME 7.12. — *Il existe un entier k avec la propriété suivante. Soient $s \in \Sigma$ et p un nombre premier tels que $\text{MT}(h_s)_{\mathbb{Q}_p}$ soit scindé et $p \notin \text{Mauvais}(\text{MT}(h_s))$. Alors il existe un $g \in \text{MT}(h_s)(\mathbb{Q}_p)$ appartenant à $r(\text{MT}(h_s), \{h_s\})(\mathbb{Q}_p \otimes F)$, vérifiant les conditions du théorème 7.6 et tel que $|T_g Z \cap Z| \leq p^k$ si cette intersection est finie.*

Soit k comme dans le lemme. Comme $m(s)$ devient arbitrairement grand, le théorème de Chebotarev implique l'existence d'un élément $s \in \Sigma$ et d'un nombre premier p vérifiant les conditions du lemme, avec $p \geq N_1$ et N_2 (comme dans les théorèmes 7.5 et 7.6) et tels que la minoration de $|\Gamma_F \cdot s|$ établie dans le théorème 7.7 soit supérieure à p^k . Pour l'élément g fourni par le lemme ci-dessus, le fait que g appartient à l'image de Γ_F assure que $T_g s$ contient un conjugué de s sous Γ_F et donc que $\Gamma_F \cdot s \subset T_g Z \cap Z$. Il s'ensuit que l'intersection $T_g Z \cap Z$ ne peut pas être finie et comme Z est irréductible on en déduit que $Z \subset T_g Z$. L'élément g vérifie les conditions du critère 7.4 et la démonstration du théorème est achevée. \square

7.6. Esquisse de la démonstration du théorème 7.2

Rappelons les notations en vigueur dans le théorème 7.2. On a fixé une extension quadratique réelle F de \mathbb{Q} , posé $G = \text{Res}_{F/\mathbb{Q}} \text{GL}_{2/F}$, construit un morphisme

$$h: \mathbb{C}^\times \rightarrow G_{\mathbb{R}} = \text{GL}_{2/\mathbb{R}}$$

et défini X comme la $G(\mathbb{R})$ classe de conjugaison de h . Le groupe G admet un modèle naturel sur \mathbb{Z} , à savoir $G_{\mathbb{Z}} = \text{Res}_{\mathcal{O}_F/\mathbb{Z}} \text{GL}_{2/\mathcal{O}_F}$ et on pose $K = G_{\mathbb{Z}}(\widehat{\mathbb{Z}})$. La variété de Shimura $\text{Sh}_K(G, X)_{\mathbb{C}}$ admet alors un modèle canonique sur \mathbb{Q} . On se donne un corps de nombres F et une courbe irréductible fermée $Z \subset \text{Sh}_K(G, X)_F$ contenant un ensemble infini Σ de points spéciaux. Il faut montrer que Z est de type Hodge.

Dans [18] on travaille avec la donnée de Shimura (G, X) et avec une donnée de Shimura (G', X') associée à un sous-groupe $G' \subset G$. Ici on essayera de se ramener autant que possible à la méthode exposée ci-dessus. On passe donc à la donnée de Shimura adjointe $(G^{\text{ad}}, X^{\text{ad}})$ en remarquant que celle-ci est aussi la donnée de Shimura adjointe de (G', X') . Le groupe G^{ad} hérite d'un modèle entier $G_{\mathbb{Z}}^{\text{ad}}$ et on prend $K^{\text{ad}} = G_{\mathbb{Z}}^{\text{ad}}(\widehat{\mathbb{Z}})$. L'image de Z dans $\text{Sh}_{K^{\text{ad}}}(G^{\text{ad}}, X^{\text{ad}})$ est notée Z^{ad} . Si Z^{ad} n'est pas Hodge générique, alors le théorème 7.2 est trivialement valable et on supposera dans la suite qu'on se trouve dans le cas restant : $Z^{\text{ad}} \subset \text{Sh}_{K^{\text{ad}}}(G^{\text{ad}}, X^{\text{ad}})$ est Hodge générique. Il suffit de montrer que Z^{ad} est une composante de $\text{Sh}_{K^{\text{ad}}}(G^{\text{ad}}, X^{\text{ad}})$, car cela implique que Z^{ad} est de type Hodge. Alternativement, si Z^{ad} est une composante de $\text{Sh}_{K^{\text{ad}}}(G^{\text{ad}}, X^{\text{ad}})$, cela contredit le fait que Z^{ad} soit une courbe. En tout cas, le théorème sera prouvé.

On a $G(\mathbb{Q}_p) = \mathrm{GL}_2(K \otimes_{\mathbb{Q}} \mathbb{Q}_p)$. Le critère 7.4 sera appliqué à la correspondance de Hecke T_p^{ad} associée à l'image dans $G^{\mathrm{ad}}(\mathbb{A}_f)$ d'un élément de la forme

$$\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(K \otimes \mathbb{Q}_p) = G(\mathbb{Q}_p) \subset G(\mathbb{A}_f),$$

avec p premier. Les théorèmes 7.5 et 7.6 impliquent l'existence de constantes N_1 et N_2 tels que les conditions 7.4.1 et 7.4.2 soient vérifiées pour $p \geq N = \max(N_1, N_2)$.

Pour que la condition 7.4.3 soit vérifiée, on utilise la même idée que ci-dessus, c'est-à-dire qu'on choisira p tel que l'intersection $T_p^{\mathrm{ad}}Z^{\mathrm{ad}} \cap Z^{\mathrm{ad}}$ ne puisse pas être propre. Si cette intersection est propre, alors la borne supérieure pour $|T_p^{\mathrm{ad}}Z^{\mathrm{ad}} \cap Z^{\mathrm{ad}}|$ fournie par le corollaire 6.5 est de la forme $c_1 p^k$ pour une constante $c_1 > 0$ et un entier k .

Il reste à trouver une minoration $|T_p^{\mathrm{ad}}Z^{\mathrm{ad}} \cap Z^{\mathrm{ad}}|$ et pour cela on suivra [18]. En particulier, on se place à nouveau dans $\mathrm{Sh}_K(G, X)_F$ où il suffit maintenant de trouver une estimation de $|T_p Z \cap Z|$. La variété $\mathrm{Sh}_K(G, X)$ peut être interprétée comme espace de modules de variétés abéliennes munies d'une injection $\mathcal{O}_K \subset \mathrm{End}(A)$ vérifiant certaines conditions. Pour tout $s \in \mathrm{Sh}_K(G, X)$, on note R_s l'anneau $\mathrm{End}_{\mathcal{O}_K}(A_s)$ où A_s est la variété abélienne correspondant à s . Par des arguments similaires à ceux utilisés dans le paragraphe 7.5, Edixhoven montre dans [18, Lemma 8.1] que si $s \in Z(\overline{\mathbb{Q}})$ est un point spécial et p un nombre premier tels que R_s soit scindé au-dessus de p , on a $\Gamma_F \cdot s \subset T_p Z \cap Z$.

Le théorème 7.7 trouve un analogue dans [18, Theorem 6.2]⁴ qui donne une estimation $|\Gamma_F \cdot s| > c_2 d_s$ où $d_s = |\mathrm{discr}(R_s)|^{1/4-\varepsilon}$ et $c_2 > 0$. On peut montrer que, pour $s \in \Sigma$, le nombre d_s devient arbitrairement grand. L'hypothèse de Riemann généralisée implique une version effective du théorème de Chebotarev dont on déduit que, pour d_s assez grand, il existe des nombres premiers p tels que $R_s \otimes_{\mathbb{F}_p} \cong \mathbb{F}_p^n$ et vérifiant une estimation de valeur polynomiale en $\log d_s$. Toujours pour d_s assez grand on en déduit l'existence d'un nombre premier p tel que $c_2 d_s > c_1 p^2$. Pour un tel p on a $Z \subset T_p Z$ et le critère 7.4 implique alors le théorème.

8. SOUS-VARIÉTÉS DE DIMENSION SUPÉRIEURE

8.1. Travaux de Moonen et de Yafaev

Le premier cas particulier de la conjecture d'André–Oort qui ait été prouvé est dû à Moonen, voir le paragraphe 5.1. Ses résultats sont valables pour des sous-variétés de dimension arbitraire dans les variétés de Shimura. Les énoncés prouvés par la méthode d'Edixhoven et Yafaev traités jusqu'ici sont limités au cas des courbes contenues dans les variétés de Shimura. Toutefois, en généralisant la démonstration du théorème 7.1,

⁴ Historiquement, le théorème 7.7 est postérieur à [18].

en particulier le cas où la fonction $m(s)$ est bornée, et en adaptant les théorèmes 7.5 et 7.6, Yafaev [41] prouve la généralisation suivante du théorème 5.2.

THÉORÈME 8.1. — *Soient (G, X) une donnée de Shimura (G, X) et $K \subset G(\mathbb{A}_f)$ un sous-groupe compact ouvert. On fixe une représentation fidèle \mathbb{Q} -linéaire V de G muni d'un réseau $V_{\mathbb{Z}} \subset V$ et, en utilisant cette représentation, on identifie G avec un sous-groupe fermé de GL_n/\mathbb{Q} .*

Soit $\Sigma_X \subset X$ un ensemble de points spéciaux. Supposons qu'il existe un nombre premier p vérifiant la condition que pour tout $s \in \Sigma_X$ il existe un sous-tore $M_s \cong \mathbb{G}_{m/\mathbb{Q}_p}$ de $\mathrm{MT}(s)_{\mathbb{Q}_p}$ tel que

- *l'ensemble des poids de M_s agissant sur $V_{\mathbb{Q}_p}$ est uniformément borné en s ,*
- *pour aucun quotient non trivial T de $\mathrm{MT}(s)$, l'image de M_s dans $T_{\mathbb{Q}_p}$ ne soit triviale et*
- *l'adhérence de M_s dans $\mathrm{GL}_n/\mathbb{Z}_p$ est isomorphe à $\mathbb{G}_{m/\mathbb{Z}_p}$.*

Alors pour tout $g \in G(\mathbb{A}_f)$ les composantes irréductibles de l'adhérence de l'image de $\Sigma_X \times \{g\}$ dans $\mathrm{Sh}_K(G, X)_{\mathbb{C}}$ sont de type Hodge.

8.2. Le cas de \mathbb{C}^n

Signalons encore que la méthode d'André esquissée dans le paragraphe 5.2 permet de montrer la conjecture d'André–Oort inconditionnellement pour les courbes dans $\mathbb{A}_{\mathbb{C}}^n$. Dans [19] Edixhoven adapte sa méthode au cas d'une sous-variété de dimension quelconque de $\mathbb{A}_{\mathbb{C}}^n$ contenant un sous-ensemble dense de points spéciaux.

Dans la section précédente, le fait que la sous-variété Z est une courbe est essentiel dans la fin de la démonstration : pour appliquer le critère 7.4 il faut que $Z \subset T_g Z$, pour une correspondance de Hecke convenable T_g . Pour montrer cette inclusion dans le cas où Z est une courbe irréductible, il suffit de montrer que l'intersection ne peut pas être finie. On verra ici, de manière succincte, comment l'approche d'Edixhoven [17] peut être généralisée pour des sous-ensembles de points spéciaux dans $\mathbb{A}_{\mathbb{C}}^n$, vu comme produit n copies de la courbe modulaire. Dans la suite on note, comme dans 5.2, \mathbb{A}^1 la variété de modules (grossière) des courbes elliptiques. On a alors les trois résultats suivants, dus à Edixhoven [19].

THÉORÈME 8.2. — *Soit $\Sigma \subset \mathbb{A}^n(\mathbb{C})$ un ensemble de points spéciaux. Supposons que l'hypothèse de Riemann généralisée soit vérifiée. Alors les composantes irréductibles de l'adhérence de Σ sont de type Hodge.*

THÉORÈME 8.3. — *Soit $\Sigma \subset \mathbb{A}^n(\mathbb{C})$ un ensemble de points spéciaux tels que les variétés abéliennes (toutes produits de courbes elliptiques) correspondantes soient toutes isogènes. Alors les composantes irréductibles de l'adhérence de Σ sont de type Hodge.*

D'après la proposition 6.7, ces théorèmes impliquent :

COROLLAIRE 8.4. — *Soit (GL_2, Y) la donnée de Shimura considérée en 5.2. Alors les énoncés ci-dessus restent valables si chaque facteur \mathbb{A}^1 est remplacé par une courbe modulaire $X_i = \mathrm{Sh}_{K_i}(\mathrm{GL}_2, Y)$, pour des sous-groupes compacts ouverts $K_i \subset \mathrm{GL}_2(\mathbb{A}_f)$.*

Dans [19], les énoncés sont formulés en termes d'une description explicite des sous-variétés de type Hodge. Pour donner cette description, on décrit d'abord deux types particuliers : Un point de $\mathbb{A}_{\mathbb{C}}^1$ est de type Hodge si et seulement si c'est un point spécial. Une courbe dans $\mathbb{A}_{\mathbb{C}}^n$ est de type Hodge si et seulement si c'est une composante connexe de l'image de $\mathbb{A}_{\mathbb{C}}^1$ par un produit $(T_{g_1}, \dots, T_{g_n}) : \mathbb{A}_{\mathbb{C}}^1 \rightarrow \mathbb{A}_{\mathbb{C}}^n$ de correspondances de Hecke associées à des éléments $g_i \in \mathrm{GL}_2(\mathbb{Q})$. En général, une sous-variété de $\mathbb{A}_{\mathbb{C}}^n$ est de type Hodge si et seulement si, après une permutation des facteurs, c'est un produit de sous-variétés des deux types précédents. Cette description est utilisée pour établir la caractérisation des sous-variétés de $\mathbb{A}_{\mathbb{C}}^n$ de type Hodge donnée dans la proposition 8.6.

DÉFINITION 8.5. — *Pour $I \subset \{1, \dots, n\}$, soit $p_I : \mathbb{A}^n \rightarrow \mathbb{A}^I$ la projection sur les facteurs indexés par I . Soit $Z \subset \mathbb{A}_{\mathbb{C}}^n$ une sous-variété fermée irréductible. Alors un sous-ensemble $I \subset \{1, \dots, n\}$ est minimal pour Z si $\dim p_I Z < |I|$ mais $\dim p_J Z = |J|$ pour tout $J \subsetneq I$.*

PROPOSITION 8.6. — *Soit $Z \subset \mathbb{A}_{\mathbb{C}}^n$ une sous-variété fermée irréductible. Alors Z est de type Hodge si et seulement si, pour tout $I \subset \{1, \dots, n\}$ qui est minimal pour Z , on a $|I| \leq 2$ et $\overline{p_I Z}$ est de type Hodge.*

On passe au résumé de la démonstration des deux théorèmes. On fixe un ensemble de points spéciaux $\Sigma \subset \mathbb{A}^n(\mathbb{C})$ et on définit Z comme l'adhérence de Σ . On peut supposer que Z est irréductible. Pour tout entier m , notons T_m la correspondance de Hecke associée à l'élément $\begin{pmatrix} m & 0 \\ 0 & 1 \end{pmatrix}^n \in \mathrm{GL}_2(\mathbb{A}_f)^n$. Une des innovations principales de [19] est le théorème 8.7.

THÉORÈME 8.7. — *Si Σ vérifie les conditions d'un des théorèmes 8.2 ou 8.3, alors pour tout $s \in \Sigma$, sauf un nombre fini, il existe une courbe $C \subset Z$ de type Hodge avec $s \in C(\mathbb{C})$.*

Esquisse de la démonstration. — L'idée centrale est de couper Z avec son image $T_p Z$ pour un nombre premier p convenable. Comme dans la dernière partie de la démonstration du théorème 7.1, présentée dans 7.5, on doit choisir p de sorte que $Z_1 = Z \cap T_p Z$ contienne s et que l'orbite galoisienne de s soit assez grande pour pouvoir déduire par un argument d'intersection que Z_1 est de dimension strictement positive.

Si $\dim Z_1 = \dim Z$, alors $Z \subset T_p Z$ et on montre que Z est de type Hodge en appliquant [19, Theorem 4.1] qui est un analogue de la combinaison des théorèmes 7.5 et 7.6 avec le critère 7.4. Le théorème 8.7 est évident dans ce cas. Si $\dim Z_1 < \dim Z$, alors on répète l'opération. On ne sait pas si Z_1 contient un sous-ensemble dense de points

spéciaux, mais l'orbite de s sous l'action du groupe de Galois suffit pour appliquer les arguments ci-dessus. \square

Les théorèmes 8.2 et 8.3 sont prouvés en appliquant la proposition 8.6. Clairement, le théorème 8.7 implique que si I est minimal pour Z et $|I| \leq 2$, alors $\overline{\rho_I Z}$ est de type Hodge.

Le fait qu'il n'existe pas de sous-ensemble $I \subset \{1, \dots, n\}$ d'ordre au moins 3 qui est minimal pour Z est un peu plus difficile à montrer, voir [19, §9].

RÉFÉRENCES

- [1] A. Abbes. Hauteurs et discrétude (d'après L. Szpiro, E. Ullmo et S. Zhang). *Sém. Bourbaki* (1996/97), Exp. N° 825, *Astérisque* 245, p. 141–166, 1997.
- [2] Y. André. *G-functions and geometry*, tome E13 de *Aspects of Mathematics*. Vieweg, 1989.
- [3] Y. André. Mumford–Tate groups of mixed Hodge structures and the theorem of the fixed part. *Compos. Math.* 82(1), p. 1–24, 1992.
- [4] Y. André. Distribution des points CM sur les sous-variétés des variétés de modules de variétés abéliennes. Prépublication 120, Institut Mathématique de Jussieu, 1997.
- [5] Y. André. Finitude des couples d'invariants modulaires singuliers sur une courbe algébrique plane non modulaire. *J. reine Angew. Math.* 505, p. 203–208, 1998.
- [6] Y. André. Shimura varieties, subvarieties, and CM points, August–September 2001. Six lectures at the Franco–Taiwan arithmetic festival, <http://www.math.umd.edu/~yu/notes.shtml>.
- [7] F. Breuer. La conjecture d'André–Oort pour le produit de deux courbes modulaires de Drinfeld. *C. R. Math. Acad. Sci. Paris* 335(11), p. 867–870, 2002.
- [8] L. Clozel et E. Ullmo. Correspondances modulaires et mesures invariantes. *J. reine Angew. Math.* 558, p. 47–83, 2003.
- [9] L. Clozel et E. Ullmo. Équidistribution de sous-variétés spéciales, 2004. À paraître dans *Ann. of Math.*
- [10] L. Clozel et E. Ullmo. Équidistribution des points de Hecke. Dans *Contributions to automorphic forms, geometry, and number theory*, pages 193–254. Johns Hopkins Univ. Press, 2004.
- [11] P. Cohen et G. Wüstholz. Application of the André–Oort conjecture to some questions in transcendence. Dans *A panorama of number theory or the view from Baker's garden (Zürich, 1999)*, édité par G. Wüstholz, pages 89–106. Cambridge Univ. Press, 2002.
- [12] R. Coleman. Torsion points on curves. Dans *Galois representations and arithmetic algebraic geometry*, édité par Y. Ihara, tome 12 de *Adv. Stud. Pure Math.*, pages 235–247, 1987.

- [13] C. Cornut. Non-trivialité des points de Heegner. *C. R. Acad. Sci. Paris, Sér. I*, 334(12), p. 1039–1042, 2002.
- [14] J. de Jong et R. Noot. Jacobians with complex multiplication. Dans *Arithmetic algebraic geometry*, édité par G. van der Geer, F. Oort, et J.H.M. Steenbrink, tome 89 de *Progr. Math.*, pages 177–192. Birkhäuser, 1991.
- [15] P. Deligne. Travaux de Shimura. Sémin. Bourbaki (1970/71), Exp. N° 389, *Lecture Notes in Math.* 244, Springer-Verlag, p. 141–166, 1971.
- [16] P. Deligne. Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques. Dans *Automorphic forms, representations, and L-functions*, édité par A. Borel et W. Casselman, tome XXXIII, Part 2 de *Proc. Sympos. Pure Math.*, pages 247–289. Amer. Math. Soc., 1979.
- [17] S.J. Edixhoven. Special points on the product of two modular curves. *Compos. Math.* 114(3), p. 315–328, 1998.
- [18] S.J. Edixhoven. On the André–Oort conjecture for Hilbert modular surfaces. Dans *Moduli of abelian varieties (Texel Island, 1999)*, édité par G. van der Geer, C. Faber, et F. Oort, tome 195 de *Progr. Math.*, pages 133–155. Birkhäuser, 2001.
- [19] S.J. Edixhoven. Special points on products of modular curves. e-print, arXiv: math.NT/0302138, 2003. À paraître dans *Duke Math. J.*, Version 2, 26 Oct. 2004, <http://www.arxiv.org/abs/math.NT/0302138v2>.
- [20] S.J. Edixhoven et A. Yafaev. Subvarieties of Shimura varieties. *Ann. of Math.* 157(2), p. 621–645, 2003.
- [21] M. Hindry. Points de torsion sur les sous-variétés de variétés abéliennes. *C. R. Acad. Sci. Paris Sér. I Math.* 304(12), p. 311–314, 1987.
- [22] M. Hindry. Autour d’une conjecture de Serge Lang. *Invent. Math.* 94(3), p. 575–603, 1988.
- [23] D.-H. Jiang, J.-S. Li, et S.-W. Zhang. Periods and distribution of cycles on Hilbert modular varieties. Preprint, 2004. <http://www.math.columbia.edu/~szhang/papers/jlz05.pdf>.
- [24] J.S. Milne. Lettre à Deligne, 28–03–1990. <http://www.jmilne.org/math/Manuscripts/svd.pdf>.
- [25] B.J.J. Moonen. *Special points and linearity properties of Shimura varieties*. Thèse de doctorat, Universiteit Utrecht, 1995.
- [26] B.J.J. Moonen. Linearity properties of Shimura varieties, I. *J. Algebraic Geom.* 7, p. 539–567, 1998.
- [27] B.J.J. Moonen. Linearity properties of Shimura varieties, II. *Compos. Math.* 114, p. 3–35, 1998.
- [28] B.J.J. Moonen. Models of Shimura varieties in mixed characteristics. Dans *Galois representations in arithmetic algebraic geometry (Durham, 1996)*, édité par A.J. Scholl et R.L. Taylor, tome 254 de *London Math. Soc. Lecture Note Ser.*, pages 267–350. Cambridge Univ. Press, 1998.

- [29] R. Noot. Models of Shimura varieties in mixed characteristic. *J. Algebraic Geom.* 5(1), p. 187–207, 1996.
- [30] M.V. Nori. On subgroups of $GL_n(\mathbf{F}_p)$. *Invent. Math.* 88(2), p. 257–275, 1987.
- [31] F. Oort. Some questions in algebraic geometry, 1995.
<http://www.math.uu.nl/people/oort/>.
- [32] F. Oort. Canonical liftings and dense sets of CM-points. Dans *Arithmetic geometry (Cortona, 1994)*, tome XXXVII de *Sympos. Math.*, pages 228–234. Cambridge Univ. Press, 1997.
- [33] R. Pink. A combination of the conjectures of Mordell–Lang and André–Oort. Preprint, ETH Zürich, 2004. <http://www.math.ethz.ch/~pink/preprints.html>.
- [34] M. Raynaud. Courbes sur une variété abélienne et points de torsion. *Invent. Math.* 71(1), p. 207–233, 1983.
- [35] M. Raynaud. Sous-variétés d’une variété abélienne et points de torsion. Dans *Arithmetic and geometry, Vol. I*, édité par M. Artin et J. Tate, tome 35 de *Progr. Math.*, pages 327–352. Birkhäuser, 1983.
- [36] J-P. Serre. *Cours d’arithmétique*. Presses Universitaires de France, 1970.
- [37] E. Ullmo. Théorie ergodique et géométrie arithmétique. Dans *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, pages 197–206. Higher Ed. Press, 2002.
- [38] J. Wolfart. Werte hypergeometrischer Funktionen. *Invent. Math.* 92(1), p. 187–216, 1988.
- [39] A. Yafaev. Special points on products of two Shimura curves. *Manuscripta Math.* 104(2), p. 163–171, 2001.
- [40] A. Yafaev. A conjecture of Yves André. e-print, arXiv:math.NT/0302125, 2003. Version 1, 11 Feb. 2003, <http://www.arxiv.org/abs/math.NT/0302125v1>.
- [41] A. Yafaev. On a result of Ben Moonen on the moduli space of principally polarised abelian varieties, 2005. À paraître dans *Compos. Math.*
- [42] S.-W. Zhang. Equidistribution of CM-points on quaternion Shimura varieties. Preprint, 2004. <http://www.math.columbia.edu/~szhang/papers/ecm4.pdf>.

Rutger NOOT

IRMA

Université Louis Pasteur et CNRS

7 rue René Descartes

F-67084 Strasbourg

E-mail : rutger.noot@math.u-strasbg.fr

Web : <http://www-irma.u-strasbg.fr/~noot/>

CLASSES DE COHOMOLOGIE POSITIVES DANS LES VARIÉTÉS KÄHLÉRIENNES COMPACTES

[d'après Boucksom, Demailly, Nakayama, Păun, Peternell...]

par Olivier DEBARRE

1. INTRODUCTION

Soit X une variété⁽¹⁾ complexe compacte. On dit qu'une forme différentielle ω de type $(1, 1)$ sur X est définie positive si la forme hermitienne h définie par

$$h(x, y) = \omega(x, iy) - i\omega(x, y)$$

l'est, c'est-à-dire si elle définit une métrique hermitienne sur X . Lorsque ω est de plus fermée, on dit que h définit une métrique de Kähler sur X . La classe de cohomologie de Dolbeault de ω dans $H^{1,1}(X, \mathbf{R})$ est dite classe de Kähler, et ces classes forment un premier cône⁽²⁾ convexe ouvert $\text{Kah}(X)$ dans l'espace vectoriel réel de dimension finie $H^{1,1}(X, \mathbf{R})$. Si L est un fibré en droites holomorphe sur X , la courbure d'une métrique hermitienne sur L est une forme différentielle réelle fermée de type $(1, 1)$ dont un multiple représente la première classe de Chern de L . Pour que cette classe soit une classe de Kähler, il faut et il suffit que L soit *ample*.

Mais les éléments de $H^{1,1}(X, \mathbf{R})$ sont aussi les classes des *courants* réels fermés de type $(1, 1)$ sur X , pour lesquels on a une notion similaire, mais plus faible, de positivité (cf. § 3.1). Les classes correspondantes forment un second cône convexe fermé, plus grand, dans $H^{1,1}(X, \mathbf{R})$; on le note $\text{Pef}_{\text{tr}}(X)$. Lorsqu'on a affaire à la première classe de Chern d'un fibré en droites holomorphe sur X , cette seconde notion de positivité correspond à la positivité du courant de courbure d'une métrique singulière sur L , ou encore, du point de vue algébrique, au fait que les puissances tensorielles positives de L ont « beaucoup » de sections holomorphes⁽³⁾.

L'étude, lorsque X est une variété kählérienne compacte, de la géométrie de ces deux cônes est l'objet de cet exposé.

Les premières classes de Chern des fibrés en droites holomorphes sur X engendrent un sous-espace vectoriel de $H^{1,1}(X, \mathbf{R})$, qui peut être strict; on le note $N^1(X, \mathbf{R})$. Ses

⁽¹⁾Toutes nos variétés seront toujours, sauf mention explicite du contraire, supposées lisses.

⁽²⁾Un cône est une partie stable par multiplication par \mathbf{R}^{+*} .

⁽³⁾La positivité est en fait un peu plus faible que cette propriété. La différence entre les deux est essentielle et source de moult difficultés.

éléments sont dits classes « algébriques ». Les traces de nos deux cônes sur $N^1(X, \mathbf{R})$ sont notées $\text{Amp}(X)$ (cône *ample*) et $\text{Pef}(X)$ (cône *pseudo-effectif*) respectivement. Lorsque X est projective, le cône dual de $\text{Amp}(X)$ dans l'espace vectoriel dual $N^1(X, \mathbf{R})^\vee$ est le cône fermé engendré par les classes de courbes algébriques de X , aussi dit « cône de Mori » de X , dont on sait qu'il reflète une grande partie de la géométrie de X .

Ces cônes ont été étudiés à l'aide de techniques de géométrie algébrique (qui ne donnent accès qu'aux classes algébriques, mais sont souvent aussi valables sur les variétés singulières) et, plus récemment, de techniques d'analyse complexe, qui permettent de traiter toutes les classes, sur les variétés lisses. Pour donner un exemple, on sait depuis longtemps caractériser de façon numérique (c'est-à-dire en termes de nombres d'intersection) les classes amples : c'est le célèbre critère de Nakai (valable sur tous les schémas propres). Ce n'est que récemment que la caractérisation analogue des classes de Kähler sur les variétés kählériennes compactes (lisses) a été démontrée par Demailly et Păun (§ 2.4).

Un autre problème, de nature algébrique celui-ci, résolu récemment est la description du cône dual de $\text{Pef}(X)$ dans $N^1(X, \mathbf{R})^\vee$: Boucksom, Demailly, Păun et Peternell ont montré qu'il est engendré par les classes de courbes mobiles, c'est-à-dire dont les déformations recouvrent X (§ 4.2). La démonstration fait intervenir plusieurs constructions qu'il m'a paru utile de détailler.

La première est celle de « volume » d'une classe. Pour les classes algébriques, par exemple pour la première classe de Chern d'un fibré en droites L sur X , il s'agit essentiellement d'une mesure asymptotique de la dimension de l'espace des sections des puissances tensorielles positives de L (*cf.* § 5.1 et § 5.3 pour l'extension aux classes transcendantes). On définit ainsi une fonction continue $\text{vol} : H^{1,1}(X, \mathbf{R}) \rightarrow \mathbf{R}^+$ dont la nature reste mystérieuse.

Le second ingrédient est la décomposition de Zariski. Celui-ci exprime, lorsque X est une *surface* algébrique, toute classe algébrique comme la somme d'une classe nef (limite d'amples) de même volume et d'une classe effective dite négative, orthogonale à la partie positive (§ 5.2). Cette décomposition a des implications si importantes qu'on a longtemps—et vainement—tenté de l'étendre en dimension supérieure. Nakayama a montré que ce n'est pas possible, même si l'on permet des modifications de X .

On peut néanmoins montrer l'existence d'une décomposition « approchée » (th. 5.3 et 5.4) qui suffit pour démontrer la caractérisation du cône dual de $\text{Pef}(X)$ mentionnée plus haut. Il s'agit d'une décomposition *sur une modification de X* en la somme d'une classe ample de volume proche et d'une classe effective, qui est alors automatiquement « presque orthogonale » (en un sens convenable) à la première.

Je termine en décrivant dans le § 7 une autre extension du résultat de Zariski en dimension supérieure, due à Nakayama pour les classes algébriques, puis à Boucksom en général : la décomposition de Zariski divisorielle. Il s'agit d'une décomposition en la somme d'une classe nef « en codimension 1 » de même volume et d'une classe « exceptionnelle ».

Les applications de ces résultats sont multiples. Même si de nombreux exemples montrent que la situation en dimension supérieure est infiniment plus compliquée que ce qui se passe sur les surfaces, Huybrechts et Boucksom ont montré, comme conséquence des résultats et techniques présentés ici, que tout se passe admirablement bien pour les variétés hyperkähleriennes (définies dans l'exemple 2.4), où une forme bilinéaire définie par Beauville et Bogomolov joue le rôle du produit d'intersection sur les surfaces. Nous expliquons leurs résultats sous forme d'une série d'exemples.

Voici une autre conséquence importante du théorème de dualité (cor. 4.3) : pour qu'une variété projective ne soit pas recouverte par des courbes rationnelles, il faut et il suffit que sa classe canonique soit dans l'adhérence du cône effectif. On s'attend alors à ce qu'un multiple de la classe canonique soit effectif. C'est peut-être le moment de parler de cette classe, qui est la grande absente de ces notes. Mori et Kawamata ont su montrer l'importance et la spécificité de cette classe et l'utiliser avec brio dans leurs théorèmes du cône et de rationalité. De nombreux exemples exhibent des classes pathologiques, mais on pense (ou on espère) que la classe canonique devrait se comporter raisonnablement. Beaucoup de travail reste à faire dans cette direction !

La littérature qui concerne les questions abordées ici étant en général bien écrite, il m'a paru inutile de la recopier ici (à deux exceptions près). J'ai préféré donner à ces notes la forme d'un survol, en présentant les deux points de vue, algébrique et analytique, l'un après l'autre, le tout assaisonné de nombreux exemples.

Je remercie pour leurs précieux conseils Arnaud Beauville, Sébastien Boucksom, Jean-Pierre Demailly, Daniel Huybrechts, Robert Lazarsfeld, Mihai Păun et Mihnea Popa.

Dans tout l'exposé, X est une variété complexe compacte de dimension n .

2. CÔNE AMPLE ET CÔNE DE KÄHLER

2.1. Quelques définitions

Soit L un fibré en droites sur X (toujours sous-entendu *holomorphe*) ; on note $H^0(X, L)$ l'espace vectoriel de ses sections holomorphes. Le fibré L est ample s'il existe un plongement de X dans un espace projectif \mathbf{P}^N tel que $\mathcal{O}_{\mathbf{P}^N}(1)$ se restreint à X en un puissance tensorielle positive de L . L'amplitude est une propriété numérique : elle ne dépend que de la première classe de Chern $c_1(L) \in H^2(X, \mathbf{R})$ (cf. § 2.4).

Un *diviseur* sur X est une combinaison linéaire formelle $D = \sum_E t_E E$, à coefficients entiers, d'hypersurfaces analytiques irréductibles de X . Il est dit *effectif* lorsque tous les coefficients sont positifs ; on écrit alors $D \geq 0$. Il est dit *premier* si une seule hypersurface irréductible apparaît dans D et qu'elle a coefficient 1 ; on confond alors souvent D avec cette hypersurface. On considérera aussi des \mathbf{Q} -diviseurs (on permet des coefficients rationnels), et même des \mathbf{R} -diviseurs. On note $[D]$ le diviseur obtenu en prenant la partie entière de ses coefficients, c'est-à-dire $\sum_E [t_E] E$; le *support* de D est la réunion

des hypersurfaces irréductibles de X affectées d'un coefficient non nul dans D . Les composantes irréductibles d'un diviseur sont par définition celles de son support.

Toute fonction méromorphe f sur X a un diviseur (celui de ses pôles et zéros), noté $\text{div}(f)$; les diviseurs de ce type sont dits *principaux*. Deux diviseurs sont *linéairement équivalents* si leur différence est principale.

À tout diviseur D sur X , on associe un fibré en droites sur X , noté L_D , dont les sections locales sont les fonctions méromorphes f dont le diviseur satisfait $\text{div}(f) + D \geq 0$. Inversement, toute section méromorphe non nulle s d'un fibré en droites L a un diviseur D_s , et L est isomorphe à L_{D_s} . On dit que D est *ample* si L_D l'est. On notera $H^0(X, D)$ au lieu de $H^0(X, L_D)$. Si D' est un autre diviseur, $L_{D'}$ est isomorphe à L_D si et seulement si D' est linéairement équivalent à D . L'ensemble $|D|$ des diviseurs linéairement équivalents à D s'identifie à l'espace projectif $\mathbf{P}(H^0(X, D))$.

Tout diviseur D a une classe de cohomologie $[D] \in H^2(X, \mathbf{R})$, et $c_1(L_D) = [D]$. Noter que l'addition des diviseurs se traduit par le produit tensoriel des fibrés en droites associés, et l'opposé par le dual.

2.2. Classes amples et classes de Kähler

Une métrique hermitienne h sur X est donnée en coordonnées locales (z_1, \dots, z_n) par

$$h = \sum_{j,k} h_{jk} dz_j \otimes d\bar{z}_k$$

où (h_{jk}) est une matrice hermitienne de fonctions locales de classe \mathcal{C}^∞ à valeurs dans \mathbf{C} , définie positive en chaque point. On lui associe la forme différentielle réelle $\omega = -\text{Im}(h)$, de type $(1, 1)$, donnée localement par

$$\omega = i \sum_{j,k} h_{jk} dz_j \wedge d\bar{z}_k$$

Une forme de ce type est dite *définie positive*. On dit que la métrique h (ou la 2-forme ω) est une métrique *de Kähler* si ω est une forme différentielle *fermée*. S'il existe une telle métrique sur la variété X , celle-ci est dite *kählérienne*. Via la théorie de Hodge, toute forme de Kähler ω a une classe de cohomologie de de Rham⁽⁴⁾

$$[\omega] \in H^{1,1}(X, \mathbf{R}) \subset H^2(X, \mathbf{R})$$

L'ensemble des classes de (formes de) Kähler est un cône convexe ouvert

$$\text{Kah}(X) \subset H^{1,1}(X, \mathbf{R})$$

⁽⁴⁾Une grande partie des constructions de cet exposé peuvent être faites sur une variété complexe compacte générale, pas nécessairement kählérienne, au prix de complications techniques. L'espace de cohomologie de Dolbeault $H^{1,1}(X, \mathbf{R})$ est remplacé par

$$H_{\partial\bar{\partial}}^1(X, \mathbf{R}) = \{ \text{formes de type } (1, 1) \text{ réelles } d\text{-fermées} \} / \{ \text{formes de type } (1, 1) \text{ réelles } \partial\bar{\partial}\text{-exactes} \}$$

Je renvoie à [Bo4] et [Bo3] les lecteurs intéressés.

On fait maintenant intervenir la structure entière de $H^2(X, \mathbf{R})$. Le groupe abélien libre de type fini

$$N^1(X) = H^{1,1}(X, \mathbf{R}) \cap (\text{Im}(H^2(X, \mathbf{Z}) \rightarrow H^2(X, \mathbf{R})))$$

est appelé groupe de Néron–Severi de X . Le théorème de Hodge nous dit que c’est l’ensemble des premières classes de Chern des fibrés en droites sur X .

Si X est une sous-variété d’un espace projectif (on dit qu’elle est *projective*), la restriction à X de la forme de Fubini–Study est une forme de Kähler sur X dont la classe est entière, c’est-à-dire dans $N^1(X)$. Inversement, si $N^1(X)$ contient une classe de Kähler, le théorème de plongement de Kodaira entraîne que la variété X est projective (donc algébrique par le théorème de Chow). Plus précisément, si $N^1(X, \mathbf{R})$ est le sous-espace vectoriel réel de $H^{1,1}(X, \mathbf{R})$ engendré par $N^1(X)$, le cône

$$\text{Amp}(X) = \text{Kah}(X) \cap N^1(X, \mathbf{R})$$

est le cône engendré par les classes (de diviseurs) *amples* sur X .

2.3. Classes nef

Si X est projective, on montre que les éléments de $\overline{\text{Amp}}(X)$ sont les classes de fibrés en droites *numériquement effectifs* (ou nef) sur X , c’est-à-dire ceux dont le degré sur toute courbe est positif. Un fibré en droites nef L a des propriétés cohomologiques asymptotiques particulières : il vérifie le théorème de Fujita ([Lz1], Theorem 1.4.40)

$$h^i(X, L^{\otimes m}) = O(m^{n-i}) \quad \text{pour tout } i \geq 0,$$

qui entraîne, avec le théorème de Riemann–Roch,

$$(1) \quad h^0(X, L^{\otimes m}) = \frac{c_1(L)^n}{n!} m^n + O(m^{n-1})$$

Par extension, lorsque X n’est que kählérienne, les éléments de $\overline{\text{Kah}}(X)$ sont aussi dits nef. L’estimation (1) reste valable ; c’est une conséquence des inégalités de Morse holomorphes de Demailly ([De1]).

En toute généralité, on dira qu’une classe $\xi \in H^{1,1}(X, \mathbf{R})$ est nef si, une forme réelle ω de type $(1, 1)$ définie positive (non nécessairement fermée) sur X étant fixée, il existe pour tout $\varepsilon > 0$ une forme ω_ε dans ξ , réelle de type $(1, 1)$, telle que $\omega_\varepsilon + \varepsilon\omega$ est positive.

2.4. Caractérisation numérique du cône de Kähler

Le critère de Nakai–Moishezon–Kleiman (cf. [Lz1], 1.2.B, pour une démonstration et un historique) caractérise numériquement le cône ample : un fibré en droites L sur une variété projective X est ample si et seulement si

$$c_1(L)^{\dim(Y)} \cdot [Y] > 0 \quad \text{pour toute sous-variété algébrique } Y \text{ de } X.$$

La caractérisation correspondante des classes amples de $N^1(X, \mathbf{R})$ (non nécessairement entières) est un joli théorème de Campana et Peternell ([CP]).

Les deux résultats suivants généralisent ces critères aux classes transcendantales.

THÉORÈME 2.1 (Demailly–Păun, [DP]). — Soit X une variété kählérienne compacte. Le cône $\text{Kah}(X)$ est une composante connexe de l'ensemble des classes $\xi \in H^{1,1}(X, \mathbf{R})$ qui vérifient

$$(2) \quad \int_Y \xi^{\dim(Y)} > 0 \quad \text{pour toute sous-variété analytique } Y \text{ de } X.$$

La classe d'une forme de Kähler ω vérifie ces propriétés : $\int_X \omega^n$ n'est autre que le volume de la variété X pour la métrique associée, il est donc strictement positif. De même la restriction de ω à une sous-variété analytique Y de X est encore une forme de Kähler, de sorte que $\int_Y \omega^{\dim(Y)}$ est aussi strictement positif.

SCHÉMA DE PREUVE — On veut montrer que le cône de Kähler est fermé dans l'ensemble des classes vérifiant la propriété (2). On utilise quelques concepts de la théorie des courants définis dans le § 3.

Soit ω une forme de Kähler sur X . Étant donnée une sous-variété analytique Y de X de codimension p , on construit tout d'abord une famille $(\omega_\varepsilon)_{\varepsilon>0}$ de formes de Kähler dans la même classe que ω , qui a la propriété que tout courant T limite faible de (ω_ε^p) quand ε tend vers 0 vérifie $T \geq cT_Y$, avec $c > 0$.

Soit ξ une forme réelle de type $(1, 1)$ dont la classe est dans l'adhérence de $\text{Kah}(X)$, donc nef, et qui vérifie $\int_X \xi^n > 0$.

Le théorème de Yau montre que l'on peut résoudre l'équation de Monge–Ampère

$$(\xi + \varepsilon\omega + i\partial\bar{\partial}\varphi_\varepsilon)^n = c_\varepsilon\omega_\varepsilon^n$$

dès que $c_\varepsilon = \int_X (\xi + \varepsilon\omega)^n / \int_X \omega_\varepsilon^n$. La forme $\alpha_\varepsilon = \xi + \varepsilon\omega + i\partial\bar{\partial}\varphi_\varepsilon$ a même classe que $\xi + \varepsilon\omega$. Le fait que l'on ait $\lim_{\varepsilon \rightarrow 0^+} c_\varepsilon > 0$ permet de montrer que toute limite faible T de (α_ε^p) quand ε tend vers 0, qui est donc dans la classe $[\xi]^p$, vérifie encore $T \geq cT_Y$, avec $c > 0$.

En appliquant cette construction au cas où Y est la diagonale de $X \times X$, on construit un courant de Kähler dans $[\xi]$. On peut dire que le théorème de Yau permet de trouver dans $[\xi]$ un courant strictement positif partout sous une hypothèse de stricte positivité en moyenne.

En raisonnant par récurrence sur n , on montre ensuite que l'hypothèse (2) entraîne que la restriction de $[\xi]$ à toute sous-variété analytique propre de X contient une forme de Kähler. On conclut, en utilisant un résultat antérieur de Păun, que la classe $[\xi]$ elle-même contient une forme de Kähler.

Pour les détails, je renvoie le lecteur à l'article original [DP] ou à [De3]. □

Exemple 2.2 (Tores complexes). — Soit X un tore complexe, c'est-à-dire le quotient de \mathbf{C}^n par un réseau Γ . L'espace $H^{1,1}(X, \mathbf{R})$ s'identifie à l'espace vectoriel réel des formes hermitiennes sur \mathbf{C}^n , et $N^1(X)$ au groupe de celles qui sont entières sur Γ . Le cône $\text{Kah}(X)$ est le cône des formes hermitiennes définies positives.

Lorsque Γ est « très général », on a $N^1(X) = 0$ et X ne contient aucune sous-variété analytique autre qu'elle-même et ses sous-ensembles finis. La condition du théorème se réduit à $\int_X \xi^n > 0$, ce qui signifie que la forme hermitienne associée est non dégénérée.

L'ensemble des classes vérifiant cette propriété n'est donc pas nécessairement connexe, mais il l'est si X est projective, comme on peut le déduire de la variante suivante du théorème.

COROLLAIRE 2.3. — *Soient X une variété kählérienne compacte et ω une forme de Kähler sur X . Le cône $\text{Kah}(X)$ est l'ensemble des classes $\xi \in H^{1,1}(X, \mathbf{R})$ telles que*

$$\int_Y \xi^k \wedge \omega^{\dim(Y)-k} > 0 \quad \text{pour toute sous-variété analytique } Y \\ \text{de } X \text{ et tout } k \in \{1, \dots, \dim(Y)\}.$$

PREUVE — L'hypothèse entraîne que $t\omega + \xi$ vérifie la condition du théorème 2.1 pour tout $t \geq 0$. Sa classe reste donc dans la même composante connexe. Comme c'est une forme de Kähler pour $t \gg 0$, il en est de même pour ξ . □

Buchdahl ([Bu], cor. 15) et Lamari ([L1], [L2]), indépendamment, avaient auparavant démontré ce critère pour toutes les surfaces complexes compactes, même non kählériennes, en utilisant des métriques de Gauduchon (c'est-à-dire des métriques dont la $(1, 1)$ forme associée ω vérifie $\partial\bar{\partial}\omega^{n-1} = 0$).

Exemple 2.4 (Variétés hyperkähleriennes). — La variété X est dite *hyperkählienne* si elle est compacte, kählérienne, simplement connexe, et que l'ensemble des 2-formes holomorphes sur X est de dimension 1, engendré par une forme σ qui est non dégénérée en chaque point ([B], [H1]). Ce sont les analogues en toute dimension (paire) des surfaces $K3$. L'espace vectoriel $H^2(X, \mathbf{R})$ est muni d'une forme bilinéaire non dégénérée q_X dite de Beauville–Bogomolov qui généralise le produit d'intersection pour une surface $K3$ ⁽⁵⁾. Elle est entière sur $H^2(X, \mathbf{Z})$ et sa signature sur $H^{1,1}(X, \mathbf{R})$ est $(1, h^{1,1}(X) - 1)$. Le cône $\text{Kah}(X)$ est contenu dans

$$\{\xi \in H^{1,1}(X, \mathbf{R}) \mid q_X(\xi, \xi) > 0\}$$

donc dans l'une de ses deux composantes connexes ; on la note $\mathcal{C}(X)$.

- **Critère de projectivité.** Huybrechts déduit du théorème ci-dessus le critère suivant ([H1])

$$X \text{ projective} \iff \exists \xi \in N^1(X) \quad q_X(\xi, \xi) > 0 \iff \mathcal{C}(X) \cap N^1(X, \mathbf{R}) \neq \emptyset.$$

Le point essentiel consiste à montrer qu'un fibré en droites L sur X vérifiant $q_X(c_1(L), c_1(L)) > 0$ satisfait la propriété (3) ci-dessous, ce qui entraîne que X est projective (cf. § 3.4).

⁽⁵⁾Sur $H^{1,1}(X, \mathbf{R})$, elle est définie par $q_X(\xi, \xi') = \int_X \xi \wedge \xi' \wedge (\sigma \wedge \bar{\sigma})^{n/2-1}$, où σ est normalisée de façon convenable.

- **Cône de Kähler.** On a ([H2], [Bo1])

$$\text{Kah}(X) = \{ \xi \in \mathcal{C}(X) \mid \int_C \xi > 0 \text{ pour toute courbe } C \text{ rationnelle sur } X \}$$

Lorsque X est une surface $K3$ algébrique, la structure du cône $\text{Pef}(X)$ est très bien connue ([Kov]).

Déformation du cône de Kähler

Je me contente de citer le résultat suivant ([DP], Theorem 5.1), qui dit que dans une famille de variétés kählériennes compactes, le cône de Kähler d’une fibre très générale est « constant ».

THÉORÈME 2.5. — *Soit $X \rightarrow S$ une famille de variétés kählériennes compactes. Pour s très général dans $S^{(6)}$, les cônes $\text{Kah}(X_s) \subset H^{1,1}(X_s, \mathbf{R})$ sont invariants par transport parallèle pour la composante de type $(1, 1)$ de la connexion de Gauss–Manin.*

3. CÔNES PSEUDO-EFFECTIFS

3.1. Courants réels de type $(1, 1)$

Un courant sur X est une forme différentielle dont les coefficients sont des distributions. Un courant T est dit *fermé* si $dT = 0^{(7)}$. Il a alors une classe $[T]$ dans $H^\bullet(X, \mathbf{R})^{(8)}$. Un courant de type $(1, 1)$

$$T = i \sum_{j,k} T_{j,k} dz_j \wedge d\bar{z}_k$$

est dit *positif* si la distribution $\sum_{j,k} \lambda_j \bar{\lambda}_k T_{j,k}$ est une mesure positive pour tous complexes $\lambda_1, \dots, \lambda_n$.

Un courant fermé T de type $(1, 1)$ s’écrit localement $i\partial\bar{\partial}\varphi$, où φ est une fonction réelle (et il est positif si et seulement si φ est pluri-sous-harmonique). Cela permet de définir son image inverse par une application holomorphe *surjective* $\mu : X' \rightarrow X$, en posant localement $\mu^*T = i\partial\bar{\partial}(\varphi \circ \mu)$ (c’est bien indépendant du choix de φ).

Toute hypersurface analytique D de X définit un courant

$$T_D : \gamma \longmapsto \int_D \gamma$$

qui est positif fermé de type $(1, 1)$. Cette définition s’étend par linéarité à tout \mathbf{R} -diviseur. On a $[T_D] = [D]$ dans $H^{1,1}(X, \mathbf{R})$.

⁽⁶⁾C’est-à-dire pour s en dehors de la réunion d’une famille dénombrable de sous-variétés analytiques de S distinctes de S .

⁽⁷⁾Comme pour les distributions, dT est le courant défini par $\langle dT, \gamma \rangle = -\langle T, d\gamma \rangle$.

⁽⁸⁾On montre que les groupes de cohomologie de de Rham ou de Dolbeault peuvent être définis soit à partir de formes différentielles, soit à partir de courants.

3.2. Métriques sur les fibrés en droites

Soit L un fibré en droites sur X . Dans une carte locale, une métrique hermitienne h sur L s'écrit

$$h(x, t) = |t|^2 e^{-2\varphi(x)} \quad \text{avec } x \in X, t \in \mathbf{C},$$

où φ (le « poids ») est une fonction de classe \mathcal{C}^∞ . La relation

$$\Theta_h(L) = \frac{i}{\pi} \partial \bar{\partial} \varphi$$

définit une forme fermée réelle de type $(1, 1)$ sur X qui est la courbure de h ; sa classe de cohomologie de de Rham est $c_1(L)$.

On parlera de *métrique singulière* sur L lorsque φ n'est que localement intégrable. On définit par la même formule son courant de courbure; sa classe est encore $c_1(L)$. Tout courant fermé réel de type $(1, 1)$ dans $c_1(L)$ qui peut s'écrire comme la somme d'une forme réelle de type $(1, 1)$ et d'un courant positif est le courant de courbure d'une métrique singulière sur L .

Le théorème de Kodaira dit que

- L est ample si et seulement s'il admet une métrique lisse à courbure définie positive.

Demailly démontre, en utilisant des estimées L^2 ([De2]; [De3], Theorem 1.2), que

- L est pseudo-effectif si et seulement s'il admet une métrique singulière à courant de courbure positif.

Attention, un fibré nef n'admet pas toujours de métrique *lisse* à courbure positive⁽⁹⁾ !

3.3. Classes effectives et pseudo-effectives

On note $\text{Eff}(X) \subset N^1(X, \mathbf{R})$ le cône convexe engendré par les classes des hypersurfaces analytiques de X et $\text{Pef}(X)$ son adhérence, souvent appelé « cône pseudo-effectif »⁽¹⁰⁾.

Pour définir l'analogue transcendant de ce cône, on s'inspire de la discussion du § précédent : une classe dans $H^{1,1}(X, \mathbf{R})$ est dite pseudo-effective si c'est la classe d'un courant fermé de type $(1, 1)$ positif. On appelle *cône pseudo-effectif transcendant* le cône convexe fermé

$$\text{Pef}_{\text{tr}}(X) \subset H^{1,1}(X, \mathbf{R})$$

des classes pseudo-effectives. Si X est projective, on a

$$\text{Pef}_{\text{tr}}(X) \cap N^1(X, \mathbf{R}) = \text{Pef}(X)$$

⁽⁹⁾C'est le cas par exemple du fibré associé à la section d'auto-intersection nulle de la surface réglée non triviale au-dessus d'une courbe elliptique ([DPS]). Il n'y a pas ce genre de problème pour les pseudo-effectifs à cause de la compacité faible des courants positifs de classe bornée.

⁽¹⁰⁾La terminologie n'est pas très heureuse, mais il semble malheureusement trop tard pour la changer.

On a ainsi un diagramme

$$\begin{array}{ccccc} \text{Amp}(X) & \subset & \text{Pef}(X) & \subset & N^1(X, \mathbf{R}) \\ \cap & & \cap & & \cap \\ \text{Kah}(X) & \subset & \text{Pef}_{\text{tr}}(X) & \subset & H^{1,1}(X, \mathbf{R}) \end{array}$$

où les carrés sont cartésiens.

Exemple 3.1 (Fibrés projectifs sur une courbe). — Soient E un fibré vectoriel de rang au moins 2 sur une courbe projective C et $\pi : \mathbf{P}(E) \rightarrow C$ le fibré projectif associé. Les espaces vectoriels $H^{1,1}(X, \mathbf{R})$ et $N^1(X, \mathbf{R})$ coïncident et sont de dimension 2 : ils sont engendré par la classe d'une section et celle d'une fibre de π . Cette dernière engendre une face commune des cônes plans $\text{Nef}(X)$ et $\text{Pef}(X)$.

Miyaoka a montré que ces cônes coïncident si et seulement si le fibré vectoriel E est semi-stable.

Exemple 3.2 (Surfaces). — Lorsque X est une surface kählérienne, on a à notre disposition le produit d'intersection sur $H^2(X, \mathbf{R})$. Le cône $\{\xi \in H^{1,1}(X, \mathbf{R}) \mid \xi^2 > 0\}$ a deux composantes connexes. On note $\mathcal{C}(X)$ celle qui contient $\text{Kah}(X)$. On a⁽¹¹⁾

$$\overline{\mathcal{C}(X)} \subset \text{Pef}_{\text{tr}}(X)$$

Exemple 3.3 (Conjecture de Nagata). — Soit X_r l'éclaté de \mathbf{P}^2 en r points très généraux. Soient E la somme des diviseurs exceptionnels et H l'image inverse d'une droite. On pose

$$t_{\text{nef}} = \min\{t \mid -E + tH \text{ nef}\} \quad , \quad t_{\text{pef}} = \min\{t \mid -E + tH \text{ pseudo-effectif}\}$$

On a $t_{\text{pef}}t_{\text{nef}} = r$ et $t_{\text{pef}} \leq \sqrt{r} \leq t_{\text{nef}}$. Nagata conjecture dans [N] que l'on a égalité pour tout $r \geq 9$.

La conjecture est démontrée lorsque r est un carré parfait. Les valeurs de t_{pef} et t_{nef} sont connues pour $r \leq 9$; elles sont rationnelles. McDuff et Polterovich ont montré dans [McP] que ce problème était intimement lié au problème des empilements symplectiques, ce qui leur permet de déterminer, pour $r \leq 9$, le meilleur taux de remplissage de la variété symplectique \mathbf{P}^2 à l'aide de r boules symplectiques standard de même rayon. Pour $r \geq 9$, ce problème a été résolu dans [Bi1] (*cf.* aussi [Bi2]).

Exemple 3.4 (Variétés hyperkähleriennes). — Avec les définitions et notations du § 2.4, Huybrechts déduit du théorème 2.1 que $\mathcal{C}(X)$ est contenu dans l'intérieur du cône $\text{Pef}_{\text{tr}}(X)$. On a donc

$$\text{Nef}_{\text{tr}}(X) \subset \overline{\mathcal{C}(X)} \subset \text{Pef}_{\text{tr}}(X)$$

et tous ces cônes sont égaux pour une déformation très générale de X ([H2]).

⁽¹¹⁾Pour les classes algébriques, c'est une conséquence du théorème de Riemann-Roch. Le cas général est démontré dans [Bo3], § 4.2.1.

3.4. Classes grandes

On dit qu'un élément de $H^{1,1}(X, \mathbf{R})$ est *grand* (« big » en anglais) s'il est dans l'intérieur du cône pseudo-effectif. On note

$$\text{Big}_{\text{tr}}(X) = \text{Pef}_{\text{tr}}^{\circ}(X) \subset H^{1,1}(X, \mathbf{R}) \quad \text{et} \quad \text{Big}(X) = \text{Pef}^{\circ}(X) \subset N^1(X, \mathbf{R})$$

les cônes correspondants. La terminologie provient du fait que (la première classe de Chern d') un fibré en droites L est grand si et seulement s'il existe une constante $c > 0$ telle que

$$(3) \quad h^0(X, L^{\otimes m}) \geq cm^n \quad \text{pour tout } m \gg 0.$$

On peut terminer la discussion du § 3.2 : un fibré en droites L est

- grand si et seulement s'il admet une métrique singulière à courant de courbure T dit *de Kähler*⁽¹²⁾, c'est-à-dire tel que $T - \omega$ soit positif pour une forme définie positive convenable ω .

THÉORÈME 3.5. — *Soit X une variété complexe compacte.*

- *Pour que X soit une variété de Fujiki (c'est-à-dire biméromorphe à une variété kählérienne), il faut et il suffit que le cône $\text{Big}_{\text{tr}}(X)$ ne soit pas vide.*
- *Pour que X soit une variété de Moishezon (c'est-à-dire biméromorphe à une variété projective), il faut et il suffit que le cône $\text{Big}(X)$ ne soit pas vide.*

Le premier point est un autre résultat de [DP], pour la démonstration duquel on utilise la cohomologie $\partial\bar{\partial}$ (cf. note 4). Le second est facile : les multiples suffisamment grands d'un fibré en droites grand définissent une application birationnelle entre X et une sous-variété (projective) d'un espace projectif. Rappelons qu'une variété kählérienne compacte de Moishezon est projective.

4. CÔNES DUAUX

Supposons X kählérienne. La dualité de Poincaré permet d'identifier $H^{n-1,n-1}(X, \mathbf{R})$ (que l'on notera parfois $H_{1,1}(X, \mathbf{R})$) au dual de $H^{1,1}(X, \mathbf{R})$, et le sous-espace vectoriel $N_1(X, \mathbf{R})$ engendré par

$$N_1(X) = H^{n-1,n-1}(X, \mathbf{R}) \cap (\text{Im}(H^{2n-2}(X, \mathbf{Z}) \rightarrow H^{2n-2}(X, \mathbf{R})))$$

au dual de $N^1(X, \mathbf{R})$.

Étant donné un cône \mathcal{C} dans un espace vectoriel réel H de dimension finie, on définit son dual par

$$\mathcal{C}^{\vee} = \{\xi^{\vee} \in H^{\vee} \mid \langle \xi^{\vee}, \xi \rangle \geq 0 \text{ pour tout } \xi \in \mathcal{C}\}$$

C'est un cône convexe fermé dans H^{\vee} . Si \mathcal{C} n'est pas d'intérieur vide, on a $\mathcal{C}^{\vee} = \overline{(\mathcal{C})}^{\vee} = (\mathcal{C}^{\circ})^{\vee}$. Si \mathcal{C} est convexe, on a $(\mathcal{C}^{\vee})^{\vee} = \overline{\mathcal{C}}$.

⁽¹²⁾Là encore, la terminologie est mauvaise : la classe d'un courant de Kähler est grande, pas de Kähler !

4.1. Dual du cône ample et du cône de Kähler

Lorsque X est projective, le cône $\text{Amp}(X)^\vee$ est le cône fermé engendré par les classes des courbes algébriques dans X (critère de Kleiman). L'analogie « transcendant » est le corollaire 2.3 : si X est kählérienne, le cône $\text{Kah}(X)^\vee \subset H_{1,1}(X, \mathbf{R})$ est le cône convexe fermé engendré par les

$$[Y] \wedge \omega^{\dim(Y)-1}$$

où ω décrit l'ensemble des formes de Kähler sur X , et Y celui des sous-variétés analytiques de X .

4.2. Dual du cône pseudo-effectif

Supposons X projective. Une courbe (irréductible réduite) C dans X est dite *mobile* s'il existe une modification $\mu : X' \rightarrow X$ et des diviseurs très amples H'_1, \dots, H'_{n-1} sur X' tels que

$$C = \mu_*(H'_1 \cap \dots \cap H'_{n-1})$$

Les déformations de C recouvrent X . De plus, étant donnée une hypersurface Z de X , il existe une déformation de C qui n'est pas contenue dans Z , ce qui montre qu'une classe mobile est positive sur tout élément de $\text{Pef}(X)$.

THÉORÈME 4.1 ([BDPP]). — *Soit X une variété projective. Le cône $\text{Pef}(X)^\vee \subset N_1(X, \mathbf{R})$ est le cône convexe fermé engendré par les classes de courbes mobiles.*

Un diviseur qui est de degré positif sur toute courbe mobile est donc pseudo-effectif. La preuve sera donnée dans le § 6.

Lorsque X est une variété kählérienne compacte, on conjecture que le cône $\text{Pef}_{\text{tr}}(X)^\vee$ est le cône convexe fermé engendré par les classes des courants du type

$$\mu_*(\omega'_1 \wedge \dots \wedge \omega'_{n-1})$$

où $\mu : X' \rightarrow X$ est une composition d'éclatements à centre lisse et les ω'_j sont des formes de Kähler sur X' . Cette conjecture a été démontrée dans des cas particuliers ([BDPP], Corollary 10.13), entre autres lorsque X est une surface ou une variété hyperkählienne.

Exemple 4.2 (Surfaces). — Lorsque X est une surface, on a $H_{1,1}(X, \mathbf{R}) = H^{1,1}(X, \mathbf{R})$. Les classes mobiles sont nef, et elles engendrent le cône $\text{Nef}(X)$. Avec les notations de l'exemple 3.2, on a ([Bo3], Theorem 4.1)

$$\text{Nef}_{\text{tr}}(X) = \text{Pef}_{\text{tr}}(X)^\vee \subset \overline{\mathcal{C}(X)} = \overline{\mathcal{C}(X)}^\vee \subset \text{Pef}_{\text{tr}}(X) = \text{Nef}_{\text{tr}}(X)^\vee$$

COROLLAIRE 4.3. — *Pour qu'une variété projective X soit recouverte par des courbes rationnelles, il faut et il suffit que sa classe canonique K_X ne soit pas pseudo-effective.*

PREUVE — Si K_X n'est pas pseudo-effectif, il existe par le théorème 4.1 une courbe dont les déformations recouvrent X et sur laquelle K_X est de degré strictement négatif. Étant donnée une courbe C telle que $K_X \cdot C < 0$, un résultat de Miyaoka et Mori ([MM]), basé sur le célèbre « lemme de cassage » de ce dernier, dit qu'il existe une courbe *rationnelle* dans X qui passe par un point donné arbitraire de C . Dans notre cas, cela entraîne que X est recouverte par des courbes rationnelles.

Inversement, si X est recouverte par des courbes rationnelles, le fibré normal à une de ces courbes C générale est engendré par ses sections, de sorte que $K_X \cdot C$ est strictement négatif. Soit H un \mathbf{Q} -diviseur ample vérifiant $(K_X + H) \cdot C < 0$. Tout diviseur effectif de classe $m(K_X + H)$, avec $m > 0$, doit donc contenir C . Il est donc nul ; cela montre que $K_X + H$ n'est pas grand, donc que K_X n'est pas pseudo-effectif. \square

Plus généralement, on conjecture (ce serait une conséquence du programme du modèle minimal de Mori) que X est recouverte par des courbes rationnelles si et seulement si sa dimension de Kodaira est $-\infty$, c'est-à-dire si $H^0(X, mK_X) = 0$ pour tout $m > 0$. L'implication directe est facile. Pour la réciproque, il reste donc à montrer que si K_X est pseudo-effectif, un de ses multiples a une section non nulle. Noter que c'est très particulier au diviseur canonique : il existe des classes pseudo-effectives dont aucun multiple (entier) n'est effectif.

Campana et Păun, dans un travail récent, déduisent du corollaire précédent que toute variété projective X admettant une application dominante $\mathbf{C}^n \rightarrow X$ de degré fini est recouverte par des courbes rationnelles (on conjecture l'unirationalité de X). Cela s'applique en particulier aux compactifications de \mathbf{C}^n , c'est-à-dire aux variétés complexes compactes dont un ouvert de Zariski est analytiquement isomorphe à \mathbf{C}^n . Lorsque $n = 2$, ce résultat est dû à Kodaira ([Ko], Theorem 5).

COROLLAIRE 4.4 (Campana–Păun). — *Toute compactification projective de \mathbf{C}^n est recouverte par des courbes rationnelles.*

PREUVE — Soient X une telle compactification et $\varphi : \mathbf{C}^n \rightarrow X$ une fonction holomorphe qui induit un isomorphisme analytique entre \mathbf{C}^n et un ouvert de Zariski U de X . Une métrique de Kähler ω sur X induit une métrique hermitienne h sur ω_X .

Supposons ω_X pseudo-effectif ; il existe alors une fonction $f \in L^1(X)$ telle que

$$(4) \quad \Theta_h(\omega_X) + \frac{i}{\pi} \partial \bar{\partial} f \geq 0$$

comme courant, et f est semi-continue supérieurement, donc majorée sur X . Le jacobien Jac_φ induit une section holomorphe du fibré en droites $\varphi^* \omega_X^{-1}$ sur \mathbf{C}^n . On a donc

$$\varphi^* \Theta_h(\omega_X) = \frac{i}{\pi} \partial \bar{\partial} \log \|\text{Jac}_\varphi\|^2$$

En prenant l'image inverse de (4) par φ , on obtient

$$i \partial \bar{\partial} \log \left(\|\text{Jac}_\varphi\|^2 e^{f \circ \varphi} \right) \geq 0$$

ce qui signifie que la fonction $\tau = \|\text{Jac}_\varphi\|^2 e^{f \circ \varphi}$ est pluri-sous-harmonique sur \mathbf{C}^n . Le théorème de changement de variables donne par ailleurs

$$\int_{\mathbf{C}^n} \tau d\lambda \leq e^{\sup_X f} \int_{\mathbf{C}^n} \|\text{Jac}_\varphi\|^2 d\lambda = e^{\sup_X f} \int_{\mathbf{C}^n} \varphi^* \omega^n = e^{\sup_X f} \int_X \omega^n$$

où λ est la mesure de Lebesgue sur \mathbf{C}^n . En conclusion, la fonction positive τ est intégrable sur \mathbf{C}^n par rapport à λ . Des arguments standard de convexité montrent que τ est identiquement nulle, ce qui est absurde (la fonction f ne peut valoir $-\infty$ sur un ouvert non vide de X).

Il s'ensuit que ω_X n'est pas pseudo-effectif, donc que X est recouverte par des courbes rationnelles (cor. 4.3). □

Je voudrais aussi mentionner une conjecture de Green et Griffiths ([De4], Conjecture 3.6), qui énonce qu'une variété est de type général si et seulement si elle n'est pas « mesure hyperbolique », c'est-à-dire s'il n'existe pas de suite d'applications holomorphes $f_r : \Delta(1)^{n-1} \times \Delta(r) \rightarrow X$ telle que $\|\text{Jac}_{f_r}(0)\| \geq 1$ pour tout $r \in \mathbf{N}$ (on a noté $\Delta(r)$ le disque complexe de centre 0 et de rayon r).

Le sens direct est dû à Kobayashi–Ochiai et Griffiths, indépendamment ([De4], Corollary 3.5). Pour la réciproque, on notera qu'une variété recouverte par des courbes rationnelles n'est pas mesure hyperbolique. Il reste donc à exclure le cas où K_X est sur le bord du cône pseudo-effectif.

5. VOLUME ET DÉCOMPOSITIONS DE ZARISKI

La démonstration du théorème 4.1 utilise la notion de « volume » d'un fibré en droites sur une variété kählérienne compacte X .

5.1. Volume d'un fibré en droites

Soit L un fibré en droites sur une variété kählérienne compacte X . On montre que la suite

$$\frac{h^0(X, L^{\otimes m})}{m^n/n!}$$

converge lorsque m tend vers $+\infty$. On note $\text{vol}(L)$ sa limite (le « volume » de L). Celle-ci ne dépend que de la première classe de Chern de L .

Comme de plus $\text{vol}(L^{\otimes m}) = m \text{vol}(L)$, on peut étendre par homogénéité le volume à une fonction sur $N^1(X, \mathbf{Q})$, qui se trouve être localement lipschitzienne ([Lz1], Theorem 2.2.44). On l'étend par continuité en une fonction continue $\text{vol} : N^1(X, \mathbf{R}) \rightarrow \mathbf{R}$. Elle vérifie

$$\xi \in \text{Big}(X) \iff \text{vol}(\xi) > 0$$

Si L est un fibré en droites nef sur X , l'estimation (1) entraîne $\text{vol}(L) = c_1(L)^n$. On a donc, par continuité de la fonction volume,

$$(5) \quad \text{vol}(\xi) = \xi^n \quad \text{pour tout } \xi \in \text{Nef}(X).$$

On aura aussi besoin des propriétés suivantes du volume : pour tout ξ dans $N^1(X, \mathbf{R})$, pour toute modification $\mu : X' \rightarrow X$ et pour tout $e \in \text{Pef}(X)$, on a ([Lz1], Exemples 2.2.48 et 2.2.49)

$$(6) \quad \text{vol}(\xi + e) \geq \text{vol}(\xi) \quad , \quad \text{vol}(\mu^* \xi) = \text{vol}(\xi)$$

Enfin, la fonction volume satisfait l'inégalité

$$\text{vol}(\xi_1 + \cdots + \xi_n)^{\frac{1}{n}} \geq \text{vol}(\xi_1)^{\frac{1}{n}} + \cdots + \text{vol}(\xi_n)^{\frac{1}{n}}$$

pour toutes classes ξ_1, \dots, ξ_n dans $\text{Pef}(X, \mathbf{R})$ ⁽¹³⁾.

Exemple 5.1. — Soit X l'éclaté de \mathbf{P}^n en un point. Soient E le diviseur exceptionnel et H l'image inverse dans X d'un hyperplan. On a $H^{1,1}(X, \mathbf{R}) = N^1(X, \mathbf{R})$ et cet espace vectoriel est engendré par les classes $[H]$ et $[E]$. Nos cônes sont donc plans. On a

$$\begin{aligned} \text{Amp}(X) &= \{x[H] - y[E] \mid x > y > 0\} \\ \text{Eff}(X) &= \{x[H] - y[E] \mid x \geq 0, x \geq y\} \end{aligned}$$

On a d'autre part

$$H^n = 1 \quad , \quad (-E)^n = -1 \quad , \quad H \cdot E = 0$$

Le volume est donné sur $\text{Nef}(X)$ par

$$\text{vol}(x[H] - y[E]) = (x[H] - y[E])^n = x^n - y^n$$

Si p et q sont des entiers positifs, qE est fixe dans le système linéaire $|pH + qE|$. On a donc $\text{vol}(pH + qE) = \text{vol}(pH) = p^n$. On en déduit

$$\text{vol}(x[H] - y[E]) = \begin{cases} x^n - y^n & \text{si } x \geq y \geq 0 ; \\ x^n & \text{si } x \geq 0, y \leq 0 ; \\ 0 & \text{sinon.} \end{cases}$$

Plus généralement, si X est une variété torique, $\text{Pef}(X)$ est un cône polyédral qui admet une partition en un nombre fini de sous-cônes polyédraux sur chacun desquels le volume est une fonction polynomiale ([ELMNP]).

Nombre d'auto-intersection mobile. On peut donner du volume d'un diviseur grand D l'interprétation géométrique suivante. Soit $\text{Base}(D)$ le lieu base de D (c'est-à-dire l'intersection de tous les éléments du système linéaire $|D|$). On pose

$$D^{[n]} = \text{Card}(D_1 \cap \cdots \cap D_n \cap (X - \text{Base}(D)))$$

⁽¹³⁾Cette inégalité se déduit de l'inégalité dite de Teissier–Hovanskii,

$$(7) \quad (\xi_1 \cdots \xi_n)^n \geq (\xi_1^n)^{\frac{1}{n}} \cdots (\xi_n^n)^{\frac{1}{n}}$$

valable pour toutes classes nef ξ_1, \dots, ξ_n dans $N^1(X, \mathbf{R})$, qui s'obtient à partir du théorème de l'indice de Hodge.

où D_1, \dots, D_n sont des éléments généraux de $|D|$. Si $|D|$ est sans point base, cela vaut donc D^n . On a alors ([Lz2], Définition 11.4.10; comparer avec (5))

$$\mathrm{vol}(D) = \lim_{m \rightarrow +\infty} \frac{(mD)^{[n]}}{m^n}$$

5.2. Décompositions de Zariski

On dit qu'un diviseur D sur X a une *décomposition de Zariski* s'il existe des \mathbf{Q} -diviseurs P et N , respectivement nef et effectif, tels que $D = P + N$ et que l'inclusion

$$H^0(X, [mP]) \hookrightarrow H^0(X, mD)$$

soit bijective pour tout entier $m > 0$.

Zariski établit dans [Z] l'existence d'une telle décomposition lorsque X est une *surface algébrique* et que D est effectif⁽¹⁴⁾. C'est d'ailleurs sans doute la première apparition des \mathbf{Q} -diviseurs (la décomposition n'est pas nécessairement entière, comme le montre l'exemple ci-dessous). Fujita étend dans [F] ce résultat aux diviseurs pseudo-effectifs.

Si on a une telle décomposition, $\mathrm{vol}(D) = P^n$. Sur une surface projective, le volume prend donc des valeurs rationnelles sur les points rationnels; de plus, c'est localement une fonction polynomiale quadratique ([BKS]).

Exemple 5.2 ([Lz1], Exemple 2.3.20). — Soit X la surface obtenue en éclatant trois points colinéaires sur \mathbf{P}^2 . Soient E_1, E_2 et E_3 les diviseurs exceptionnels et H l'image inverse dans X d'une droite. Le diviseur

$$D = 3H - 2E_1 - 2E_2 - 2E_3$$

est grand. Sa décomposition de Zariski est donnée par

$$P = \frac{1}{2}(3H - E_1 - E_2 - E_3) \quad , \quad N = \frac{3}{2}(H - E_1 - E_2 - E_3)$$

et $\mathrm{vol}(D) = P^2 = 3/2$.

Cutkosky a construit un exemple de fibré en droites sur une variété projective de dimension 3 dont le volume est irrationnel ([C]; [CS]; [Lz1], Exemple 2.3.8). Il existe aussi une variété projective de dimension 3 pour laquelle la fonction volume n'est pas localement polynomiale ([BKS]). Le résultat de Zariski ne s'étend donc pas tel quel aux variétés de dimension supérieure. On pourrait simplement demander qu'il existe une modification $\mu : X' \rightarrow X$ et une décomposition

$$\mu^*D = P + N$$

en somme de \mathbf{R} -diviseurs avec P nef, N effectif (sur X'), telles que l'inclusion

$$H^0(X', [mP]) \hookrightarrow H^0(X', m\mu^*D) = H^0(X, mD)$$

⁽¹⁴⁾Zariski caractérise en fait P —la partie positive—et N —la partie négative—de la façon suivante à l'aide du produit d'intersection entre diviseurs :

- P est orthogonal à chaque composante de N ;
- la matrice d'intersection des composantes de N est définie négative.

soit bijective pour tout entier $m > 0$. Hélas, même une telle décomposition n'existe pas toujours ([Ny1])⁽¹⁵⁾. On doit se contenter du résultat suivant ([F]).

THÉORÈME 5.3 (Décomposition de Zariski approchée). — *Soient X une variété projective et $\xi \in \text{Big}(X)$. Étant donné $\varepsilon > 0$, il existe une modification $\mu : X' \rightarrow X$ avec*

$$\mu^*\xi = a + e$$

où a est ample et e effectif (sur X'), avec

$$\text{vol}(\xi) \geq \text{vol}(a) \geq \text{vol}(\xi) - \varepsilon$$

Je renvoie à [Lz2], § 11.4 pour la preuve. Il faut noter que si ce résultat semble proche de l'existence d'une « vraie » décomposition de Zariski, il apporte beaucoup moins d'informations. En particulier, contrairement à celle-ci, il n'est d'aucune utilité pour étudier, pour un diviseur D sur X , l'anneau $\bigoplus_{m \geq 0} H^0(X, mD)$, le but originel de Zariski.

Dans la décomposition de Zariski sur une surface, les deux morceaux sont orthogonaux (cf. note 14). Dans une décomposition de Zariski approchée, ils restent « presque orthogonaux » (en un sens convenable). C'est l'objet du théorème suivant ([BDPP]), qui avait été essentiellement conjecturé par Nakamaye.

THÉORÈME 5.4. — *Soit X une variété projective. Soit $\xi \in \text{Big}(X)$ et soit $\mu : X' \rightarrow X$ une modification telle que $\mu^*\xi = a + e$, avec a ample et e effectif. Soit enfin h une classe ample sur X telle que $h \pm \xi$ est ample. On a*

$$(a^{n-1} \cdot e)^2 \leq 20h^n(\text{vol}(\xi) - \text{vol}(a))$$

PREUVE — Nous suivons [BDPP], § 5 et [Lz2], § 11.4. Pour $0 \leq t \leq 1$, on a $\text{vol}(\xi) = \text{vol}(a + e) \geq \text{vol}(a + te)$. Nous allons étudier la fonction $t \mapsto \text{vol}(a + te)$ sur l'intervalle $[0, 1]$. Écrivons $a + te$ comme différence de deux classes amples

$$a + te = p_t - q_t, \quad \text{avec } p_t = a + t\mu^*(\xi + h) \quad \text{et} \quad q_t = t(a + \mu^*h).$$

LEMME 5.5. — *Soit X une variété projective de dimension n . Pour tous ξ et ξ' dans $\text{Nef}(X)$, on a*

$$\text{vol}(\xi - \xi') \geq \xi^n - n\xi^{n-1}\xi'$$

PREUVE RAPIDE DU LEMME — Par continuité et homogénéité, on peut supposer que ξ et ξ' sont les classes de diviseurs D et E très amples. On choisit alors E_1, \dots, E_m distincts dans le système linéaire $|E|$. La suite exacte

$$0 \rightarrow H^0(X, m(D - E)) \rightarrow H^0(X, mD) \rightarrow \bigoplus_{j=1}^m H^0(E_j, D|_{E_j})$$

⁽¹⁵⁾En revanche, il est conjecturé qu'un diviseur canonique possède toujours une décomposition de Zariski. C'est équivalent au fait que l'anneau canonique est de type fini ([K]).

permet, avec l'estimation (1), de conclure⁽¹⁶⁾. □

On en déduit

$$\text{vol}(a + te) \geq p_t^n - np_t^{n-1}q_t$$

On a d'une part

$$p_t^n = (a + t(a + e + \mu^*h))^n \geq a^n + nta^{n-1}(a + e + \mu^*h)$$

et d'autre part

$$\begin{aligned} p_t^{n-1}q_t &= t(a + t(a + e + \mu^*h))^{n-1}(a + \mu^*h) \\ &= t \sum_{k=0}^{n-1} t^k a^{n-1-k} (a + e + \mu^*h)^k (a + \mu^*h) \binom{n-1}{k} \end{aligned}$$

On vérifie que si $a_1, \dots, a_n, a'_1, \dots, a'_n$ sont nefes et $a'_1 - a_1, \dots, a'_n - a_n$ effectifs, on a

$$(8) \quad a_1 \cdots a_n \leq (a_1 + a'_1) \cdots (a_n + a'_n)$$

Comme $\mu^*h - a$, $2\mu^*h - (a + e + \mu^*h)$ et $2\mu^*h - (a + \mu^*h)$ sont nefes, on en déduit

$$\begin{aligned} p_t^{n-1} \cdot q_t &\leq ta^{n-1}(a + \mu^*h) + \sum_{k=1}^{n-1} (2t)^{k+1} \mu^*h^n \binom{n-1}{k} \\ &\leq ta^{n-1}(a + \mu^*h) + 4(n-1)(1+2t)^{n-2}t^2h^n \end{aligned}$$

Pour $0 \leq t \leq \frac{1}{10n}$, on a $4(n-1)(1+2t)^{n-2} \leq 5n$, de sorte que

$$\text{vol}(a + te) \geq a^n + nta^{n-1}e - 5n^2t^2h^n$$

Le membre de droite de cette inégalité est maximal pour $t = t_0 = \frac{a^{n-1}e}{10nh^n}$, qui est bien $\leq \frac{1}{10n}$ par (8). On en déduit

$$\text{vol}(\xi) = \text{vol}(a + e) \geq \text{vol}(a + t_0e) \geq a^n + \frac{(a^{n-1}e)^2}{20h^n}$$

ce qui montre le théorème. □

⁽¹⁶⁾L'analogie de ce lemme pour des classes de Kähler quelconques n'est pas connu. On a quand même ([BDPP], Theorem 10.4)

$$\text{vol}(\xi - \xi') \geq \xi^n - \frac{(n+1)^2}{4} \xi^{n-1} \xi'$$

lorsque ξ est une classe de Kähler et ξ' une classe ample.

5.3. Volume d'une classe réelle de type $(1, 1)$

On peut partir de la décomposition de Zariski approchée pour *définir* le volume d'une classe réelle de type $(1, 1)$ sur une variété kählérienne compacte. Pour ξ dans $H^{1,1}(X, \mathbf{R})$, on pose $\text{vol}(\xi) = 0$ si $\xi \notin \text{Big}_{\text{tr}}(X)$, et sinon⁽¹⁷⁾

$$\text{vol}(\xi) = \sup_{\omega} \int_X \omega^n$$

où le supremum est pris sur tous les courants de Kähler T dans ξ à singularités analytiques⁽¹⁸⁾, toutes les modifications $\mu : X' \rightarrow X$ et toutes les décompositions $\mu^*T = \omega + [E]$ avec ω forme de Kähler sur X' et E \mathbf{Q} -diviseur effectif sur X' .

Lorsque ω est une forme de Kähler, on a $\text{vol}(\omega) = \omega^n$ (c'est donc le volume de X pour la métrique associée, d'où la terminologie). Cette définition du volume coïncide avec la définition précédente sur $\text{Pef}(X)$.

THÉORÈME 5.6 (Boucksom, [Bo2]). — *La fonction*

$$\text{vol} : H^{1,1}(X, \mathbf{R}) \longrightarrow \mathbf{R}$$

ainsi définie est continue, et

$$\xi \in \text{Big}_{\text{tr}}(X) \iff \text{vol}(\xi) > 0$$

La démonstration repose sur la même technique de dégénérescence d'équation de Monge-Ampère utilisée dans la preuve du théorème 2.1.

6. DÉMONSTRATION DU THÉORÈME 4.1

Nous suivons [BDPP], § 5 et [Lz2], § 11.4. On a déjà vu que le cône $\text{Pef}(X)$ est contenu dans le cône dual de l'ensemble des classes mobiles. Si l'inclusion est stricte, il existe une classe ξ dans la frontière de $\text{Pef}(X)$ et dans l'intérieur du cône dual. Soit h une classe ample telle que $h \pm 2\xi$ est ample. Il existe $\varepsilon > 0$ tel que $\xi - \varepsilon h$ soit encore dans le cône dual; on a donc

$$(9) \quad (\xi - \varepsilon h) \cdot \gamma \geq 0$$

⁽¹⁷⁾C'est la version de [BDPP], Définition 3.2. Boucksom montre que la classe grande ξ contient un courant T_{\min} positif « à singularités minimales » (non unique) et que $\text{vol}(\xi)$ est l'intégrale sur l'ouvert où les coefficients de T_{\min} sont bornés du courant T_{\min}^n (bien défini sur ce lieu).

⁽¹⁸⁾Les courants de Kähler sont définis dans le § 3.4. Un courant T réel de type $(1, 1)$ est dit à « singularités analytiques », ou à « pôles logarithmiques », s'il s'écrit localement comme la somme d'une forme différentielle de type $(1, 1)$ et d'un courant du type $ic\partial\bar{\partial}\log(|f_1|^2 + \cdots + |f_r|^2)$, où f_1, \dots, f_r sont des fonctions holomorphes et $c > 0$. Sur la modification de X obtenue en éclatant l'idéal (global) engendré (localement) par f_1, \dots, f_r , l'image inverse de T est somme d'une forme lisse et d'un diviseur.

pour toute courbe mobile γ . Prenons $\delta \in]0, \frac{1}{2}[$. Comme ξ est pseudo-effective, on a $\text{vol}(\xi + \delta h) \geq \text{vol}(\delta h) = \delta^n h^n > 0$. Il existe donc une décomposition de Zariski approchée (th. 5.3)

$$(10) \quad \mu_\delta^*(\xi + \delta h) = a_\delta + e_\delta \quad , \quad \text{vol}(a_\delta) \geq \max\{\text{vol}(\xi + \delta h) - \delta^{2n}, \frac{1}{2}\delta^n h^n\}$$

où $\mu_\delta : X'_\delta \rightarrow X$ est une modification. Considérons la classe mobile

$$\gamma_\delta = \mu_{\delta*}(a_\delta^{n-1})$$

En utilisant la formule de projection et l'inégalité (7), on obtient

$$h \cdot \gamma_\delta = \mu_\delta^* h \cdot a_\delta^{n-1} \geq (h^n)^{\frac{1}{n}} (a_\delta^n)^{\frac{n-1}{n}}$$

D'autre part,

$$\xi \cdot \gamma_\delta \leq (\xi + \delta h) \cdot \gamma_\delta = \mu_\delta^*(\xi + \delta h) \cdot a_\delta^{n-1} = (a_\delta + e_\delta) \cdot a_\delta^{n-1}$$

Comme $h \pm (\xi + \delta h)$ est ample, le théorème 5.4 entraîne, avec (10),

$$e_\delta \cdot a_\delta^{n-1} \leq (\text{vol}(\xi + \delta h) - \text{vol}(a_\delta)) \sqrt{20h^n} \leq \delta^n \sqrt{20h^n} \leq \delta (2a_\delta^n)^{\frac{n-1}{n}} \sqrt{20h^n}$$

On en déduit, en utilisant de nouveau (10),

$$(11) \quad \frac{\xi \cdot \gamma_\delta}{h \cdot \gamma_\delta} \leq \frac{a_\delta^n + \delta^n \sqrt{20h^n}}{(h^n)^{\frac{1}{n}} (a_\delta^n)^{\frac{n-1}{n}}} \leq \frac{a_\delta^n + a_\delta^n \frac{2}{h^n} \sqrt{20h^n}}{(h^n)^{\frac{1}{n}} (a_\delta^n)^{\frac{n-1}{n}}} \leq C (a_\delta^n)^{\frac{1}{n}}$$

où C est une constante strictement positive. On a $a_\delta^n = \text{vol}(a_\delta) \leq \text{vol}(\xi + \delta h)$ et $\text{vol}(\xi) = 0$ puisque ξ est sur la frontière du cône pseudo-effectif; on obtient donc par continuité de la fonction volume $\lim_{\delta \rightarrow 0^+} a_\delta^n = 0$, ce qui, avec (11), contredit (9), et termine la démonstration du théorème 6.

7. DÉCOMPOSITION DE ZARISKI DIVISORIELLE

Même si un diviseur n'a en général pas de décomposition de Zariski, on peut quand même (à la suite de Nakayama et Boucksom) obtenir une décomposition où la partie positive est « nef en codimension 1 ».

7.1. Lieu non nef pour les classes algébriques

Soit D un diviseur grand sur X . Nous voulons définir la notion de « nef en un point » x de X . Il se trouve que la bonne façon de faire n'est pas de demander que D soit de degré positif sur les courbes passant par x , mais de considérer les multiplicités des éléments des systèmes linéaires $|mD|$ en x . Tout d'abord, on pose

$$\text{mult}_x |D| = \min_{E \in |D|} \text{mult}_x E$$

Cet entier est nul si et seulement si x n'est pas dans le lieu base de $|D|$. Goodman montre ([G], Proposition 8) que D est nef si et seulement si

$$\liminf_{m \rightarrow +\infty} \frac{\text{mult}_x |mD|}{m} = 0$$

pour tout x dans X . La limite inférieure est en fait une limite; on la note $\text{mult}_x \|D\|$. Elle ne dépend que de la classe $[D]$, et $\text{mult}_x \|mD\| = m \text{mult}_x \|D\|$ pour tout entier positif m . On peut donc par homogénéité définir $\text{mult}_x \xi$ pour toute classe rationnelle grande ξ , et enfin définir une fonction

$$\text{mult}_x : \text{Big}(X) \longrightarrow \mathbf{R}^+$$

qui est continue ([Ny2], Lemma 1.7). Le *lieu non nef* de ξ est alors

$$\{x \in X \mid \text{mult}_x \xi > 0\}$$

Donnons une version plus géométrique de cette construction. On définit le *lieu base stable* d'un diviseur D comme l'intersection ensembliste

$$\mathbf{B}(D) = \bigcap_{m>0} \text{Base } |mD|$$

Pour tout entier positif m , on a $\mathbf{B}(D) = \mathbf{B}(mD)$. On peut donc définir le lieu base stable d'un \mathbf{Q} -diviseur, mais on n'a pas en général $\mathbf{B}(D) = \mathbf{B}(D')$ lorsque $[D] = [D']$ ([ELMNP], Example 1.1). Le résultat suivant rend les calculs pratiques possibles dans certains cas. C'est à Nakamaye qu'on doit l'idée de perturber les classes pour améliorer le comportement du lieu base stable ([Nm1], [Nm2]).

THÉORÈME 7.1 ([ELMNP]). — *Le lieu non nef d'un \mathbf{R} -diviseur D grand sur une variété projective X est égal à son lieu base restreint*

$$\mathbf{B}_-(D) = \bigcup_{\substack{H \text{ } \mathbf{R}\text{-diviseur ample} \\ D+H \text{ } \mathbf{Q}\text{-diviseur}}} \mathbf{B}(D+H) \subset \mathbf{B}(D)$$

Ce lieu est une union dénombrable de sous-variétés de X ; il contient la réunion des courbes sur lesquelles D est de degré strictement négatif, mais peut être plus grand ([BDPP], Remark 6.3). Il est défini pour tout \mathbf{R} -diviseur D (pas nécessairement grand) et ne dépend que de $[D]$ ([Lz1], Lemma 10.3.1). On a ([ELMNP], Example 1.21)

- $\mathbf{B}_-(D) = \emptyset$ si et seulement si D est nef;
- $\mathbf{B}_-(D) \neq X$ si et seulement si D est pseudo-effectif.

Pour tout \mathbf{R} -diviseur D , on pose aussi

$$\mathbf{B}_+(D) = \bigcap_{\substack{H \text{ } \mathbf{R}\text{-diviseur ample} \\ D+H \text{ } \mathbf{Q}\text{-diviseur}}} \mathbf{B}(D-H) \supset \mathbf{B}(D)$$

C'est une sous-variété de X qui ne dépend que de la classe $[D]$.

Exemple 7.2 (Surfaces projectives). — Si X est une surface projective, D a une décomposition de Zariski $D = P + N$ (§ 5.2). On a $\mathbf{B}_-(D) = \text{Supp}(N)$ et $\mathbf{B}_+(D) = \mathbf{B}_+(P)$ est la réunion de toutes les courbes irréductibles de X sur lesquelles P est de degré 0; il contient donc le support de N ([ELMNP], Exemples 1.11 et 4.6).

L'introduction de ce dernier lieu est justifiée par le résultat suivant de Nakamaye ([Nm1]; [Lz1], Theorem 10.3.5).

THÉORÈME 7.3. — *Si D est un diviseur nef et grand, $\mathbf{B}_+(D)$ est la réunion de toutes les sous-variétés irréductibles Y de X pour lesquelles*

$$D^{\dim(Y)} \cdot Y = 0$$

7.2. Lieu non nef général

On peut étendre les définitions précédentes au cas où X est une variété kählérienne compacte (et même une variété complexe compacte quelconque; [Bo3], Definition 3.3). La notion de multiplicité sera remplacée par celle de *nombre de Lelong*.

Le nombre de Lelong en un point x de X d'un courant fermé réel positif T de type $(1, 1)$, qui s'écrit localement $T = \frac{i}{\pi} \partial \bar{\partial} \varphi$, est défini par

$$\nu(T, x) = \liminf_{z \rightarrow x} \frac{\varphi(z)}{\log |z - x|} \geq 0$$

Si D est un diviseur effectif, un résultat de Thie ([T]) entraîne $\nu(T_D, x) = \text{mult}_x D$. Enfin, pour tout $\xi \in \text{Big}_{\text{tr}}(X)$, on pose⁽¹⁹⁾

$$\nu(\xi, x) = \inf_{\substack{T \in \xi \\ T \text{ positif}}} \nu(T, x)$$

On appelle ce réel positif la *multiplicité minimale* de ξ en x . Si $\xi \in \text{Big}(X)$, cette définition coïncide avec celle de multiplicité [Bo3], Theorem 5.4) :

$$\nu(\xi, x) = \text{mult}_x \xi$$

On peut donc définir le *lieu non nef* d'un élément quelconque ξ de $H^{1,1}(X, \mathbf{R})$ comme⁽²⁰⁾

$$\mathbf{B}_-(\xi) = \begin{cases} \{x \in X \mid \nu(\xi, x) > 0\} & \text{si } \xi \in \text{Pef}(X); \\ X & \text{sinon.} \end{cases}$$

L'analogue du théorème 7.1 est l'égalité suivante ([BDPP], § 6) :

$$\mathbf{B}_-(\xi) = \bigcup_{\varepsilon > 0} \bigcap \text{Supp}(E)$$

⁽¹⁹⁾Avec les notations de la note 17, on a $\nu(\xi, x) = \nu(T_{\min}, x)$. D'autre part, on peut comme dans [Bo3], Proposition 3.2, étendre cette définition au cas $\xi \in \text{Pef}_{\text{tr}}(X)$; la fonction $\nu(\cdot, x)$ ainsi définie n'est pas continue.

⁽²⁰⁾Dans [Bo3], Definition 3.3, il est noté $E_{nn}(\xi)$.

où, une forme de Kähler ω sur X étant fixée, l'intersection est prise sur tous les courants T de classe $\xi + \varepsilon\omega$ et toutes les modifications $\mu : X' \rightarrow X$ pour lesquelles $\mu^*T = \beta + [E]$, où E est un diviseur effectif dans X' et β une forme positive fermée.

Boucksom définit dans [Bo3], Definition 3.18, pour toute classe transcendante ξ , le lieu $\mathbf{B}_+(\xi)$ (appelé « lieu non kählérien » et noté $E_{nK}(\xi)$ dans *loc. cit.*).

7.3. Encore des cônes

On peut séparer les éléments de $H^{1,1}(X, \mathbf{R})$ selon la dimension de leur lieu non nef. Cela définit ainsi, pour chaque $k \in \{0, \dots, n-1\}$, un cône ouvert convexe

$$\mathbf{C}_k(X) = \{\xi \in H^{1,1}(X, \mathbf{R}) \mid \dim \mathbf{B}_+(\xi) \leq k\}$$

et son adhérence

$$\overline{\mathbf{C}_k(X)} = \overline{\{\xi \in H^{1,1}(X, \mathbf{R}) \mid \dim \mathbf{B}_-(\xi) \leq k\}}$$

On a⁽²¹⁾

$$\text{Nef}_{\text{tr}}(X) = \overline{\mathbf{C}_0(X)} \subset \overline{\mathbf{C}_1(X)} \subset \dots \subset \overline{\mathbf{C}_{n-1}(X)} = \text{Pef}_{\text{tr}}(X)$$

On peut dire des éléments de $\mathbf{C}_k(X)$ qu'ils sont « nef en codimension $n - k - 1$ »⁽²²⁾.

Le cône $\overline{\mathbf{C}_{n-2}(X)} \cap N^1(X, \mathbf{R})$ a tout d'abord été considéré par Kawamata dans [K], sous le nom de « movable cone », puis étudié dans [Ny2]. Boucksom étend la définition aux classes transcendantes dans [Bo3], et le nomme « modified nef cone ». Il montre qu'il est engendré par les images directes de classes de Kähler sur des modifications de X . C'est aussi le cône engendré par les classes grandes ξ telle que $\text{mult}_D(\xi) = 0$ pour tout diviseur premier D dans X .

Exemple 7.4 (Produits symétriques d'une courbe). — Soit C une courbe projective lisse de genre $g \geq 2$. Son n -ième produit symétrique $X = \text{Sym}^n C$ est une variété projective (lisse) de dimension n . Supposons $n > 1$. Lorsque C est très générale, l'espace vectoriel $N^1(X, \mathbf{R})$ est de dimension 2, engendré par la classe ample x du diviseur $\text{Sym}^{n-1} C$ et par la classe nef θ de l'image inverse par l'application d'Abel–Jacobi $\text{Sym}^n C \rightarrow J^n C$ de la polarisation canonique de la jacobienne $J^n C$. On pose

$$t_k = \inf\{t \in \mathbf{R} \mid -\theta + tx \in \mathbf{C}_k(X)\} > 0$$

La « grande diagonale » Δ_C a pour classe $2(-\theta + (g + n - 1)x)$ et est sur la frontière du cône pseudo-effectif de X ([Kou], Theorem 3), de sorte que

$$t_{n-1} = g + n - 1$$

⁽²¹⁾L'égalité de gauche provient du fait que $\mathbf{B}_+(\xi)$ n'a jamais de point isolé.

⁽²²⁾Il serait intéressant de déterminer les cônes duaux (cf. § 4) des cônes $\mathbf{C}_k(X)$ pour prolonger ce qui a été fait pour $k = 0$ et $k = n - 1$ dans les § 4.1 et 4.2. On pourrait penser par analogie que le dual de $\mathbf{C}_k(X)$ est le cône engendré par les classes de courbes dont les déformations recouvrent une sous-variété de X de dimension $\geq k + 1$, mais c'est déjà faux pour $k = n - 2$. Une caractérisation complète du cône dual de $\mathbf{C}_k(X)$ vient d'être donnée dans [Pa] lorsque X est une variété torique.

L'argument de Kouvidakis prouve en fait que la classe tout diviseur premier de X autre que Δ_C s'écrit $c(-\theta + tx)$ avec $c > 0$ et $t \geq g + n - 1 + \frac{2}{n}(g + 1)$. Cela entraîne $\mathbf{B}_+(-\theta + tx) \supset \Delta_C$ pour $t < g + n - 1 + \frac{2}{n}(g + 1)$, de sorte que

$$t_{n-2} \geq g + n - 1 + \frac{2}{n}(g + 1) > t_{n-1}$$

La demi-droite $\mathbf{R}^+(-\theta + t_0x)$ est orthogonale à la « petite diagonale » $\delta_C^{(23)}$. Comme $x \cdot \delta_C = n$ et $\theta \cdot \delta_C = n^2g$, on obtient $\mathbf{B}_-(-\theta + tx) \supset \delta_C$ pour $t < ng$ et

$$t_0 = ng.$$

Calculer les pentes

$$u_k = \inf\{t \in \mathbf{R} \mid \theta + tx \in \mathbf{C}_k(X)\} \leq 0$$

des autres faces de ces cônes est en général beaucoup plus difficile. Lorsque $n > g$, elles valent toutes 0 car θ est nef mais pas grand. Lorsque C a un g_d^1 avec $d \leq n$, la classe θ est nef mais pas ample, donc $u_0 = 0$. Lorsque $g = 2n$ et que C est générale, on a $u_0 = -2$ ([P], Theorem 1.1).

7.4. Décomposition de Zariski divisorielle

L'idée est de « projeter » une classe pseudo-effective ξ sur le cône $\overline{\mathbf{C}_{n-2}(X)}$ en lui soustrayant des multiples convenables des diviseurs contenus dans son lieu non nef $\mathbf{B}_-(\xi)$. Pour tout diviseur premier D de X , on note⁽²⁴⁾

$$\nu(\xi, D) = \inf_{x \in D} \nu(\xi, x)$$

On montre que ces nombres sont nuls sauf pour un nombre fini de diviseurs premiers D ([Ny2], Corollary III.1.11; [Bo3], § 3.3). On pose alors⁽²⁵⁾

$$N_\xi = \sum_{D \text{ diviseur premier}} \nu(\xi, D)D$$

On notera que N_ξ est un \mathbf{R} -diviseur, pas seulement une classe. Le théorème suivant rassemble les résultats de [Ny2], III, § 1, dans le cas algébrique et de [Bo3], § 3.2, 3.3 et 3.7, en général.

⁽²³⁾Cela résulte du fait que le morphisme

$$\begin{aligned} C^n &\longrightarrow (J^0 C)^{\binom{n}{2}} \\ (p_1, \dots, p_n) &\longmapsto (p_i - p_j)_{1 \leq i < j \leq n} \end{aligned}$$

contracte exactement la petite diagonale et que l'image inverse de la polarisation produit descend en une classe nef non ample sur X ([P], Proposition 2.2).

⁽²⁴⁾Lorsque ξ est dans $N^1(X, \mathbf{R})$, la notation de [Ny2] est $\sigma_D(\xi)$.

⁽²⁵⁾La notation de Nakayama est $N_\sigma(\xi)$.

THÉORÈME 7.5. — Soit ξ une classe pseudo-effective sur une variété kählérienne compacte X . L'application

$$\begin{aligned} \rho : \text{Pef}_{\text{tr}}(X) &\longrightarrow \overline{\mathbf{C}_{n-2}(X)} \\ \xi &\longmapsto \xi - [N_\xi] \end{aligned}$$

est concave, homogène de degré 1 et continue sur $\text{Big}_{\text{tr}}(X)$. Elle conserve le volume.

Si $\xi \in \overline{\mathbf{C}_{n-2}(X)}$, on a $\rho(\xi) = \xi$.

Si $\xi \notin \overline{\mathbf{C}_{n-2}(X)}$, on a $\rho(\xi) \in \partial\overline{\mathbf{C}_{n-2}(X)}$.

La décomposition $\xi = \rho(\xi) + [N_\xi]$ est dite *décomposition de Zariski divisorielle* de ξ . Le lien avec la décomposition de Zariski telle qu'elle est définie dans le § 5.2 est le suivant : si D est un diviseur grand sur X , la décomposition $D = P_D + N_{[D]}$, avec $P_D = D - N_{[D]}$, est l'unique décomposition de D en somme de \mathbf{R} -diviseurs respectivement nef en codimension un et effectif, telle que l'inclusion

$$H^0(X, [mP_D]) \hookrightarrow H^0(X, mD)$$

est bijective pour tout entier $m > 0$ ([Bo3], Theorem 5.5).

Si $\xi \in \overline{\mathbf{C}_{n-1}(X)} - \overline{\mathbf{C}_{n-2}(X)}$, les diviseurs premiers qui apparaissent avec des coefficients non nuls dans N_ξ sont très rigides au sens où leurs combinaisons linéaires sont déterminées par leur classe de cohomologie ; ils sont en particulier en nombre fini. Plus généralement, on dit qu'un \mathbf{R} -diviseur effectif E est *exceptionnel* s'il vérifie les conditions équivalentes suivantes ([Ny2], § 1.a ; [Bo3], Proposition 3.13)

- $N_{[E]} = E$;
- les classes des composantes irréductibles de E sont linéairement indépendantes dans $N^1(X, \mathbf{R})$ et le cône convexe qu'elles engendrent ne rencontre pas $\overline{\mathbf{C}_{n-2}(X)}$.

Il peut cependant y avoir une infinité (dénombrable) de diviseurs exceptionnels.

PROPOSITION 7.6 ([Bo3], Proposition 3.15). — Soit E un \mathbf{R} -diviseur effectif exceptionnel. La classe $[E]$ contient un unique courant positif, à savoir T_E . En particulier, si E est un diviseur, on a $h^0(X, mE) = 1$ pour tout $m \geq 0$.

Exemple 7.7 (Surfaces projectives). — Si X est une surface projective, la décomposition de Zariski divisorielle est la décomposition de Zariski. Un \mathbf{R} -diviseur effectif E est exceptionnel si et seulement si la matrice des nombres d'intersection de ses composantes irréductibles est définie négative ([Bo3], Theorem 4.5). C'est équivalent à dire qu'il existe une surface projective Y et un morphisme $X \rightarrow Y$ qui contracte chaque composante connexe de $\text{Supp}(E)$ sur un point et qui est un isomorphisme en dehors de ce lieu.

Exemple 7.8 (Variétés hyperkählériennes). — On a (cf. exemple 3.4)

$$\text{Nef}_{\text{tr}}(X) \subset \overline{\mathbf{C}_{n-2}(X)} \subset \overline{\mathcal{C}(X)} \subset \text{Pef}_{\text{tr}}(X)$$

et les cônes $\overline{C_{n-2}(X)}$ et $\text{Pef}_{\text{tr}}(X)$ sont duaux pour la forme quadratique de Beauville–Bogomolov ([H3]). Plus précisément, une classe ξ de $\overline{\mathcal{C}(X)}$ est dans $\overline{C_{n-2}(X)}$ si $q_X(\xi, [D]) \geq 0$ pour tout diviseur D de X recouvert par des courbes rationnelles.

Un \mathbf{R} -diviseur effectif de composantes irréductibles E_1, \dots, E_r est exceptionnel si et seulement si la matrice $(q_X(E_j, E_k))$ est définie négative, et tous les diviseurs exceptionnels sont recouverts par des courbes rationnelles ([Bo3], Theorem 4.5 et Proposition 4.7).

On a $q_X(\rho(\xi), [N_\xi]) = 0$ et $\text{vol}(\xi) = \rho(\xi)^n$ ([Bo3], Proposition 4.12 et Theorem 4.8). Si ξ est rationnelle, N_ξ est un \mathbf{Q} -diviseur et son volume est donc rationnel.

Exemple 7.9 (Produits symétriques d’une courbe). — Dans la situation de l’exemple 7.4, la grande diagonale est un diviseur exceptionnel.

RÉFÉRENCES

- [BKS] T. BAUER, A. KÜRONYA, T. SZEMBERG – *Zariski chambers, volumes, and stable base loci*, J. reine angew. Math. **576** (2004), 209–233.
- [B] A. BEAUVILLE – *Variétés Kähleriennes dont la première classe de Chern est nulle*, J. Differential Geom. **18** (1983), 755–782.
- [Bi1] P. BIRAN – *Symplectic packing in dimension 4*, Geom. Funct. Anal. **7** (1997), 420–437.
- [Bi2] P. BIRAN – *From Symplectic Packing to Algebraic Geometry and Back*, Proceedings of the 3rd European Congress of Mathematics (Barcelona 2000), Vol II, 507–524, Progr. Math. **202**, Birkhäuser 2001.
- [Bo1] S. BOUCKSOM – *Le cône kählérien d’une variété hyperkählérienne*, C. R. Acad. Sci. Paris **333** (2001), 935–938.
- [Bo2] S. BOUCKSOM – *On the volume of a line bundle*, International Journal of Mathematics **13** (2002), 1043–1063.
- [Bo3] S. BOUCKSOM – *Divisorial Zariski decompositions on compact complex manifolds*, Ann. Sci. École Norm. Sup. **37** (2004), 45–76.
- [Bo4] S. BOUCKSOM – *Cônes positifs des variétés complexes compactes*. Thèse, Grenoble, 2002.
- [BDPP] S. BOUCKSOM, J.-P. DEMAILLY, M. PĂUN, T. PETERNELL – *The pseudo-effective cone of a compact Kähler manifold and varieties of negative Kodaira dimension*, eprint math.AG/0405285.
- [Bu] N. BUCHDAHL – *On compact Kähler surfaces*, Ann. Inst. Fourier **49** (1999), 287–302.
- [CP] F. CAMPANA, T. PETERNELL – *Algebraicity of the ample cone of projective varieties*, J. reine angew. Math. **407** (1990), 160–166.

- [C] S. CUTKOSKY – *Zariski decomposition of divisors on algebraic varieties*, Duke Math. J. **53** (1986), 149–156.
- [CS] S. CUTKOSKY, V. SRINIVAS – *On a problem of Zariski on dimensions of linear systems*, Ann. of Math. **137** (1993), 531–559.
- [De1] J.-P. DEMAILLY – *Champs magnétiques et inégalités de Morse pour la d'' -cohomologie*, C. R. Acad. Sci. Paris **301** (1985), 119–122.
- [De2] J.-P. DEMAILLY – *Singular Hermitian metrics on positive line bundles*, Complex algebraic varieties (Bayreuth, 1990), 87–104, Lecture Notes in Math. **1507**, Springer, Berlin, 1992.
- [De3] J.-P. DEMAILLY – *On the geometry of positive cones of projective and Kähler varieties*, The Fano Conference, Proceedings, Torino 2002, 395–422, édité par Collino, A. Conte et M. Marchiso, Torino, 2004.
- [De4] J.-P. DEMAILLY – *Algebraic criteria for Kobayashi hyperbolic projective varieties and jet differentials*, Algebraic geometry—Santa Cruz 1995, 285–360, Proc. Sympos. Pure Math. **62**, Part 2, Amer. Math. Soc., Providence, RI, 1997.
- [DP] J.-P. DEMAILLY, M. PĂUN – *Numerical characterization of the Kähler cone of a compact Kähler manifold*, Ann. of Math. **159** (2004), 1247–1274.
- [DPS] J.-P. DEMAILLY, T. PETERNELL, M. SCHNEIDER – *Compact complex manifolds with numerically effective tangent bundles*, J. Algebraic Geom. **3** (1994), 295–345.
- [ELMNP] L. EIN, R. LAZARSFELD, M. MUSTAŢĂ, M. NAKAMAYE, M. POPA – *Asymptotic Invariants of Base Loci*, eprint math.AG/0308116.
- [F] T. FUJITA – *On Zariski problem*, Proc. Japan Acad. Ser. A Math. Sci. **55** (1979), 106–110.
- [G] J. GOODMAN – *Affine open subsets of algebraic varieties and ample divisors*, Ann. of Math. **89** (1969), 160–183.
- [H1] D. HUYBRECHTS – *Compact hyperkähler manifolds : basic results*, Invent. Math. **135** (1999), 63–113.
- [H2] D. HUYBRECHTS – *Erratum. Compact hyperkähler manifolds : basic results*, Invent. Math. **152** (2003), 209–212.
- [H3] D. HUYBRECHTS – *The Kähler cone of a compact hyperkähler manifold*, Math. Ann. **326** (2003), 499–513.
- [K] Y. KAWAMATA – *The Zariski decomposition of log-canonical divisors*, Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985), 425–433, Proc. Sympos. Pure Math. **46**, Part 1, Amer. Math. Soc., Providence, RI, 1987.
- [Ko] K. KODAIRA – *Holomorphic mappings of polydiscs into compact complex manifolds*, J. Differential Geometry **6** (1971), 33–46.

- [Kou] A. KOUVIDAKIS – *Divisors on symmetric products of curves*, Trans. Amer. Math. Soc. **337** (1993), 117–128.
- [Kov] S. KOVÁCS – *The cone of curves of a K3 surface*, Math. Ann. **300** (1994), 681–691.
- [L1] A. LAMARI – *Courants kählériens et surfaces compactes*, Ann. Inst. Fourier **49** (1999), 263–285.
- [L2] A. LAMARI – *Le cône kählérien d’une surface*, J. Math. Pures Appl. **78** (1999), 249–263.
- [Lz1] R. LAZARSFELD – *Positivity in algebraic geometry I*, Ergebnisse der Mathematik und ihrer Grenzgebiete **48**, Springer-Verlag, Heidelberg, 2004.
- [Lz2] R. LAZARSFELD – *Positivity in algebraic geometry II*, Ergebnisse der Mathematik und ihrer Grenzgebiete **49**, Springer-Verlag, Heidelberg, 2004.
- [McP] D. McDUFF, L. POLTEROVICH – *Symplectic packings and algebraic geometry*, Invent. Math. **115** (1994), 405–434.
- [MM] Y. MIYAOKA, S. MORI – *A Numerical Criterion for Uniruledness*, Ann. of Math. **124** (1986), 65–69.
- [N] M. NAGATA – *On the 14-th problem of Hilbert*, Amer. J. Math. **33** (1959), 766–772.
- [Nm1] M. NAKAMAYE – *Stable base loci of linear series*, Math. Ann. **318** (2000), 837–847.
- [Nm2] M. NAKAMAYE – *Base loci of linear series are numerically determined*, Trans. Amer. Math. Soc. **355** (2003), 551–566.
- [Ny1] N. NAKAYAMA – *A counterexample to the Zariski-decomposition conjecture*, (en japonais) Hodge theory and Algebraic Geometry, Proceedings, Hokkaido Univ. (1994), Infinite Analysis Lecture Notes **19**, Kyoto University, 1994, 96–101.
- [Ny2] N. NAKAYAMA – *Zariski-decomposition and Abundance*, preprint RIMS–1142, Kyoto Univ., 1997, remanié en 2004.
- [P] G. PACIENZA – *On the nef cone of symmetric products of a generic curve*, Amer. J. Math. **125** (2003), 1117–1135.
- [Pa] S. PAYNE – *Stable base loci, movable curves, and small modifications, for toric varieties*, eprint arXiv :math.AG/0506622.
- [T] P. THIE – *The Lelong number of a point of a complex analytic set*, Math. Ann. **172** (1967), 269–312.
- [Z] O. ZARISKI – *The theorem of Riemann-Roch for high multiples of an effective divisor on an algebraic surface*, Ann. of Math. **76** (1962), 560–615.

Olivier DEBARRE

Institut de Recherche Mathématique Avancée

Université Louis Pasteur

7, rue René Descartes

F-67084 Strasbourg Cedex

E-mail : debarre@math.u-strasbg.fr

**PROGRESSIONS ARITHMÉTIQUES
DANS LES NOMBRES PREMIERS**
[d'après B. Green et T. Tao]

par **Bernard HOST**

1. INTRODUCTION

1.1. Le résultat

Le but de cet exposé est de présenter un travail récent et spectaculaire de B. Green et T. Tao où ils montrent :

THÉORÈME 1.1 ([GT]). — *L'ensemble des nombres premiers contient des progressions arithmétiques de toutes longueurs.*

En fait Green et Tao montrent un résultat plus fort : la conclusion du théorème reste valable si on remplace l'ensemble des nombres premiers par un sous-ensemble de densité relative positive. De plus, la méthode employée permet de déterminer explicitement pour tout k un entier N (très grand) tel que l'ensemble des nombres premiers plus petits que N contienne une progression arithmétique de longueur $k + 1$.

Le théorème 1.1 répond à une question fort ancienne bien que difficile à dater exactement. Très peu de résultats partiels étaient connus jusqu'ici ; citons celui de van der Corput [vdC] qui a montré en 1939 l'existence d'une infinité de progressions de longueur 3 dans les nombres premiers.

En 1923, Hardy et Littlewood [HL] ont proposé une conjecture très générale sur la répartition de certaines configurations dans les nombres premiers, qui entraînerait une version quantitative précise du théorème 1.1 si elle s'avérait exacte. Ce même théorème suivrait aussi d'une résolution positive donnée à une conjecture proposée par Erdős et Turán [ET] en 1936 :

CONJECTURE. — *Tout sous-ensemble E de \mathbb{N}^* vérifiant $\sum_{n \in E} \frac{1}{n} = +\infty$ contient des progressions arithmétiques de toutes longueurs.*

Cette conjecture reste totalement ouverte et les méthodes de Green et Tao ne permettent pas de s'en approcher. Dans une direction voisine, Szemerédi a montré en 1975

l'existence de progressions sous l'hypothèse plus forte de la densité positive. Rappelons que la densité d'un ensemble d'entiers $E \subset \mathbb{N}$ est :

$$d^*(E) = \limsup_{N \rightarrow \infty} \frac{1}{N} \text{Card}(E \cap [0, N - 1]) .$$

Le théorème de Szemerédi s'énonce :

THÉORÈME DE SZEMERÉDI ([S]). — *Tout ensemble d'entiers de densité positive contient des progressions arithmétiques de toutes longueurs.*

Il peut aussi s'exprimer en termes d'ensembles finis d'entiers :

VERSION FINIE DU THÉORÈME DE SZEMERÉDI. — *Pour tout entier $k \geq 2$ et tout réel $\delta > 0$ il existe un entier $N = N(k, \delta)$ tel que tout sous-ensemble E de $[0, N[$ ayant au moins δN éléments contienne une progression arithmétique de longueur $k + 1$.*

Ce théorème ne peut évidemment pas être utilisé directement puisque les nombres premiers ont une densité nulle. Cependant il tient une place centrale dans la démonstration.

1.2. La méthode

Le travail de Green et Tao comporte deux parties très différentes.

La première partie, qui est la plus longue, contient la démonstration d'une extension de la version finie du théorème de Szemerédi (théorème 2.2).

Dans ce dernier théorème, la quantité $|E|/N \geq \delta$ peut être vue comme la moyenne sur $[0, N[$ de la fonction indicatrice de E . L'idée naturelle est de remplacer cette fonction indicatrice par une fonction nulle en dehors de l'ensemble des nombres premiers, mais alors cette fonction ne peut pas être choisie majorée par 1 sinon sa moyenne deviendrait arbitrairement petite pour N grand. Green et Tao montrent un théorème de Szemerédi modifié (théorème 2.2) qui s'applique à une fonction majorée par un « poids pseudo-aléatoire », c'est-à-dire par une fonction de moyenne 1 dont les corrélations sont voisines de celles qu'on obtiendrait en tirant au hasard et indépendamment les valeurs aux points $0, 1, \dots, N - 1$ (section 2.2). Cette utilisation d'une majoration fait penser à la méthode du crible.

La démonstration de ce « théorème de Green-Tao Szemerédi » est écrite dans le langage des probabilités. Comme tous les espaces de probabilité sont finis et munis de la mesure uniforme, on pourrait dire qu'elle utilise seulement des arguments de dénombrement. Cette façon de voir serait formellement correcte mais trop réductrice. En fait la démarche de Green et Tao s'inspire directement de la théorie ergodique, et plus précisément de la démonstration ergodique du théorème de Szemerédi donnée par Furstenberg ([F], voir aussi [FKO]). Dans les deux cas, le cœur de la preuve est un résultat de décomposition (proposition 3.4) consistant à écrire une fonction comme la somme de son espérance conditionnelle sur une σ -algèbre bien choisie et d'un reste. L'espérance conditionnelle est « lissée » et dans le cas considéré par Green et Tao elle est même uniformément

bornée, ce qui permet d'utiliser le théorème de Szemerédi classique. Le reste se comporte comme une oscillation aléatoire et sa contribution dans les calculs est négligeable. Les ergodiciciens reconnaîtront la façon dont les « facteurs » interviennent dans de nombreux problèmes. Pour les autres, nous ajoutons que l'article n'utilise aucun résultat provenant de la théorie ergodique et que sa lecture ne demande aucune connaissance dans ce domaine.

Cette inspiration ergodique dans une démonstration combinatoire est encore plus apparente dans la nouvelle démonstration que T. Tao vient de donner du théorème de Szemerédi [T1]. Nous ne pensons pas que cette démarche soit artificielle. Jusqu'à présent les relations entre ces domaines se résumaient pratiquement au principe de correspondance de Furstenberg qui permet de montrer, à partir de théorèmes ergodiques, des résultats combinatoires dont beaucoup n'ont aujourd'hui pas d'autre preuve. Il apparaît depuis peu des ressemblances de plus en plus prononcées quoiqu'encore mal comprises entre les objets et les méthodes des deux théories. Nous reviendrons dans ces notes sur ce point qui mérite sans doute d'être approfondi.

Une fois démontré le théorème de Szemerédi modifié, il reste à construire un poids pseudo-aléatoire adapté au problème posé. Il s'agit donc ici de théorie des nombres. Dans cette partie de l'article [GT] les auteurs utilisent une fonction de von Mangoldt tronquée et font appel à des outils sophistiqués provenant des travaux de Goldston et Yıldırım [GY] mais, dans une note non publiée [T2], T. Tao explique comment l'argument peut être modifié pour n'utiliser que les propriétés les plus élémentaires des nombres premiers et de la fonction ζ . C'est cette approche que nous suivons ici en nous inspirant de notes manuscrites de J.-C. Yoccoz.

Dans cet exposé, qui ne contient aucune démonstration complète, on se propose de présenter de façon assez détaillée l'organisation de la preuve et de donner une idée des méthodes employées à chaque étape. Le lecteur pressé pourra se limiter à la section 2 qui contient la formulation précise des définitions et résultats correspondant aux deux grandes parties auxquelles on vient de faire allusion, encore que la définition des normes de Gowers (sous-sections 3.1 et 3.2) ait son intérêt propre. Le résultat de décomposition (proposition 3.4) est énoncé dans la sous-section 3.4 et montré dans la section 4. La deuxième partie de la preuve, c'est-à-dire la construction du poids pseudo-aléatoire, est contenue dans la section 5.

1.3. Conventions et notations

Quand f est une fonction définie sur un ensemble fini A , l'*espérance* de f sur A , notée $\mathbb{E}(f(x) \mid x \in A)$ ou $\mathbb{E}(f \mid A)$, est la moyenne arithmétique de f sur A ; la même est utilisée pour les fonctions de plusieurs variables.

Dans toute la suite, $k \geq 2$ est un entier que nous considérons comme une constante. L'objectif est de montrer l'existence d'une progression arithmétique de longueur $k + 1$ dans les nombres premiers. La progression est cherchée dans l'intervalle $[0, N[$, où N est

un (grand) entier qu'il est souvent nécessaire de supposer premier. On identifie $[0, N[$ au groupe $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$.

Il est crucial dans la preuve de contrôler la manière dont toutes les estimations dépendent de N et nous adoptons les conventions suivantes. Dans chaque énoncé, N est supposé fixé mais toutes les constantes sont indépendantes de N . Nous notons $o(1)$ une quantité tendant vers 0 quand N tend vers l'infini, uniformément par rapport à tous les paramètres sauf éventuellement ceux notés en indice. La notation $O(1)$ est employée avec une signification similaire.

2. POIDS PSEUDO-ALÉATOIRES ET THÉORÈME DE GREEN-TAO SZEMERÉDI

Green et Tao généralisent une formulation classique du théorème de Szemerédi, qui est celle sous laquelle Gowers [G] l'a redémontré récemment.

THÉORÈME 2.1. — *Pour tout réel $\delta > 0$ il existe une constante $c(\delta) > 0$ tel que, pour toute fonction $f: \mathbb{Z}_N \rightarrow \mathbb{R}$ avec*

$$0 \leq f(x) \leq 1 \text{ pour tout } x \text{ et } \mathbb{E}(f \mid \mathbb{Z}_N) \geq \delta$$

on ait

$$(1) \quad \mathbb{E}(f(x)f(x+t) \dots f(x+kt) \mid x, t \in \mathbb{Z}_N) \geq c(\delta) .$$

La version finie du théorème de Szemerédi se déduit de ce théorème en prenant pour f la fonction indicatrice d'un sous-ensemble de $[0, N[$. Green et Tao s'affranchissent de la condition $f \leq 1$ en la remplaçant par l'hypothèse que f est majorée par un *poids pseudo-aléatoire*; cette notion sera définie plus loin.

2.1. Les deux ingrédients de la preuve du théorème 1.1

Nous appelons «théorème de Green-Tao Szemerédi» l'extension suivante du théorème de Szemerédi :

THÉORÈME 2.2. — *Soit $\nu: \mathbb{Z}_N \rightarrow \mathbb{R}^+$ un poids pseudo-aléatoire (voir la sous-section 2.2). Pour tout réel $\delta > 0$ il existe une constante $c'(\delta) > 0$ satisfaisant la propriété suivante. Pour toute fonction $f: \mathbb{Z}_N \rightarrow \mathbb{R}$ telle que*

$$0 \leq f(x) \leq \nu(x) \text{ pour tout } x \text{ et } \mathbb{E}(f \mid \mathbb{Z}_N) \geq \delta$$

on a

$$(2) \quad \mathbb{E}(f(x)f(x+t) \dots f(x+kt) \mid x, t \in \mathbb{Z}_N) \geq c'(\delta) - o(1) .$$

La démonstration de ce théorème, qui occupe une part importante de l'article de Green et Tao, est résumée dans les sections 3 et 4. Pour l'appliquer aux nombres premiers, il faut une fonction f et un poids ν convenables dont l'existence est donnée par le théorème suivant.

THÉORÈME 2.3. — *Il existe une constante positive δ , un poids pseudo-aléatoire $\nu: \mathbb{Z}_N \rightarrow \mathbb{R}^+$ et une fonction $f: \mathbb{Z}_N \rightarrow \mathbb{R}$ avec*

f est nulle en dehors de l'ensemble des nombres premiers ;

$$0 \leq f(x) \leq \nu(x) \text{ pour tout } x ;$$

$$\mathbb{E}(f \mid \mathbb{Z}_N) \geq \delta ;$$

$$\|f\|_{L^\infty} = O(1) \log N .$$

La construction de f et ν est faite dans la section 5. Nous montrons maintenant comment le théorème 1.1 découle des théorèmes 2.2 et 2.3.

Démonstration. — Soient δ , f et ν comme dans le théorème 2.3. Il existe un intervalle $J \subset [0, N[$, de longueur plus petite que $N/2$ et tel que $\mathbb{E}(\mathbf{1}_J f \mid \mathbb{Z}_N) \geq \delta/3$. Nous utilisons le théorème 2.2 avec la fonction f remplacée par $\mathbf{1}_J f$ et le réel δ remplacé par $\delta/3$.

La contribution dans l'espérance (2) des termes où $t = 0$ est majorée par $N^{-1} \|f\|_{L^\infty}^{k+1} = o(1)$, et est donc inférieure à $c'(\delta)$ si N est assez grand. Il existe donc dans ce cas $x, t \in \mathbb{Z}_N$ avec $t \neq 0$ tels que $f(x)f(x+t) \dots f(x+kt) \neq 0$. Rappelons que dans cette expression $x, x+t, \dots, x+kt$ sont considérés comme des éléments de \mathbb{Z}_N et que donc l'addition est modulo N . Si nous considérons x et t comme des entiers appartenant à l'intervalle $[0, N[$ nous obtenons que f est non nulle aux points $x, x+t \bmod N, \dots, x+kt \bmod N$. Comme elle est nulle en dehors de l'intervalle J de longueur $< N/2$, tous ces entiers appartiennent à cet intervalle et on en déduit facilement qu'ils forment une progression arithmétique non triviale de longueur $k+1$. Enfin, f est nulle en dehors de l'ensemble des nombres premiers et on a bien une progression formée de nombres premiers. \square

2.2. Définition des poids pseudo-aléatoires

Dans les théorèmes précédents nous avons considéré un *poids pseudo-aléatoire* comme une fonction définie sur \mathbb{Z}_N . Il s'agit plus précisément de la donnée, pour chaque nombre premier N , d'une fonction $\nu = \nu_N: \mathbb{Z}_N \rightarrow \mathbb{R}^+$, de sorte que soient satisfaites deux conditions asymptotiques appelées *condition sur les formes linéaires* et *condition sur les corrélations*.

La condition sur les formes linéaires. — Ici m_0, t et L sont des constantes entières (ne dépendant que de k) que nous n'explicitons pas.

Soient $m \leq m_0$ un entier et ψ_1, \dots, ψ_m des applications de \mathbb{Z}_N^t dans \mathbb{Z}_N de la forme

$$(3) \quad \psi_i(\mathbf{x}) = b_i + \sum_{j=1}^t L_{i,j} x_j$$

où $\mathbf{x} = (x_1, \dots, x_t)$ et

- pour tout i , b_i est un entier;
- pour tous i, j , $L_{i,j}$ est un entier avec $|L_{i,j}| \leq L$;
- aucun des vecteurs $(L_{i,j})_{1 \leq j \leq t} \in \mathbb{Z}^t$ n'est nul et ces vecteurs sont deux à deux non colinéaires

alors la condition sur les formes linéaires stipule que

$$(4) \quad \mathbb{E}(\nu(\psi_1(\mathbf{x})) \dots \nu(\psi_m(\mathbf{x})) \mid \mathbf{x} \in \mathbb{Z}_N^t) = 1 + o(1) .$$

D'après nos conventions, la quantité $o(1)$ tend vers 0 quand N tend vers l'infini indépendamment du choix des fonctions ψ_i et en particulier du choix des b_i , qui ne sont pas supposés bornés. Remarquons que la condition des formes linéaires entraîne que la même majoration reste valable s'il y a moins de t variables. En particulier,

$$\mathbb{E}(\nu \mid \mathbb{Z}_N) = 1 + o(1) .$$

La condition des corrélations. — Ici encore, q_0 est une constante entière que nous n'explicitons pas.

La condition sur les corrélations stipule qu'il existe une fonction $\tau : \mathbb{Z}_N \rightarrow \mathbb{R}^+$ avec

$$\text{pour tout } p \geq 1, \mathbb{E}(\tau^p(x) \mid x \in \mathbb{Z}_N) = O_p(1)$$

telle que, pour tout $q \leq q_0$ et tous $h_1, \dots, h_q \in \mathbb{Z}_N$, distincts ou confondus, on ait

$$(5) \quad \mathbb{E}(\nu(x + h_1)\nu(x + h_2) \dots \nu(x + h_q) \mid x \in \mathbb{Z}_N) \leq \sum_{1 \leq i < j \leq q} \tau(h_i - h_j) .$$

Nous remarquons que, si ν est un poids pseudo-aléatoire, alors $(1 + \nu)/2$ en est également un. On peut donc sans perte de généralité se restreindre au cas où $\nu(x) > 0$ pour tout x .

3. LES NORMES DE GOWERS

3.1. La définition

Il y a quelques années Gowers a proposé une nouvelle preuve du théorème de Szemerédi [G] à base d'analyse harmonique et de combinatoire. Dans sa démonstration il a introduit une suite de normes sur l'espace $\mathcal{C}(\mathbb{Z}_N)$ des fonctions sur \mathbb{Z}_N à valeurs réelles et les a utilisées pour contrôler les espérances qui apparaissent dans le théorème 2.1. Green et Tao les utilisent également pour contrôler les espérances du théorème 2.2. Nous donnons ici leur définition.

Pour $f \in \mathcal{C}(\mathbb{Z}_N)$ on définit par récurrence les quantités $\|f\|_{U^d}$, $d \geq 1$, par

$$\|f\|_{U^1} = |\mathbb{E}(f \mid \mathbb{Z}_N)| ;$$

$$\|f\|_{U^{d+1}} = \left(\mathbb{E}(\|f \cdot f_t\|_{U^d}^{2^d} \mid t \in \mathbb{Z}_N) \right)^{1/2^{d+1}} \text{ pour } d \geq 1$$

où f_t est la fonction $x \mapsto f(x + t)$.

Ces quantités peuvent être aussi données par une formule close.

Pour $\boldsymbol{\omega} = (\omega_1, \dots, \omega_d) \in \{0,1\}^d$ et $\mathbf{t} = (t_1, \dots, t_d) \in \mathbb{Z}_N^d$ notons

$$\boldsymbol{\omega} \cdot \mathbf{t} = \omega_1 t_1 + \omega_2 t_2 + \dots + \omega_d t_d .$$

On a alors

$$(6) \quad \|f\|_{U^d} = \left(\mathbb{E} \left(\prod_{\boldsymbol{\omega} \in \{0,1\}^d} f(x + \boldsymbol{\omega} \cdot \mathbf{t} \mid x \in \mathbb{Z}_N, \mathbf{t} \in \mathbb{Z}_N^d) \right) \right)^{1/2^d} .$$

On vérifie alors facilement que $\|f\|_{U^2}$ est la norme ℓ^4 de la transformée de Fourier de f et que $\|f\|_{U^{d+1}} \geq \|f\|_{U^d}$ pour tout $d \geq 1$. De plus on obtient :

PROPOSITION 3.1 (Inégalité de Cauchy-Schwarz-Gowers). — *Si $f_{\boldsymbol{\omega}}$, $\boldsymbol{\omega} \in \{0,1\}^d$, sont 2^d fonctions réelles sur \mathbb{Z}_N on a*

$$(7) \quad \left| \mathbb{E} \left(\prod_{\boldsymbol{\omega} \in \{0,1\}^d} f_{\boldsymbol{\omega}}(x + \boldsymbol{\omega} \cdot \mathbf{t} \mid x \in \mathbb{Z}_N, \mathbf{t} \in (\mathbb{Z}_N)^d) \right) \right| \leq \prod_{\boldsymbol{\omega} \in \{0,1\}^d} \|f_{\boldsymbol{\omega}}\|_{U^d} .$$

On en déduit :

PROPOSITION 3.2. — *Pour $d \geq 2$ l'application $f \mapsto \|f\|_{U^d}$ est une norme sur $\mathcal{C}(\mathbb{Z}_N)$.*

On peut facilement étendre ces définitions au cas des fonctions à valeurs complexes.

3.2. Commentaires

Pour $d > 2$ la norme $\|\cdot\|_{U^d}$ est assez difficile à interpréter car elle ne peut apparemment pas être exprimée au moyen des normes classiques. La définition n'est pas simplifiée par l'usage de la transformée de Fourier ; par exemple, la norme $\|\cdot\|_{U^3}$ d'une fonction est la même que celle de sa transformée de Fourier (à une normalisation près).

Cette difficulté provient sans doute du fait que ces normes ont un aspect non commutatif. En effet, il est clairement possible de définir des normes similaires sur l'espace $\mathcal{C}_K(G)$ des fonctions continues à support compact sur un groupe abélien localement compact G . Mais il est sans doute moins évident que la définition de la norme $\|\cdot\|_{U^d}$ s'étend au cas où G est localement compact nilpotent d'ordre $d - 1$, et qu'elle peut même être définie sur $\mathcal{C}_K(G/\Gamma)$ lorsque Γ est un sous-groupe fermé d'un groupe G de ce type.

D'une manière indépendante, des semi-normes $\|\cdot\|_d$, $d \geq 1$, ont récemment été introduites [HK] en théorie ergodique dans l'étude de questions relatives au théorème de Szemerédi où elles servent également à contrôler des espérances ressemblant à celles du théorème 2.1. La définition de ces semi-normes, nettement plus compliquée, ne sera pas donnée ici mais elle est formellement assez similaire à celle des normes de Gowers. Ces semi-normes ont une interprétation simple : elles sont liées à l'existence de quotients du système munis d'une structure d'espace homogène d'un groupe de Lie nilpotent.

Si on admet que les ressemblances de plus en plus nombreuses qui apparaissent entre les deux théories ne sont pas fortuites, il est alors possible de conjecturer que les normes

de Gowers s'interprètent au moyen d'une sorte de transformée de Fourier nilpotente, même lorsque le groupe est abélien.

3.3. Normes de Gowers et progressions arithmétiques

La proposition suivante généralise un résultat analogue de Gowers établi sous l'hypothèse plus forte que toutes les fonctions sont bornées par 1. Sa démonstration consiste en une suite ingénieuse d'applications de l'inégalité de Cauchy-Schwarz, de changements de variables et de la condition sur les formes linéaires.

PROPOSITION 3.3. — *Soient ν un poids pseudo-aléatoire et f_0, f_1, \dots, f_k des fonctions sur \mathbb{Z}_N vérifiant*

$$|f_j(x)| \leq 1 + \nu(x) \text{ pour tout } x \in \mathbb{Z}_N \text{ et tout } j \text{ avec } 0 \leq j \leq k .$$

Alors

$$\left| \mathbb{E} \left(\prod_{j=0}^k f_j(x + jt) \mid x, t \in \mathbb{Z}_N \right) \right| \leq 2^{k+1} \inf_{0 \leq j \leq k} \|f_j\|_{U^k} + o(1) .$$

L'utilisation que font Green et Tao de cette proposition est très différente de la manière dont Gowers utilise le résultat analogue pour les fonctions bornées.

Ce dernier procède par dichotomie.

Soient f une fonction sur \mathbb{Z}_N et $c = \mathbb{E}(f \mid \mathbb{Z}_N)$. Si $\|f - c\|_{U^k}$ est petit, alors l'espérance (1) est peu différente de l'espérance obtenue en remplaçant f par c et elle est donc grande. Si au contraire cette norme est grande, alors Gowers montre que la restriction de f à un sous-ensemble pas trop petit de \mathbb{Z}_N présente des régularités qui sont ensuite exploitées.

Green et Tao utilisent une décomposition, où les normes de Gowers jouent un rôle très proche de celui joué par les semi-normes $\|\cdot\|_d$ dans [HK]. Lorsque f est une fonction majorée par un poids pseudo-aléatoire, elle peut s'écrire (essentiellement) comme la somme d'une fonction ayant une petite norme et d'une fonction bornée qui est son espérance conditionnelle sur une σ -algèbre (proposition 3.4). La proposition 3.3 permet alors de borner la contribution provenant de la fonction de petite norme. En théorie ergodique on écrit chaque fonction comme somme de son espérance conditionnelle sur une σ -algèbre adaptée et d'une fonction de semi-norme nulle. On utilise ensuite le fait que cette σ -algèbre a une interprétation «géométrique» assez simple.

3.4. σ -algèbres sur \mathbb{Z}_N et un résultat de décomposition

\mathbb{Z}_N étant fini, toute σ -algèbre \mathcal{B} sur \mathbb{Z}_N est définie par une partition de cet ensemble: les éléments de \mathcal{B} sont les réunions d'atomes de cette partition et les fonctions \mathcal{B} -mesurables sont les fonctions constantes sur chaque atome. Quand f est une fonction sur \mathbb{Z}_N , son espérance conditionnelle par rapport à \mathcal{B} est la fonction \mathcal{B} -mesurable définie par

$$\text{si } A \text{ est l'atome de } \mathcal{B} \text{ contenant } x, \mathbb{E}(f \mid \mathcal{B})(x) = \mathbb{E}(f \mid A) = \frac{\mathbb{E}(\mathbf{1}_A f \mid \mathbb{Z}_N)}{\mathbb{E}(\mathbf{1}_A \mid \mathbb{Z}_N)} .$$

La proposition suivante est la clé de la démonstration du théorème 2.2.

PROPOSITION 3.4. — *Soit ν un poids pseudo-aléatoire. Pour tout $\varepsilon > 0$ il existe un entier $N_0(\varepsilon)$ tel que pour tout $N > N_0(\varepsilon)$ on ait la propriété suivante.*

Soit f une fonction sur \mathbb{Z}_N avec $0 \leq f(x) \leq \nu(x)$ pour tout x . Alors il existe une σ -algèbre \mathcal{B} sur \mathbb{Z}_N , un sous-ensemble Ω de \mathbb{Z}_N appartenant à \mathcal{B} avec

$$(8) \quad \mathbb{E}(\nu \mathbf{1}_\Omega \mid \mathbb{Z}_N) = o_\varepsilon(1) ;$$

$$(9) \quad \|(1 - \mathbf{1}_\Omega) \mathbb{E}(\nu - 1 \mid \mathcal{B})\|_{L^\infty} = o_\varepsilon(1) ;$$

$$(10) \quad \|(1 - \mathbf{1}_\Omega) (f - \mathbb{E}(f \mid \mathcal{B}))\|_{U^k} \leq \varepsilon .$$

On donne dans la section 4 un résumé de la preuve de cette proposition. Admettant ce résultat pour le moment, nous indiquons comment on peut en déduire le théorème 2.2.

3.5. Démonstration du théorème 2.2 à partir des propositions 3.1 et 3.4

Soient ν, f et δ comme dans le théorème. Soient $\varepsilon > 0$ un paramètre suffisamment petit et \mathcal{B}, Ω comme dans la proposition 3.4. Nous supposons que N est suffisamment grand. Posons

$$g = (1 - \mathbf{1}_\Omega) \mathbb{E}(f \mid \mathcal{B}) \text{ et } h = (1 - \mathbf{1}_\Omega) (f - \mathbb{E}(f \mid \mathcal{B})) .$$

Comme $f \leq \nu$ nous avons

$$(11) \quad \mathbb{E}(g \mid \mathbb{Z}_N) \geq \mathbb{E}(f \mid \mathbb{Z}_N) - \mathbb{E}(\mathbf{1}_\Omega \mathbb{E}(\nu \mid \mathcal{B}) \mid \mathbb{Z}_N) = \mathbb{E}(f \mid \mathbb{Z}_N) - \mathbb{E}(\mathbf{1}_\Omega \nu \mid \mathbb{Z}_N) \geq \delta - o_\varepsilon(1)$$

car $\Omega \in \mathcal{B}$ et d'après (8). De plus

$$(12) \quad 0 \leq g \leq (1 - \mathbf{1}_\Omega) \mathbb{E}(\nu \mid \mathcal{B}) \leq 1 + o_\varepsilon(1)$$

d'après (9). Ainsi, $|h| \leq f + g \leq 1 + \nu + o_\varepsilon(1)$ et par ailleurs $\|h\|_{U^k} \leq \varepsilon$ d'après (10).

Comme $0 \leq g + h \leq f$, l'espérance (2) apparaissant dans le théorème est minorée par la même espérance avec f remplacée par $g + h$. Cette dernière expression s'écrit comme somme de 2^{k+1} espérances de la forme

$$(13) \quad \mathbb{E}(f_0(x) f_1(x+t) \dots f_k(x+kt) \mid x, t \in \mathbb{Z}_N)$$

où chacune des fonctions f_i , $0 \leq i \leq k$, est égale à g ou à h . Ainsi, $|f_i| \leq 1 + \nu + o_\varepsilon(1)$ pour tout i .

Le terme principal est celui où toutes les fonctions f_i sont égales à g ; en effet la majoration (12) permet d'utiliser le théorème de Szemerédi (théorème 2.1) et la minoration (11) entraîne donc que ce terme est minoré par $c(\delta - o_\varepsilon(1))$. Tous les autres termes ont une valeur absolue majorée par $2^{k+1}\varepsilon + o_\varepsilon(1)$ d'après la proposition 3.3.

En choisissant ε assez petit nous obtenons donc la minoration annoncée de l'espérance (2), avec $c'(\delta) = c(\delta)$. □

Remarque. — Le résultat obtenu est plus fort que ce qui est réellement nécessaire, à savoir $c'(\delta) > 0$. Il serait sans doute possible de modifier la démonstration en affaiblissant les conditions imposées à ν tout en conservant la propriété annoncée.

4. DÉMONSTRATION DE LA PROPOSITION 3.4

Cette section est la plus technique de ces notes et les lecteurs qui ne seraient pas intéressés par les détails sont invités à passer directement à la suivante.

4.1. Les fonctions duales

Soit f une fonction réelle sur \mathbb{Z}_N . Pour $x \in \mathbb{Z}_N$ définissons

$$\mathcal{D}f(x) = \mathbb{E} \left(\prod_{\substack{\boldsymbol{\omega} \in \{0,1\}^k \\ \boldsymbol{\omega} \neq \mathbf{0}}} f(x + \boldsymbol{\omega} \cdot \mathbf{t}) \mid \mathbf{t} \in \mathbb{Z}_N^k \right)$$

où $\mathbf{0}$ représente l'élément $(0, 0, \dots, 0)$ de $\{0, 1\}^k$. $\mathcal{D}f$ est appelée la *fonction duale* (d'ordre k) de f .

Écrivons $\langle \cdot, \cdot \rangle$ le produit scalaire sur $\mathcal{C}(\mathbb{Z}_N)$ donné par $\langle f, g \rangle = \mathbb{E}(fg \mid \mathbb{Z}_N)$. La définition des normes et l'inégalité de Cauchy-Schwarz-Gowers entraînent immédiatement :

LEMME 4.1. — *Pour toute fonction $f \in \mathcal{C}(\mathbb{Z}_N)$,*

$$\|f\|_{U^k} = \langle f, \mathcal{D}f \rangle = \sup \{ |\langle f, \mathcal{D}g \rangle| ; g \in \mathcal{C}(\mathbb{Z}_N), \|g\|_{U^k} \leq 1 \} .$$

Ainsi, la boule unité pour la norme duale de $\|\cdot\|_{U^k}$ est l'enveloppe convexe de l'ensemble $\{\mathcal{D}f ; \|f\|_{U^k} \leq 1\}$. Cette norme duale n'est malheureusement pas une norme d'algèbre (la norme d'un produit n'est pas majorée par le produit des normes), ce qui simplifierait beaucoup la démonstration. Dans sa preuve du théorème de Szemerédi [T1], Tao construit une norme d'algèbre qui est majorée par la norme duale. Cette construction est formellement très proche de la construction de la «tour d'extensions isométriques» de Furstenberg.

Pour comprendre le rôle joué par les fonctions duales, imaginons la situation où nous avons une fonction f telle que $\|f\|_{U^k}$ soit «grande» et que $\|\mathcal{D}f\|_{L^2}$ ne soit pas «trop grande». Supposons aussi que nous savons construire une σ -algèbre \mathcal{B} par rapport à laquelle $\mathcal{D}f$ est mesurable au moins approximativement. Alors, comme le produit scalaire de f et $\mathcal{D}f$ est grand, l'espérance $\mathbb{E}(f \mid \mathcal{B})$ aura une norme L^2 assez grande. Cette méthode est utilisée de manière itérative dans les sous-sections suivantes pour construire la σ -algèbre de la proposition 3.4.

4.2. Poids pseudo-aléatoires et fonctions duales

Dans toute la suite de cette section, ν désigne un poids pseudo-aléatoire et nous étudions les propriétés des fonctions duales des fonctions majorées par ν ou par $1 + \nu$.

Rappelons que la condition sur les formes linéaires entraîne que $\mathbb{E}(\nu \mid \mathbb{Z}_N) = 1 + o(1)$. On a plus précisément

$$(14) \quad \|\nu - 1\|_{U^k} = o(1) .$$

Nous obtenons de même :

LEMME 4.2. — Si f est une fonction sur \mathbb{Z}_N vérifiant $|f| \leq 1 + \nu$ alors $\|\mathcal{D}f\|_{L^\infty} \leq 2^{2^k-1} + o(1)$.

Nous notons désormais I un intervalle fermé borné de \mathbb{R} tel que $\mathcal{D}f(x) \in I$ pour tout x et toute fonction f avec $|f| \leq 1 + \nu$.

PROPOSITION 4.3. — Soient $m \geq 1$ un entier et f_1, \dots, f_m des fonctions sur \mathbb{Z}_N vérifiant $|f_i| \leq 1 + \nu$ pour tout i , et soit Φ une fonction continue sur le cube I^m . Alors la fonction ψ sur \mathbb{Z}_N définie par

$$\psi(x) = \Phi(\mathcal{D}f_1(x), \dots, \mathcal{D}f_m(x))$$

satisfait la relation

$$\langle \nu - 1, \psi \rangle = o_{m, \Phi}(1) .$$

De plus, cette estimation est uniforme en Φ si l'on impose à cette fonction de rester dans un compact au sens de la convergence uniforme.

Pour montrer cette proposition on se ramène facilement au cas où $\Phi(x_1, x_2, \dots, x_m) = x_1 x_2 \dots x_m$ et on utilise la condition des corrélations. C'est le seul endroit de la preuve où cette condition est utilisée.

4.3. Construction d'une σ -algèbre

Nous introduisons ici une construction qui sera utilisée de manière répétée dans la section suivante pour montrer la proposition 3.4. Ici $\varepsilon > 0$ est un paramètre et $\sigma \in]0, 1/2[$ est un paramètre accessoire qui devra être choisi soigneusement en fonction de ε . On se donne une fonction continue $\psi: \mathbb{R} \rightarrow [0, 1]$, à support dans $[0, 1]$ et égale à 1 sur $[\sigma, 1 - \sigma]$. Nous supposons toujours que N est suffisamment grand.

Soit f une fonction sur \mathbb{Z}_N avec $|f| \leq 1 + \nu$ et notons $F = \mathcal{D}f$.

Soient $\alpha \in]0, 1]$ et \mathcal{B} la σ -algèbre dont les atomes sont les ensembles A de la forme

$$(15) \quad A = \{x \in \mathbb{Z}_N ; \varepsilon^{2^{k+1}}(n + \alpha) \leq F(x) < \varepsilon^{2^{k+1}}(n + 1 + \alpha)\}$$

où n est un entier tel que cet ensemble ne soit pas vide. Le paramètre α est introduit pour éviter les effets de bord : il pourrait en effet arriver que les valeurs de la fonction F s'accroissent près des points $n\varepsilon^{2^{k+1}}$ mais, pour un choix convenable de α , l'ensemble

$$E = \bigcup_{n \in \mathbb{Z}} \{x \in \mathbb{Z}_N ; \varepsilon^{2^{k+1}}(n + \alpha - \sigma) \leq F(x) \leq \varepsilon^{2^{k+1}}(n + \alpha + \sigma)\}$$

vérifie

$$(16) \quad \mathbb{E}(\mathbf{1}_E(1 + \nu) \mid \mathbb{Z}_N) = \sigma O(1) .$$

Par construction, on a clairement,

$$(17) \quad \|F - \mathbb{E}(F \mid \mathcal{B})\|_{L^\infty} \leq \varepsilon^{2^{k+1}}$$

Comme F est bornée (lemme 4.2), le nombre d'atomes de \mathcal{B} est un $O_\varepsilon(1)$.

Appelons un atome A de \mathcal{B} *mauvais* si $\mathbb{E}((1 + \nu)\mathbf{1}_A \mid \mathbb{Z}_N) < \sigma^{1/2}$ et notons Ω la réunion des mauvais atomes. Alors $\Omega \in \mathcal{B}$ et

$$(18) \quad \mathbb{E}((1 + \nu)\mathbf{1}_\Omega \mid \mathbb{Z}_N) = \sigma^{1/2} O_\varepsilon(1) .$$

Soient maintenant A un *bon* atome, n l'entier correspondant dans la définition (15) et $J = [\varepsilon^{2^{k+1}}(n + \alpha), \varepsilon^{2^{k+1}}(n + 1 + \alpha)[$. Posons $\Phi_A(x) = \psi(\varepsilon^{-2^{k+1}}(x - n - \alpha))$. La proposition 4.3 permet de majorer $\Phi_A \circ F$ et la propriété (16) permet de contrôler le terme d'erreur $\mathbf{1}_A - \Phi_A \circ F = (\mathbf{1}_J - \Phi_A) \circ F$. Nous obtenons

$$\mathbb{E}((\nu(x) - 1)\mathbf{1}_A(x) \mid x \in A) = \sigma^{1/2} O_\varepsilon(1) + o_{\varepsilon, \sigma}(1) .$$

On en déduit que

$$(19) \quad \|(1 - \mathbf{1}_\Omega) \mathbb{E}(\nu - 1 \mid \mathcal{B})\|_{L^\infty} = o_\varepsilon(1) .$$

Supposons maintenant qu'au lieu d'une fonction f nous avons une famille finie (f_1, \dots, f_m) de fonctions vérifiant toutes $|f_i| \leq 1 + \nu$. Alors par la même méthode nous pouvons construire une σ -algèbre \mathcal{B} et un ensemble $\Omega \in \mathcal{B}$ vérifiant les propriétés (18) et (19) et tels que l'approximation uniforme (17) soit valable pour chacune des fonctions $F_i = \mathcal{D}f_i$, tous les termes $o(1)$ et $O(1)$ dépendant aussi du nombre m de fonctions.

4.4. Une récurrence

Soit maintenant f une fonction avec $0 \leq f \leq \nu$. Nous allons utiliser la construction précédente de façon répétée, construisant de proche en proche une suite (f_j) de fonctions, une suite (\mathcal{B}_j) de σ -algèbres et une suite (Ω_j) d'ensembles appartenant à \mathcal{B}_j .

Posons $f_1 = f$ et soient \mathcal{B}_1 la σ -algèbre grossière $\{\emptyset, \mathbb{Z}_N\}$ et $\Omega_1 = \emptyset$.

Supposons que les constructions ont été faites jusqu'au rang j . Nous posons

$$f_{j+1} = (1 - \mathbf{1}_{\Omega_j})(f - \mathbb{E}(f \mid \mathcal{B}_j))$$

et distinguons deux cas.

- Si $\|f_{j+1}\|_{U^k} \leq \varepsilon$ nous arrêtons l'algorithme;
- sinon nous employons la méthode précédente avec la famille de fonctions (f_1, \dots, f_{j+1}) pour définir la σ -algèbre \mathcal{B}_{j+1} et l'ensemble Ω_{j+1} et nous itérons l'algorithme, ce qui est possible car $|f_{j+1}| \leq 1 + \nu$, à multiplication près par un terme de la forme $1 + \sigma^{1/2} O_\varepsilon(1)$.

Montrons que cet algorithme s'arrête après un nombre borné d'étapes. S'il ne s'arrête pas à l'étape j alors $\|f_{j+1}\|_{U^k}^{2^k} = \mathbb{E}(f_{j+1} \mathcal{D}f_{j+1}) \geq \varepsilon^{2^k}$. Imaginons pour simplifier que le petit ensemble Ω_j est vide. Nous aurions alors $f_{j+1} = f - \mathbb{E}(f \mid \mathcal{B}_j)$ et comme par construction $\mathcal{D}f_{j+1}$ est uniformément proche de $\mathbb{E}(\mathcal{D}f_{j+1} \mid \mathcal{B}_{j+1})$, cette inégalité nous permettrait de minorer

$$\mathbb{E}\left((f - \mathbb{E}(f \mid \mathcal{B}_j)) \mathbb{E}(\mathcal{D}f_{j+1} \mid \mathcal{B}_{j+1})\right) = \mathbb{E}\left((\mathbb{E}(f \mid \mathcal{B}_{j+1}) - \mathbb{E}(f \mid \mathcal{B}_j)) \mathbb{E}(\mathcal{D}f_{j+1} \mid \mathcal{B}_{j+1})\right)$$

et donc aussi $\|\mathbb{E}(f \mid \mathcal{B}_{j+1}) - \mathbb{E}(f \mid \mathcal{B}_j)\|_{L^2}$ en appliquant l'inégalité de Cauchy-Schwarz. Les calculs précis permettent en fait de borner inférieurement la quantité positive

$$\|(1 - \mathbf{1}_{\Omega_{j+1}}) \mathbb{E}(f \mid \mathcal{B}_{j+1})\|_{L^2}^2 - \|(1 - \mathbf{1}_{\Omega_j}) \mathbb{E}(f \mid \mathcal{B}_j)\|_{L^2}^2 .$$

Comme $\|(1 - \mathbf{1}_{\Omega_j}) \mathbb{E}(f \mid \mathcal{B}_j)\|_{L^2}$ est majoré par $\|f\|_{L^2}$, cela prouve que l'algorithme s'arrête en temps borné.

La dernière σ -algèbre et le dernier ensemble Ω construits vérifient alors les propriétés de la proposition 3.4. \square

5. UN POIDS PSEUDO-ALÉATOIRE

Il nous reste à construire une fonction f nulle en dehors des nombres premiers et un poids pseudo-aléatoire ν vérifiant les hypothèses du théorème 2.2.

Dans cette section nous notons P l'ensemble des nombres premiers et la lettre p désigne toujours un nombre premier. Rappelons la définition de deux fonctions classiques.

- ϕ est la fonction indicatrice d'Euler : pour tout entier $x > 0$, $\phi(x)$ est le nombre d'entiers compris entre 1 et x et premiers avec x .
- μ est la fonction de Möbius :

$$\mu(x) = \begin{cases} 1 & \text{si } x = 0 ; \\ (-1)^\ell & \text{si } x \text{ est le produit de } \ell \text{ nombres premiers distincts ;} \\ 0 & \text{sinon.} \end{cases}$$

5.1. La fonction f et le poids ν

Nous nous donnons une fois pour toutes

- une fonction $\chi: \mathbb{R} \rightarrow \mathbb{R}^+$, de classe \mathcal{C}^∞ , à support dans $[-1,1]$, avec $\chi(0) > 0$ et $\int_0^\infty (\chi'(x))^2 dx = 1$;
- $w = w(N)$ une fonction entière tendant vers l'infini très lentement, par exemple de l'ordre de $\log \log N$.

De plus nous notons

- $W = W(N)$ le produit des nombres premiers $\leq w(N)$;
- $R = N^\alpha$ où α est une constante positive suffisamment petite.

Soit b un entier qui sera défini plus bas ; définissons la fonction f sur $[0, N[$ par

$$(20) \quad f(n) = \begin{cases} \frac{\phi(W)}{W} \log(Wn + b) & \text{si } Wn + b \geq R \text{ et } Wn + b \text{ est premier} \\ 0 & \text{sinon.} \end{cases}$$

Rappelons l'estimation élémentaire de Tchébychev pour le nombre $\pi(x)$ de nombres premiers inférieurs ou égaux à x :

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}$$

où c_1 et c_2 sont des constantes positives. On en déduit (sans utiliser le théorème de Dirichlet) qu'on peut choisir $b \in [1, W[$, premier avec W et tel que pour N assez grand on ait $\mathbb{E}(f(n) \mid n \in [0, N]) \geq \delta$, où $\delta > 0$ est une constante.

Pour tout n entier nous posons

$$(21) \quad \lambda(n) = \sum_{d|n} \mu(d) \chi\left(\frac{\log d}{\log R}\right)$$

et nous définissons

$$(22) \quad \nu(n) = \frac{\phi(W)}{W} \log R \cdot \lambda^2(Wn + b) .$$

On vérifie facilement que pour tout $n \in [0, N[$ on a $0 \leq f(n) \leq c\nu(n)$ pour une certaine constante positive c . Ainsi la fonction f vérifie (à une normalisation près) les propriétés annoncées dans le théorème 2.3.

Il reste à montrer que ν est un poids pseudo-aléatoire, c'est-à-dire que cette fonction vérifie la condition sur les formes linéaires et la condition sur les corrélations. Nous ne donnons pas ici la preuve de ces propriétés et nous contentons de montrer comment obtenir l'estimation

$$\mathbb{E}(\nu \mid \mathbb{Z}_N) = 1 + o(1) .$$

La méthode qui suit n'est sans doute pas la plus simple possible mais elle contient les principaux ingrédients utilisés dans la démonstration complète et éclaire le rôle joué par le paramètre R et par la fonction χ . Intuitivement, le rôle du paramètre w est d'éliminer les perturbations produites par les petits nombres premiers. Soit en effet $p \in P$. Les nombres premiers se répartissent dans $p - 1$ classes de congruence modulo p , ce qui cause une irrégularité d'ordre $1/p$, non négligeable si p est trop petit devant N .

5.2. Une réécriture

$\mathbb{E}(\nu \mid \mathbb{Z}_N)$ est le produit par $C = W^{-1}\phi(W) \log R$ de

$$(23) \quad \sum_{d, d'} \mu(d)\mu(d') \chi\left(\frac{\log d}{\log R}\right) \chi\left(\frac{\log d'}{\log R}\right) \mathbb{E}(\mathbf{1}_{d, d'}(x) \mid x \in \mathbb{Z}_N)$$

où d, d' sont des entiers positifs et

$$\mathbf{1}_{d, d'}(x) = \begin{cases} 1 & \text{si PPCM}(d, d') \text{ divise } Wx + b; \\ 0 & \text{sinon.} \end{cases}$$

Nous évaluons $\mathbb{E}(\mathbf{1}_{d, d'} \mid \mathbb{Z}_N)$. À cause des facteurs $\mu(d)$ et $\mu(d')$ nous pouvons nous restreindre au cas où d et d' sont sans carré et écrire

$$d = \prod_p p^{\omega_p} ; \quad d' = \prod_p p^{\omega'_p}$$

où $\omega = (\omega_p ; p \in P)$ et $\omega' = (\omega'_p ; p \in P)$ sont des suites à valeurs dans $\{0,1\}$. Définissons $E_p(\eta)$ pour $p \in P$ et $\eta \in \{0,1\}$ par

$$(24) \quad E_p(0) = 1 ; E_p(1) = 0 \text{ si } p \leq w \text{ et } E_p(1) = \frac{1}{p} \text{ si } p > w .$$

Nous avons

$$(25) \quad \mathbb{E}(\mathbf{1}_{d,d'} | \mathbb{Z}_N) = \prod_p E_p(\max(\omega_p, \omega'_p)) + N^{-1} O(1) .$$

En effet, s'il existe un premier p avec $p \leq w$ et $\max(\omega_p, \omega'_p) = 1$ alors $\mathbf{1}_{d,d'}(x) = 0$ pour tout x puisque p divise W et est premier avec b . Dans le cas contraire la proportion des $x \in \mathbb{Z}_N$ tels que PPCM(d, d') divise $Wx + b$ est $1/\text{PPCM}(d, d')$ à un $N^{-1} O(1)$ près.

Nous comprenons maintenant le rôle de la troncature effectuée par la fonction χ dans la définition de ν : pour tous les d considérés, les entiers $1, \dots, N-1$ se répartissent suffisamment uniformément dans les classes de congruence modulo d . En effet, la somme (23) contient au maximum R^2 termes non nuls et la somme des erreurs $N^{-1} O(1)$ est donc de la forme $R^2 N^{-1} O(1)$; le produit par C de cette expression est un $o(1)$ puisque R est une petite puissance de N .

Ainsi, en reportant l'estimation (25) dans (23) et en remplaçant la fonction de Möbius par sa définition, nous obtenons

$$(26) \quad \mathbb{E}(\nu | \mathbb{Z}_N) = o(1) + C \sum_{\omega, \omega'} \chi\left(\frac{\sum_p \omega_p \log p}{\log R}\right) \chi\left(\frac{\sum_p \omega'_p \log p}{\log R}\right) \prod_p (-1)^{\omega_p + \omega'_p} E_p(\max(\omega_p, \omega'_p)) .$$

5.3. Transformée de Fourier

Dans l'article [GT], Green et Tao utilisent une troncature brutale de la somme (21) définissant la fonction λ , comme dans les travaux de Goldston et Yıldırım. L'emploi de la fonction lisse χ permet de se servir de la transformée de Fourier. Nous écrivons

$$(27) \quad \chi(x) = \int \tau(t) e^{-x(1+it)} dt$$

où τ est une fonction à décroissance rapide. La somme dans la formule (26) se met alors aisément sous la forme

$$(28) \quad \iint \tau(t) \tau(t') \prod_p Z_p(t, t') dt dt'$$

où, en notant

$$z = \frac{1+it}{\log R} \text{ et } z' = \frac{1+it'}{\log R}$$

nous avons

$$Z_p(t, t') = \sum_{\eta, \eta' \in \{0,1\}} (-1)^{\eta+\eta'} E_p(\max(\eta, \eta')) p^{-\eta z - \eta' z'} .$$

En remplaçant les E_p par leurs valeurs (24) nous obtenons

$$Z_p(t, t') = \begin{cases} 1 & \text{si } p \leq w; \\ 1 - p^{-1-z} - p^{-1-z'} + p^{-1-z-z'} & \text{si } p > w. \end{cases}$$

Il vient alors

$$\begin{aligned} \prod_p Z_p(t, t') &= \prod_{p>w} (1 - p^{-1-z} - p^{-1-z'} + p^{-1-z-z'}) \\ &= (1 + o(1)) \prod_{p>w} \frac{(1 - p^{-1-z})(1 - p^{-1-z'})}{1 - p^{-1-z-z'}} \end{aligned}$$

car w tend vers l'infini avec N et donc

$$(29) \quad \prod_p Z_p(t, t') = (1 + o(1)) \frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')} \left(\prod_{p \leq w} \frac{(1 - p^{-1-z})(1 - p^{-1-z'})}{1 - p^{-1-z-z'}} \right)^{-1}.$$

Nous écrivons l'intégrale (28) comme somme de l'intégrale pour t, t' appartenant à l'intervalle $J = [-(\log R)^{1/2}, (\log R)^{1/2}]$ et d'un reste qui est de la forme $C^{-1}o(1)$ à cause de la décroissance rapide de la fonction τ .

Soient maintenant t et t' appartenant à J .

Comme W est très petit devant $(\log R)^{1/2}$ nous avons

$$\prod_{p \leq w} \frac{(1 - p^{-1-z})(1 - p^{-1-z'})}{1 - p^{-1-z-z'}} = (1 + o(1)) \prod_{p \leq w} (1 - p^{-1}) = (1 + o(1)) \frac{\phi(W)}{W}.$$

D'autre part, l'estimation élémentaire

$$\zeta(1+s) \sim \frac{1}{s} \text{ quand } s \rightarrow 0 \text{ avec } \operatorname{Re}(s) > 0$$

nous donne

$$\frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')} = (1 + o(1)) \frac{zz'}{z+z'} = (1 + o(1)) \frac{1}{\log R} \frac{(1+it)(1+it')}{2+it+it'}.$$

En reportant ces valeurs dans (29) nous obtenons

$$\iint_{J \times J} \prod_p Z_p(t, t') = C^{-1} (1 + o(1)) \iint_{J \times J} \tau(t)\tau(t') \frac{(1+it)(1+it')}{2+it+it'} dt dt'.$$

Comme τ est à décroissance rapide, l'intégrale de la deuxième fonction en dehors de $J \times J$ est de un $C^{-1}o(1)$ et nous obtenons

$$\iint \prod_p Z_p(t, t') = C^{-1} (1 + o(1)) \iint \tau(t)\tau(t') \frac{(1+it)(1+it')}{2+it+it'} dt dt'.$$

Cette dernière intégrale s'écrit

$$\int_0^{+\infty} ds \iint \tau(t)\tau(t')(1+it)(1+it')e^{-s(2+it+it')} dt dt' = \int_0^{+\infty} (\chi'(s))^2 ds = 1$$

et on a donc bien $\mathbb{E}(\nu | \mathbb{Z}_N) = 1 + o(1)$. □

RÉFÉRENCES

- [ET] P. Erdős & T. Turán. On some sequences of integers. *J. London Math. Soc.* **11** (1936), 261–264.
- [F] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, **31** (1977), 204–256.
- [FKO] H. Furstenberg, Y. Katznelson & D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc.* **7** (1982), 527–552.
- [GY] D. Goldston & C.Y. Yıldırım. Small gaps between primes. Prépublication.
- [G] T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
- [GT] B. Green & T. Tao. The primes contain arbitrarily long arithmetic progressions. À paratre dans *Annals of Maths*. Prépublication disponible à : [http://arxiv.org \(math.NT/0404188\)](http://arxiv.org/math.NT/0404188) (2004).
- [HK] B. Host & B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.*, **161** (2005), 397–488.
- [HL] G. H. Hardy & J. E. Littlewood. Some problems of “partition numerorum” III: on the expression of a number as a sum of primes. *Acta. Math.* **44** (1923), 1–70.
- [S] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 199–245.
- [T1] T. Tao. A quantitative ergodic proof of Szemerédi's theorem. Prépublication (2004) disponible à : [http://arxiv.org \(math.CO/0405221\)](http://arxiv.org/math.CO/0405221)
- [T2] T. Tao. A remark on Goldston-Yıldırım correlation estimates. Prépublication disponible à : <http://www.math.ucla.edu/~tao/preprints>
- [vdC] J.G. van der Corput. Über Summen von Primzahlen und Primzahlquadraten. *Math. Ann.* **116** (1939), 1–50.

Bernard HOST

Université Marne-La-Vallée

UMR 8050 du CNRS

UFR de Mathématiques

Cité Descartes

5, boulevard Descartes

Champs-sur-Marne

F-77454 MARNE-LA-VALLÉE Cédex 2

E-mail : Bernard.Host@univ-mlv.fr

THE MUMFORD CONJECTURE
[after Madsen and Weiss]

by **Geoffrey POWELL**

1. INTRODUCTION

The Mumford conjecture concerns the cohomology of the moduli space \mathcal{M}_g of smooth projective curves of genus g : Mumford constructed tautological classes κ_i , for $i \geq 1$, in the Chow ring $CH^i(\mathcal{M}_g)$ with rational coefficients, which yield a natural morphism of algebras $\mathbb{Q}[\kappa_i] \rightarrow CH^*(\mathcal{M})$, in which $CH^*(\mathcal{M})$ denotes the Chow ring of the moduli spaces, stabilized with respect to the genus. The conjecture asserts that the above morphism is an isomorphism [17, 8].

The conjecture can be reformulated in terms of the stable cohomology of the mapping class groups (or Teichmüller modular groups) Γ_g [5, 16]. The mapping class group Γ_g is the discrete group of isotopy classes of orientation-preserving diffeomorphisms of a smooth, oriented surface of genus g . The group cohomology $H^*(B\Gamma_g)$ of the mapping class groups stabilizes in a given degree for sufficiently large genus. The stable value identifies with the cohomology of the space $B\Gamma_\infty$, which is the homotopy colimit of the system of classifying spaces $B\Gamma_{g,2}$ of the mapping class groups of curves with two marked points, stabilized with respect to maps induced by group morphisms $\Gamma_{g,2} \rightarrow \Gamma_{g+1,2}$.

The moduli space \mathcal{M}_g can be constructed, as an analytic space, as the quotient of the action of the group Γ_g upon Teichmüller space, \mathcal{T}_g . Teichmüller space is contractible and the action has finite isotropy groups, hence the Mumford conjecture can be restated in terms of the Mumford-Morita-Miller characteristic classes [14, 15], $\kappa_i \in H^{2i}(B\Gamma_\infty; \mathbb{Q})$.

CONJECTURE 1.1. — *The classes $\kappa_i \in H^{2i}(B\Gamma_\infty; \mathbb{Q})$ induce an isomorphism of algebras $\tilde{\alpha} : \mathbb{Q}[\kappa_i] \rightarrow H^*(B\Gamma_\infty; \mathbb{Q})$.*

The algebra $H^*(B\Gamma_\infty; \mathbb{Q})$ has a Hopf algebra structure, induced by a multiplicative structure of geometric origin on the classifying space $B\Gamma_\infty$. The classes κ_i are primitive and non-trivial, thus the morphism $\tilde{\alpha}$ is a monomorphism of Hopf algebras.

The space $B\Gamma_\infty$ has a structure which enriches the multiplicative structure; namely, the space $B\Gamma_\infty$ has a perfect fundamental group, hence the Quillen plus construction applies to yield a morphism $B\Gamma_\infty \rightarrow B\Gamma_\infty^+$, which induces an isomorphism in homology and such that $B\Gamma_\infty^+$ has trivial fundamental group. Tillmann [21] showed that the

space $\mathbb{Z} \times B\Gamma_\infty^+$ is an infinite loop space, hence it represents the degree zero part of a generalized cohomology theory; the identification of the associated cohomology theory is a problem of *stable* homotopy theory.

The construction of the Mumford-Morita-Miller characteristic classes uses integration along the fibre of powers of the first Chern class of the orientation bundle of the universal oriented surface bundle. This can be interpreted in terms of the Gysin morphism, which is of topological origin, via the Pontrjagin-Thom construction. Madsen and Tillmann [11] constructed a morphism of infinite loop spaces

$$\alpha_\infty : \mathbb{Z} \times B\Gamma_\infty^+ \rightarrow \Omega^\infty(\mathbb{C}\mathbf{P}_{-1}^\infty)$$

which lifts the construction of $\tilde{\alpha}$. The infinite loop space $\Omega^\infty(\mathbb{C}\mathbf{P}_{-1}^\infty)$ is constructed from the Thom spectrum which is associated to the complements of the canonical line bundles on complex projective space.

The rational cohomology of the space $\Omega^\infty(\mathbb{C}\mathbf{P}_{-1}^\infty)$ is isomorphic to the rational cohomology of the space $\mathbb{Z} \times BU$, where BU denotes the classifying space of the infinite unitary group. The cohomology algebra $H^*(BU; \mathbb{Q})$ is isomorphic to the polynomial algebra $\mathbb{Q}[\kappa_i]$, where the classes κ_i can be taken to be Chern classes, hence the Mumford conjecture is implied by the following result, which is referred to as the generalized Mumford conjecture.

THEOREM 1.2 ([12]). — *The morphism $\alpha_\infty : \mathbb{Z} \times B\Gamma_\infty^+ \rightarrow \Omega^\infty(\mathbb{C}\mathbf{P}_{-1}^\infty)$ is a homotopy equivalence.*

The cohomology of the space $\Omega^\infty(\mathbb{C}\mathbf{P}_{-1}^\infty)$ with coefficients in a finite field \mathbb{F}_p has been calculated [4], using techniques of algebraic topology. The above theorem therefore yields a calculation of the stable cohomology of the mapping class groups $H^*(B\Gamma_\infty; \mathbb{F}_p)$, for any prime p .

1.1. Methods of proof

Madsen and Weiss reformulate the generalized Mumford conjecture using certain generalized bundle theories; these are local in nature and their classifying spaces can be constructed from realization spaces associated to sheaves of sets. In particular, they give an interpretation of a modification of the morphism α_∞ introduced in [11] as the realization of a morphism of sheaves.

Let \mathfrak{X} denote the category of smooth manifolds, without boundary and with a countable basis and consider sheaves of sets on \mathfrak{X} . There is a natural notion of homotopy on the sections of a sheaf, termed *concordance*; if \mathcal{F} is a sheaf and X is a smooth manifold, then concordance is an equivalence relation on the sections $\mathcal{F}(X)$, which is induced by elements of $\mathcal{F}(X \times \mathbb{R})$, in the usual way. The set of equivalence classes for the concordance relation is written as $\mathcal{F}[X]$. The contravariant functor $X \mapsto \mathcal{F}[X]$ is represented on the homotopy category of topological spaces by a space $|\mathcal{F}|$; namely $\mathcal{F}[X] \cong [X, |\mathcal{F}|]$, where the right hand side denotes homotopy classes of morphisms of topological spaces.

A morphism $f : \mathcal{E} \rightarrow \mathcal{F}$ of sheaves is defined to be a weak equivalence if the induced morphism $|f| : |\mathcal{E}| \rightarrow |\mathcal{F}|$ is a weak equivalence⁽¹⁾ between the representing spaces.

There are two principal techniques which are used to show that a morphism between sheaves is a weak equivalence: to exhibit explicit concordances so as to obtain an isomorphism of concordance classes or to use the relative surjectivity criterion of Proposition A.7 to show that a morphism is a weak equivalence.

The classifying space $B\Gamma_g$ classifies bundles with fibres which are closed oriented surfaces, hence the source of the morphism α_∞ is related to bundles of closed oriented surfaces. This motivates consideration of the sheaf \mathcal{V} with sections over X the set of pairs (π, f) , where $\pi : E \rightarrow X$ is a smooth submersion with 3-dimensional oriented fibres and $f : E \rightarrow \mathbb{R}$ is a smooth morphism such that (π, f) is a proper submersion. Ehresmann’s fibration lemma implies that this is a bundle of smooth surfaces on $X \times \mathbb{R}$.

The definition of \mathcal{V} can be weakened: let $h\mathcal{V}$ denote the sheaf with sections over X the set of pairs (π, \hat{f}) , where $\pi : E \rightarrow X$ is as before and \hat{f} is a smooth section of the fibrewise 1-jet bundle $J_\pi^1(E, \mathbb{R}) \rightarrow E$, subject to the condition that the morphism $(\pi, f) : E \rightarrow X \times \mathbb{R}$ is a proper submersion, where f denotes the underlying smooth map, $f : E \rightarrow \mathbb{R}$, of \hat{f} . There is a morphism of sheaves $\alpha : \mathcal{V} \rightarrow h\mathcal{V}$, induced by jet prolongation, which induces a morphism of topological spaces $|\alpha| : |\mathcal{V}| \rightarrow |h\mathcal{V}|$, which is related to the morphism α_∞ .

These definitions generalize; namely it is expedient to allow mild fibrewise singularities over $X \times \mathbb{R}$, by considering smooth sections of the fibrewise 2-jet bundle $J_\pi^2(E, \mathbb{R}) \rightarrow E$ and permitting fibrewise critical points which are of Morse type. This gives sheaves $\mathcal{W}, h\mathcal{W}$, where \mathcal{W} corresponds to the integrable situation, as above. Similarly, there are sheaves $\mathcal{W}_{\text{loc}}, h\mathcal{W}_{\text{loc}}$ which correspond to the local situation around the singular sets and these sheaves form a commutative diagram

$$(1) \quad \begin{array}{ccccc} \mathcal{V} & \longrightarrow & \mathcal{W} & \longrightarrow & \mathcal{W}_{\text{loc}} \\ j_\pi^2 \downarrow & & j_\pi^2 \downarrow & & \downarrow j_\pi^2 \\ h\mathcal{V} & \longrightarrow & h\mathcal{W} & \longrightarrow & h\mathcal{W}_{\text{loc}}. \end{array}$$

The first main theorem of Vassiliev on the space of functions with moderate singularities is used to show the following Theorem, which motivates the strategy of proof.

THEOREM 1.3. — *The morphism $j_\pi^2 : \mathcal{W} \rightarrow h\mathcal{W}$ is a weak equivalence.*

This result is used in conjunction with the following, which is proved using bordism theory.

THEOREM 1.4. —

1. *The morphism $j_\pi^2 : \mathcal{W}_{\text{loc}} \rightarrow h\mathcal{W}_{\text{loc}}$ is a weak equivalence.*

⁽¹⁾(induces an isomorphism on homotopy groups)

2. The sequence of representing spaces $|h\mathcal{V}| \rightarrow |h\mathcal{W}| \rightarrow |h\mathcal{W}_{\text{loc}}|$ is a homotopy fibre sequence⁽²⁾ of infinite loop spaces.
3. There is a homotopy equivalence $|h\mathcal{V}| \simeq \Omega^\infty(\mathbb{C}\mathbf{P}_1^\infty)$.

Let F denote the homotopy fibre of $|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|$, then it follows formally from the homotopy invariance of the homotopy fibre construction that there is a homotopy equivalence $F \xrightarrow{\simeq} |h\mathcal{V}|$. Standard methods of homotopy theory imply that the generalized Mumford conjecture follows from:

THEOREM 1.5. — *There exists a morphism $\mathbb{Z} \times B\Gamma_\infty \rightarrow F$ which induces an isomorphism in homology with integral coefficients.*

The proof of this theorem involves replacing the singularities inherent in \mathcal{W} by ones in standard form and then stratifying by critical sheets; after stratification, the concordance relation is imposed by a homotopical gluing construction, the homotopy colimit over a suitable category. The proof of the theorem relies on foundational results from homotopy theory together with the homological stability results of Harer; in particular, the proof uses closed surfaces with boundary.

1.2. Approximations

Much of the material of [12] is developed for bundles of manifolds of arbitrary dimension, d , and with a general notion of orientation, the Θ -orientation. For the presentation of this text, the general notion of orientation has been suppressed and the integer d is usually taken to be two.

To avoid set-theoretic difficulties, [12] uses the notion of graphic morphisms with respect to a fixed set in the definitions of the sheaves which are considered; moreover set-theoretic caveats are required in various proofs. All such details have been suppressed in this text.

2. MAPPING CLASS GROUPS

2.1. Orientation-preserving diffeomorphisms

Let F be a smooth, compact, oriented surface with boundary ∂F , then F is classified, up to diffeomorphism, by its genus g and the number b of boundary components; write $F_{g,b}$ for a representative of the diffeomorphism class.

The topological group of orientation-preserving diffeomorphisms of F which fix the boundary is written $\text{Diff}^\circ(F; \partial F)$ and $\text{Diff}_e^\circ(F; \partial F)$ denotes the connected component which contains the identity, so that there is a canonical monomorphism of topological groups, $\text{Diff}_e^\circ(F; \partial F) \rightarrow \text{Diff}^\circ(F; \partial F)$.

⁽²⁾A sequence of pointed spaces $F \rightarrow E \rightarrow B$ is a homotopy fibre sequence if F is weakly equivalent to the homotopy fibre of $E \rightarrow B$. The homotopy fibre can be defined explicitly as the fibre product $E \times_B PB$, where $PB \rightarrow B$ is the path space fibration over B .

DEFINITION 2.1. — For g, b non-negative integers, the mapping class group $\Gamma_{g,b}$ is the discrete group of path components, $\Gamma_{g,b} := \pi_0(\text{Diff}^\circ(F_{g,b}; \partial F_{g,b}))$.

Earle and Eells [2] proved that the topological group $\text{Diff}^\circ_e(F; \partial F)$ is contractible, for F a smooth, compact, oriented surface of genus $g \geq 2$.

COROLLARY 2.2. — For $g \geq 2$ an integer, there is a homotopy equivalence $B\Gamma_{g,b} \simeq B\text{Diff}^\circ(F_{g,b}; \partial F_{g,b})$. In particular, the classifying space $B\Gamma_{g,b}$ classifies isomorphism classes of oriented $F_{g,b}$ -bundles.

There is a model for the classifying space $B\Gamma_{g,b}$ constructed from Teichmüller space, for strictly positive b . Let $\mathcal{H}(F)$ denote the space of hyperbolic metrics on the surface F with geodesic boundary such that each boundary circle has unit length. The hyperbolic model for the moduli space of Riemann surfaces of topological type F is given by

$$\mathcal{M}(F) := \mathcal{H}(F)/\text{Diff}^\circ(F; \partial F).$$

Teichmüller space is defined as the quotient $\mathcal{T}(F) := \mathcal{H}(F)/\text{Diff}^\circ_e(F; \partial F)$.

THEOREM 2.3 ([2, 3]). — Let $F := F_{g,b}$ be a smooth, compact, oriented surface of genus $g > 1$, with b boundary components. The following statements hold.

1. The space $\mathcal{H}(F)$ is contractible.
2. The space $\mathcal{T}(F)$ is contractible and homeomorphic to $\mathbb{R}^{6g-6+2b}$.
3. If $b > 0$, the action of $\Gamma_{g,b}$ on Teichmüller space $\mathcal{T}(F)$ is free and $B\Gamma_{g,b} \simeq \mathcal{M}(F)$.
4. If $b = 0$, the action of Γ_g on Teichmüller space $\mathcal{T}(F_{g,0})$ has finite isotropy groups, hence there is a rational homotopy equivalence $B\Gamma_g \simeq_{\mathbb{Q}} \mathcal{M}(F_{g,0})$.

In particular, the above establishes the relation between the moduli space of Riemann surfaces $\mathcal{M}(F)$ and the mapping class group.

2.2. Stabilization

There are two basic gluing constructions which allow stabilization. Recall that $F_{1,2}$ is diffeomorphic to a torus with the interiors of two disjoint disks removed and $F_{0,1}$ is diffeomorphic to a disk; gluings of smooth manifolds provide concatenation diffeomorphisms of oriented surfaces $F_{g+1,b} \cong F_{g,b} \cup_{S^1} F_{1,2}$ and $F_{g,b} \cong F_{g,b+1} \cup_{S^1} F_{0,1}$.

There are induced natural morphisms of groups $\Gamma_{g,b} \rightarrow \Gamma_{g+1,b}$, $\Gamma_{g,b+1} \rightarrow \Gamma_{g,b}$ which induce morphisms between the integral cohomologies of the respective classifying spaces. The stability theorems of Harer and Ivanov imply that these are isomorphisms in a stable range.

THEOREM 2.4 ([7, 9]). — The natural morphisms induce isomorphisms in cohomology $H^*(B\Gamma_{g,b}) \cong H^*(B\Gamma_{g+1,b}) \cong H^*(B\Gamma_{g+1,b+1})$, for $* < g/2 - 1$.

Let $\Gamma_{\infty,b}$ denote the colimit of the direct system of groups $\Gamma_{*,b}$.

COROLLARY 2.5. — The group $H^*(B\Gamma_{\infty,b})$ is isomorphic to the inverse limit of the system $H^*(B\Gamma_{*,b})$ and is independent of b .

2.3. Infinite loop space structure

The classifying space $B\Gamma_{g,b}$ has perfect fundamental group for $g > 1$ (see [6] for the case $g \geq 2$). Hence, the Quillen plus construction $B\Gamma_{g,b}^+$ is defined, with trivial fundamental group, and there is a canonical morphism $B\Gamma_{g,b} \rightarrow B\Gamma_{g,b}^+$ which induces an isomorphism in homology.

The following theorem motivated the generalized Mumford conjecture [11].

THEOREM 2.6 ([21]). — *The space $\mathbb{Z} \times B\Gamma_{\infty,b}^+$ has the structure of an infinite loop space⁽³⁾.*

3. GENERALIZED BUNDLE THEORY

3.1. Submersions, fibrewise tangent bundles and jet bundles

A smooth map $\pi : E \rightarrow X$ between smooth manifolds is a submersion if the morphism of tangent bundles $TE \rightarrow TX$ is surjective on fibres. The vertical tangent bundle $T^\pi E$ of the submersion is a vector bundle on E and there is a short exact sequence $T^\pi E \rightarrow TE \rightarrow \pi^*TX$ of vector bundles on E .

The fibres of a smooth submersion are smooth manifolds of codimension equal to the dimension of the base. Ehresmann's fibration lemma (see [1] for example) states:

THEOREM 3.1. — *Let $\pi : E \rightarrow X$ be a smooth submersion which is proper, then π is a smooth fibre bundle of manifolds.*

The following submersion theorem of Phillips' for open manifolds is used.

THEOREM 3.2 ([18, Theorem A]). — *Let M, W be smooth manifolds such that M is open. If there exists a smooth surjection of tangent bundles $TM \twoheadrightarrow TX$, then there exists a smooth submersion $M \rightarrow X$. Moreover, any two smooth submersions with differentials homotopic through vector bundle surjections are homotopic through submersions.*

For $\pi : E \rightarrow X$ a smooth submersion, let $J_\pi^k(E, \mathbb{R})$ denote the fibrewise k -jet bundle; this is a smooth vector bundle on E which is a quotient of the k -jet bundle $J^k(E, \mathbb{R})$ of which the fibre at z identifies with $J^k(E_{\pi(z)}, \mathbb{R})$, the k -jet bundle of the fibre.

The definition implies that, for k a non-negative integer, there exists a natural surjection $J_\pi^{k+1}(E, \mathbb{R}) \rightarrow J_\pi^k(E, \mathbb{R})$ of vector bundles on E . The bundle $J_\pi^0(E, \mathbb{R})$ is canonically isomorphic to the trivial bundle of rank one on E , which implies the first statement of the following Lemma.

⁽³⁾A pointed topological space $(X, *)$ has an infinite loop space structure if there exist a sequence of pointed spaces $(X_n, *)$, and weak homotopy equivalences $X_n \simeq \Omega X_{n+1}$, for $n \geq 0$, such that $(X, *) = (X_0, *)$. (For technical reasons, suppose all the spaces are CW complexes).

⁽³⁾(for the purposes of this exposition, suppose without boundary)

LEMMA 3.3. — *Let $\pi : E \rightarrow X$ be a smooth submersion and let k be a non-negative integer, then the following statements hold.*

1. *A smooth section \hat{f} of $J_\pi^k(E, \mathbb{R})$ induces a smooth function $f : E \rightarrow \mathbb{R}$.*
2. *A smooth morphism $f : E \rightarrow \mathbb{R}$ induces a smooth section $j_\pi^k(f)$ of the bundle $J_\pi^k(E, \mathbb{R})$.*

There is a canonical identification of the fibrewise 1-jet bundle $J_\pi^1(E, \mathbb{R})$ with $J^0(E, \mathbb{R}) \oplus T^\pi E^*$, where $T^\pi E^*$ is the fibrewise cotangent bundle. In particular, there is a canonical surjection $J_\pi^2(E, \mathbb{R}) \rightarrow T^\pi E^*$, corresponding to the linear part of the fibrewise jet bundle. A section of $J_\pi^2(E, \mathbb{R})$ with vanishing linear part has well-defined quadratic part, which corresponds to a quadratic form on $T^\pi E$, by choice of connection.

DEFINITION 3.4. — *Let \hat{f} be a smooth section of the fibrewise 2-jet bundle $J_\pi^2(E, \mathbb{R})$, where $\pi : E \rightarrow X$ is a smooth submersion.*

1. *The section \hat{f} is fibrewise non-singular if the linear part is a non-vanishing section of $T^\pi E^*$.*
2. *The section \hat{f} is fibrewise Morse if it has non-degenerate quadratic part whenever the linear part vanishes.*

The submanifold of singular jets in $J_\pi^2(E, \mathbb{R})$ is written $\Sigma_\pi(E, \mathbb{R})$.

DEFINITION 3.5. — *Let $\pi : E \rightarrow X$ be a smooth submersion and let \hat{f} be a smooth section of the fibrewise 2-jet bundle. The fibrewise singularity set $\Sigma(\pi, \hat{f}) \subset E$ is the inverse image $\hat{f}^{-1}\Sigma_\pi(E, \mathbb{R})$ of the submanifold of singular jets.*

If \hat{f} identifies with the fibrewise 2-jet prolongation $j_\pi^2(f)$ of a smooth function f , the fibrewise singularity set $\Sigma(\pi, \hat{f})$ is written $\Sigma(\pi, f)$.

A smooth section of the fibrewise 2-jet bundle which is fibrewise Morse is equivalent to a smooth section of the fibrewise cotangent bundle which is transverse to the zero section; this implies the following result.

LEMMA 3.6 ([12]). — *Let $\pi : E \rightarrow X$ be a smooth submersion and let \hat{f} be a smooth section of the fibrewise 2-jet bundle, which is fibrewise Morse. Then $\Sigma(\pi, \hat{f})$ is a smooth submanifold of E and the morphism π restricts to a local diffeomorphism⁽⁴⁾ $\pi|_{\Sigma(\pi, \hat{f})} : \Sigma(\pi, \hat{f}) \rightarrow X$.*

⁽⁴⁾or étale map

3.2. The sheaves $h\mathcal{V}$, $h\mathcal{W}$, $h\mathcal{W}_{\text{loc}}$

Let \mathfrak{X} denote the category of smooth manifolds, without boundary and with a countable basis; all sheaves considered in this section are sheaves of sets on \mathfrak{X} .

The lower row of the diagram of sheaves is defined in terms of analytic data.

DEFINITION 3.7. —

1. Let $h\mathcal{W}$ denote the sheaf which has sections over X the set of pairs (π, \hat{f}) , where $\pi : E \rightarrow X$ is a smooth submersion of fibre dimension three, with oriented fibrewise tangent bundle. The morphism \hat{f} is a smooth section of the fibrewise 2-jet bundle $J_\pi^2(E, \mathbb{R})$ such that the following conditions are satisfied:
 - (a) the morphism $(\pi, f) : E \rightarrow X \times \mathbb{R}$ is proper, where f denotes the underlying smooth morphism of \hat{f} ;
 - (b) the section \hat{f} is fibrewise Morse.
2. Let $h\mathcal{V}$ denote the subsheaf of $h\mathcal{W}$ with sections over X given by pairs (π, \hat{f}) for which the section \hat{f} is fibrewise non-singular.

There is a variant of the sheaf $h\mathcal{W}$, in which the properness hypothesis is weakened.

DEFINITION 3.8. — Let $h\mathcal{W}_{\text{loc}}$ denote the sheaf which has sections over X the set of pairs (π, \hat{f}) , where $\pi : E \rightarrow X$ is a smooth submersion of fibre dimension three with oriented fibrewise tangent bundle. The morphism \hat{f} is a smooth section of the fibrewise 2-jet bundle $J_\pi^2(E, \mathbb{R})$ such that the following conditions are satisfied:

1. the morphism $(\pi, f) : E \rightarrow X \times \mathbb{R}$ restricts to a proper morphism $\Sigma(\pi, \hat{f}) \rightarrow X \times \mathbb{R}$;
2. the section \hat{f} is fibrewise Morse.

By construction, there are canonical morphisms of sheaves $h\mathcal{V} \rightarrow h\mathcal{W} \rightarrow h\mathcal{W}_{\text{loc}}$.

3.3. The sheaves \mathcal{V} , \mathcal{W} , \mathcal{W}_{loc}

The upper row of diagram (1) is obtained by imposing the condition that the smooth sections are integrable.

DEFINITION 3.9. —

1. Let \mathcal{W} denote the sheaf with sections over X the set of pairs (π, f) such that $\pi : E \rightarrow X$ is a smooth submersion of fibre dimension three, $f : E \rightarrow \mathbb{R}$ is a smooth morphism and $(\pi, j_\pi^2 f)$ belongs to $h\mathcal{W}(X)$.
2. Let \mathcal{V} denote the subsheaf of \mathcal{W} which is defined by the cartesian square of sheaves:

$$\begin{array}{ccc} \mathcal{V} & \longrightarrow & \mathcal{W} \\ \downarrow & & \downarrow \\ h\mathcal{V} & \longrightarrow & h\mathcal{W}. \end{array}$$

3. Let \mathcal{W}_{loc} denote the sheaf with sections over X the set of pairs (π, f) , where $\pi : E \rightarrow X$ is a smooth submersion of fibre dimension three and $f : E \rightarrow \mathbb{R}$ is a smooth morphism such that $(\pi, j_\pi^2 f)$ is an element of $h\mathcal{W}_{\text{loc}}$.

Explicit examples of sections of \mathcal{W} are given by the constructions of Section 6. The definitions yield diagram (1) in which the vertical morphisms are induced by fibrewise 2-jet prolongation.

4. BORDISM

This section addresses the identification of the sequence of representing spaces $|h\mathcal{V}| \rightarrow |h\mathcal{W}| \rightarrow |h\mathcal{W}_{\text{loc}}|$ by using bordism-theoretic arguments⁽⁵⁾.

4.1. Oriented bundles with Morse-like functions

DEFINITION 4.1. — *Let d, n be non-negative integers.*

1. *Let $\mathcal{G}(d, n)$ be the space of oriented d -dimensional subspaces of \mathbb{R}^{d+n} .
The space $\mathcal{G}(d, n)$ is a classifying space for oriented d -dimensional vector bundles with a morphism of the total space to \mathbb{R}^{d+n} which restricts to a linear embedding on each fibre.*
2. *Let $\mathcal{GW}(d, n)$ be the space of triples (V, l, q) , where V is an object of $\mathcal{G}(d, n)$, $l : V \rightarrow \mathbb{R}$ is a linear map and $q : V \rightarrow \mathbb{R}$ is a quadratic form, which is non-degenerate if $l = 0$.
The space $\mathcal{GW}(d, n)$ is a classifying space for oriented d -dimensional vector bundles equipped with the additional structure:

 - (a) *a morphism of the total space of the bundle to \mathbb{R}^{d+n} , which is a linear embedding on each fibre;*
 - (b) *a morphism from the total space to \mathbb{R} which restricts to a Morse-like function on each fibre; namely, on each fibre, the morphism has the form $l+q : V \rightarrow \mathbb{R}$, where l is linear and q is a quadratic form, which is non-degenerate if l is zero.**
3. *Let $\Sigma(d, n)$ be the subspace of $\mathcal{GW}(d, n)$ which is given by triples of the form $(V, 0, q)$, where q is a non-degenerate quadratic form.*
4. *Let $\mathcal{GV}(d, n)$ be the complement $\mathcal{GW}(d, n) \setminus \Sigma(d, n)$.*

There is a diagram of inclusions

$$(2) \quad \mathcal{GV}(d, n) \hookrightarrow \mathcal{GW}(d, n) \hookrightarrow \Sigma(d, n).$$

Remark 4.2. —

1. There is a map $\mathcal{GW}(d, n) \rightarrow \mathcal{G}(d, n)$ which forgets the Morse-like function. In particular, (2) is a diagram of spaces over $\mathcal{G}(d, n)$.
2. There is an inclusion $\mathcal{G}(d, n) \hookrightarrow \mathcal{GV}(d+1, n)$, which is induced by $- \oplus \mathbb{R}$ via $V \mapsto (V \oplus \mathbb{R}, l, 0)$, where l is the projection $V \oplus \mathbb{R} \rightarrow \mathbb{R}$.

⁽⁵⁾A résumé of the classical Pontrjagin-Thom correspondence between bordism and generalized homology theories is given in [10, Chapter 1]; the influential paper of Quillen, [18], provides the treatment of the cohomological theory.

3. A monomorphism $\mathbb{R}^{d+n} \hookrightarrow \mathbb{R}^{d+n+1}$ induces a natural morphism $\mathcal{G}(d, n) \rightarrow \mathcal{G}(d, n+1)$, together with compatible natural morphisms for \mathcal{GV} , \mathcal{GW} , Σ .

There are tautological oriented d -dimensional bundles T_n on $\mathcal{G}(d, n)$ and U_n on $\mathcal{GW}(d, n)$. Moreover, there is a canonical monomorphism $U_n \hookrightarrow \mathcal{GW}(d, n) \times \mathbb{R}^{d+n}$ into a trivial bundle, with complement U_n^\perp of dimension n ; the analogous statement holds for T_n .

LEMMA 4.3. —

1. The injection $\mathcal{G}(d, n) \rightarrow \mathcal{GV}(d+1, n)$ is covered by a fibrewise isomorphism of vector bundles $T_n^\perp \rightarrow U_n^\perp|_{\mathcal{GV}(d+1, n)}$.
2. The normal bundle of the embedding $\Sigma(d, n) \hookrightarrow \mathcal{GW}(d, n)$ is isomorphic to the dual bundle $U_n^*|_{\Sigma(d, n)}$.
3. The non-degenerate quadratic form, q , induces an isomorphism between $U_n^*|_{\Sigma(d, n)}$ and $U_n|_{\Sigma(d, n)}$.

4.2. Thom space constructions

The Thom space $\text{Th}(\xi)$ of a vector bundle ξ is the quotient space $\text{Th}(\xi) := D(\xi)/S(\xi)$, where the pair $(D(\xi), S(\xi))$ corresponds to the disc and sphere bundles associated to ξ .⁽⁶⁾

For $M \hookrightarrow \mathbb{R}^N$ an embedding with normal bundle ν such that the total space $E(\nu)$ embeds as a tubular neighbourhood, collapsing the complement of the tubular neighbourhood induces the Pontrjagin-Thom map $S^N \rightarrow \text{Th}(\nu)$. This construction generalizes to give the following.

LEMMA 4.4. — *Let $Y \hookrightarrow X$ be an immersion of smooth manifolds of codimension d , with normal bundle ν , and let ξ be a vector bundle on X . Then the sequence of spaces $\text{Th}(\xi|(X - Y)) \rightarrow \text{Th}(\xi) \rightarrow \text{Th}(\xi|Y \oplus \nu)$ is a homotopy cofibre sequence⁽⁷⁾.*

This applies to the embedding $\Sigma(3, n) \hookrightarrow \mathcal{GW}(3, n)$ and the vector bundle U_n^\perp on $\mathcal{GW}(3, n)$; together with Lemma 4.3, this implies:

LEMMA 4.5. — *For n a non-negative integer, there is a homotopy cofibre sequence*

$$\text{Th}(U_n^\perp|\mathcal{GV}(3, n)) \rightarrow \text{Th}(U_n^\perp|\mathcal{GW}(3, n)) \rightarrow \text{Th}(U_n^\perp \oplus U_n^*|\Sigma(3, n)).$$

The Thom space $\text{Th}(\xi \oplus \theta^1)$ is homeomorphic to the suspension $\Sigma\text{Th}(\xi)$, where θ denotes the trivial bundle of rank one. In particular, a morphism of vector bundles

⁽⁶⁾If the base space is compact, the Thom space is homeomorphic to the one point compactification of the total space.

⁽⁷⁾A sequence $A \rightarrow B \rightarrow C$ of pointed topological spaces is a homotopy cofibre sequence if C is homotopy equivalent to the mapping cone of $A \rightarrow B$.

$\xi \oplus \theta^1 \rightarrow \zeta$ induces a morphism $\mathrm{Th}(\xi) \rightarrow \Omega\mathrm{Th}(\zeta)$, by adjunction. Such morphisms induce the direct systems which define the infinite loop spaces below⁽⁸⁾:

$$\begin{aligned} \Omega^\infty \mathbf{h}\mathcal{W} &:= \operatorname{colim}_n \Omega^{2+n} \mathrm{Th}(U_n^\perp |_{\mathcal{G}\mathcal{W}(3,n)}); \\ \Omega^\infty \mathbf{h}\mathcal{V} &:= \operatorname{colim}_n \Omega^{2+n} \mathrm{Th}(U_n^\perp |_{\mathcal{G}\mathcal{V}(3,n)}); \\ \Omega^\infty \mathbf{h}\mathcal{W}_{\mathrm{loc}} &:= \operatorname{colim}_n \Omega^{2+n} \mathrm{Th}(U_n^\perp \oplus U_n^* |_{\Sigma(3,n)}). \end{aligned}$$

Lemma 4.5 implies the following result, which corresponds to the fact that the infinite loop space functor Ω^∞ sends stable homotopy cofibre sequences to homotopy fibre sequences of spaces.

PROPOSITION 4.6. — *There is a homotopy fibre sequence of infinite loop spaces*

$$\Omega^\infty \mathbf{h}\mathcal{V} \rightarrow \Omega^\infty \mathbf{h}\mathcal{W} \rightarrow \Omega^\infty \mathbf{h}\mathcal{W}_{\mathrm{loc}}.$$

4.3. The morphism $\tau : |h\mathcal{W}| \rightarrow \Omega^\infty \mathbf{h}\mathcal{W}$

The construction of a morphism $\tau : |h\mathcal{W}| \rightarrow \Omega^\infty \mathbf{h}\mathcal{W}$ uses auxiliary sheaves $h\mathcal{W}^{(r)}$ and $\mathcal{Z}^{(r)}$ which are defined as follows.

DEFINITION 4.7. — *For X a smooth closed manifold and r a non-negative integer:*

1. *let $h\mathcal{W}^{(r)}(X)$ denote the set of sections of $h\mathcal{W}(X)$ with the additional structure: a smooth embedding $\omega : E \rightarrow X \times \mathbb{R} \times \mathbb{R}^{2+r}$ over $X \times \mathbb{R}$ and a vertical tubular neighbourhood N ;*
2. *let $\mathcal{Z}^{(r)}(X)$ denote the set of maps $X \times \mathbb{R} \rightarrow \Omega^{2+r} \mathrm{Th}(U_r^\perp)$.*

The Pontrjagin-Thom construction establishes the following result.

LEMMA 4.8. — *There are morphisms of sheaves $h\mathcal{W} \longleftarrow h\mathcal{W}^{(r)} \longrightarrow \mathcal{Z}^{(r)}$.*

The sheaves $h\mathcal{W}^{(r)}$ and $\mathcal{Z}^{(r)}$ form direct systems as r varies, hence there are direct systems of the representing spaces $|h\mathcal{W}^{(r)}|$, $|\mathcal{Z}^{(r)}|$. The first statement of the following Lemma⁽⁹⁾ is a consequence of the Whitney embedding theorem.

LEMMA 4.9. — *There are weak equivalences:*

1. $\operatorname{hocolim}_r |h\mathcal{W}^{(r)}| \xrightarrow{\cong} |h\mathcal{W}|$;
2. $\operatorname{hocolim}_r |\mathcal{Z}^{(r)}| \xrightarrow{\cong} \Omega^\infty \mathbf{h}\mathcal{W}$.

The Lemma induces a morphism $\tau : |h\mathcal{W}| \rightarrow \Omega^\infty \mathbf{h}\mathcal{W}$ in the homotopy category of pointed topological spaces. This restricts to a morphism $\tau_{\mathcal{V}} : |h\mathcal{V}| \rightarrow \Omega^\infty \mathbf{h}\mathcal{V}$ and similar constructions define a morphism $\tau_{\mathrm{loc}} : |h\mathcal{W}_{\mathrm{loc}}| \rightarrow \Omega^\infty \mathbf{h}\mathcal{W}_{\mathrm{loc}}$ so that the following statement holds:

⁽⁸⁾Cf. the construction of Thom spectra, [20].

⁽⁹⁾(in which $\operatorname{hocolim}$ denotes the homotopy colimit of the direct system, which is the derived, homotopy-theoretic version of the direct limit)

LEMMA 4.10. — *There is a homotopy commutative diagram*

$$\begin{array}{ccccc}
 |h\mathcal{V}| & \longrightarrow & |h\mathcal{W}| & \longrightarrow & |h\mathcal{W}_{\text{loc}}| \\
 \tau_{\mathcal{V}} \downarrow & & \tau \downarrow & & \downarrow \tau_{\text{loc}} \\
 \Omega^{\infty}\mathbf{h}\mathcal{V} & \longrightarrow & \Omega^{\infty}\mathbf{h}\mathcal{W} & \longrightarrow & \Omega^{\infty}\mathbf{h}\mathcal{W}_{\text{loc}}.
 \end{array}$$

4.4. Bordism approach to τ

There are bordism descriptions for the cohomology theories represented by the spaces $\Omega^{\infty}\mathbf{h}\mathcal{W}$, $\Omega^{\infty}\mathbf{h}\mathcal{V}$ and $\Omega^{\infty}\mathbf{h}\mathcal{W}_{\text{loc}}$.

For example, consider the case $\Omega^{\infty}\mathbf{h}\mathcal{W}$; for X a smooth manifold, the set of homotopy classes $[X, \Omega^{\infty}\mathbf{h}\mathcal{W}]$ identifies with the set of bordism classes of triples (M, g, \hat{g}) , where M is a closed, smooth manifold of dimension $2 + \dim X$ and g is a morphism, $M \rightarrow X \times \mathcal{GW}(3, n)$, such that the projection to X induces a proper morphism, $M \rightarrow X$. The morphism \hat{g} is a pull-back of vector bundles of the form

$$\begin{array}{ccc}
 TM \times \mathbb{R} \times \mathbb{R}^j & \xrightarrow{\hat{g}} & TX \times U_{\infty} \times \mathbb{R}^j \\
 \downarrow & & \downarrow \\
 M & \xrightarrow{g} & X \times \mathcal{GW}(3, n),
 \end{array}$$

for some non-negative integer j . An obstruction theory argument (see [12]) shows that the integer j can be taken to be zero.

Let (π, \hat{f}) represent an element of $h\mathcal{W}(X)$ then, up to concordance, one may suppose that $f : E \rightarrow \mathbb{R}$ is transverse to $0 \in \mathbb{R}$, so that $M := f^{-1}(0)$ is a submanifold of dimension $2 + \dim X$ and the induced morphism $\pi : M \rightarrow X$ is proper. The section \hat{f} of the fibrewise 2-jet bundle, restricted to points of M , has the form $\hat{f}_z = l_z + q_z$ satisfying the non-degeneracy hypothesis. There is a classifying map $M \rightarrow \mathcal{GW}(3, \infty)$ and it is straightforward to verify that this defines an element of $[X, \Omega^{\infty}\mathbf{h}\mathcal{W}]$.

Conversely, given a triple (M, g, \hat{g}) representing a bordism class, so that $g : M \rightarrow X \times \mathcal{GW}(3, \infty)$, set $E := M \times \mathbb{R}$ and let π_E denote the composite morphism $E \rightarrow M \rightarrow X$. Phillips' submersion theorem (Theorem 3.2) implies that this gives rise to an element of $h\mathcal{W}(X)$.

At the level of concordance classes, this implies the following theorem.

THEOREM 4.11. — *The morphism $\tau : |h\mathcal{W}| \rightarrow \Omega^{\infty}\mathbf{h}\mathcal{W}$ is a homotopy equivalence.*

Similar considerations apply to show that $\tau_{\mathcal{V}} : |h\mathcal{V}| \rightarrow \Omega^{\infty}\mathbf{h}\mathcal{V}$ and $\tau_{\text{loc}} : |h\mathcal{W}_{\text{loc}}| \rightarrow \Omega^{\infty}\mathbf{h}\mathcal{W}_{\text{loc}}$ are homotopy equivalences.

COROLLARY 4.12. — *The sequence of spaces, $|h\mathcal{V}| \rightarrow |h\mathcal{W}| \rightarrow |h\mathcal{W}_{\text{loc}}|$, is a homotopy fibre sequence of infinite loop spaces.*

4.5. The morphism $j_\pi^2 : \mathcal{W}_{\text{loc}} \rightarrow h\mathcal{W}_{\text{loc}}$

The following result is proved in [12].

THEOREM 4.13. — *The morphism of sheaves $j_\pi^2 : \mathcal{W}_{\text{loc}} \rightarrow h\mathcal{W}_{\text{loc}}$ induces a weak equivalence between the representing spaces.*

The proof is geometric in nature, using a bordism theoretic description of $\mathcal{W}_{\text{loc}}(X)$, together with an application of Phillips’ submersion theorem, Theorem 3.2.

4.6. Identification of the space $\Omega^\infty \mathbf{h}\mathcal{V}$ and the morphism $j_\pi^2 : \mathcal{V} \rightarrow h\mathcal{V}$

The morphism $\mathcal{G}(2, n) \rightarrow \mathcal{GW}(3, n)$ gives rise to a morphism of Thom spaces $\text{Th}(T_n^\perp) \rightarrow \text{Th}(U_n^\perp|_{\mathcal{GW}(3, n)})$. The presence of the quadratic form does not affect the homotopy type of the associated infinite loop space:

LEMMA 4.14. — *There is a weak equivalence $\Omega^\infty \mathbf{h}\mathcal{V} \simeq \text{colim}_n \Omega^{2+n} \text{Th}(T_n^\perp)$.*

A standard argument involving complexification⁽¹⁰⁾ implies the following result.

PROPOSITION 4.15. — *The space $|h\mathcal{V}|$ is equivalent to the infinite loop space, $\Omega^\infty \mathbb{C}\mathbf{P}_{-1}^\infty$.*

The morphism $j_\pi^2 : \mathcal{V} \rightarrow h\mathcal{V}$ induces a map between concordance classes $\mathcal{V}[X] \rightarrow h\mathcal{V}[X]$, which is induced by a map $|\mathcal{V}| \rightarrow |h\mathcal{V}|$ between the representing spaces. The space $|h\mathcal{V}|$ is equivalent to the space $\Omega^\infty \mathbf{h}\mathcal{V}$ (remark following Theorem 4.11) and hence to $\Omega^\infty \mathbb{C}\mathbf{P}_{-1}^\infty$ by Proposition 4.15.

The sheaf \mathcal{V} can be interpreted in terms of surface bundles on $X \times \mathbb{R}$, by Ehresmann’s fibration lemma. Hence the functor $X \mapsto \mathcal{V}[X]$ is represented by the space $\amalg_F B\text{Diff}^\circ(F, \partial F)$, as F ranges over diffeomorphism classes of smooth, closed oriented surfaces (not necessarily connected). The natural transformation $\mathcal{V}[X] \rightarrow h\mathcal{V}[X]$ is represented by a map $\alpha : \amalg_F B\text{Diff}^\circ(F, \partial F) \rightarrow \Omega^\infty \mathbf{h}\mathcal{V} \simeq \Omega^\infty \mathbb{C}\mathbf{P}_{-1}^\infty$, which is related to the morphism α_∞ of the Introduction.

5. APPLICATION OF VASSILIEV’S FIRST MAIN THEOREM

Theorem 1.3 is deduced from Vassiliev’s first main theorem; the proof uses techniques from sheaf homotopy theory.

⁽¹⁰⁾(passage from the structure group SO to U)

5.1. Vassiliev’s first main theorem

DEFINITION 5.1. — *Let $\mathfrak{U} \subset J^2(\mathbb{R}^r, \mathbb{R})$ denote the space of 2-jets represented by smooth functions $f : (\mathbb{R}^r, z) \rightarrow \mathbb{R}$ such that $f(z) = 0$ and $df(z) = 0$ and $\det(d^2f(z)) = 0$.*

The space \mathfrak{U} corresponds to the space of 2-jets with singularities which have critical value zero and which are not Morse.

DEFINITION 5.2. — *For $N^r, \partial N^r$ a smooth, compact manifold with boundary and $\psi : N \rightarrow \mathbb{R}$ a smooth function such that $j^2\psi(z) \notin \mathfrak{U}$ in a neighbourhood of ∂N , define the spaces:*

1. $\Phi((N, \partial N), \psi) := \{f \in C^\infty(N, \mathbb{R}) \mid f \equiv_{\partial N} \psi, j^2f \notin \mathfrak{U}\};$
2. $h\Phi((N, \partial N), \psi) := \{\hat{f} \in \Gamma J^2(N, \mathbb{R}) \mid \hat{f} \equiv_{\partial N} j^2\psi, \hat{f} \notin \mathfrak{U}\}.$

(Here, $\equiv_{\partial N}$ indicates equality in a neighbourhood of ∂N .)

Jet prolongation defines a map $j^2 : \Phi((N, \partial N), \psi) \rightarrow h\Phi((N, \partial N), \psi)$. A special case of Vassiliev’s main theorem reads as follows:

THEOREM 5.3 ([23, 22]). — *The map $j^2 : \Phi((N, \partial N), \psi) \rightarrow h\Phi((N, \partial N), \psi)$ induces an isomorphism in integral homology.*

5.2. Indications on the proof of Theorem 1.3

The Whitehead theorem implies that it is sufficient to show that the morphism $j_\pi^2 : |\mathcal{W}| \rightarrow |h\mathcal{W}|$ induces an isomorphism on integral homology, using the fact that the spaces $|\mathcal{W}|, |h\mathcal{W}|$ are simple⁽¹¹⁾. This fact is deduced from the existence of compatible monoid structures on the spaces, together with the fact that $|h\mathcal{W}|$ has the structure of an infinite loop space, by Theorem 1.4.

The next step is to extend $\mathcal{W}, h\mathcal{W}$ to weakly equivalent sheaves $\mathcal{W}^0, h\mathcal{W}^0$ in which the Morse condition is only imposed in a neighbourhood of the critical value $f^{-1}(0)$. There is a canonical extension of the jet-prolongation morphism to a morphism $j_\pi^2 : \mathcal{W}^0 \rightarrow h\mathcal{W}^0$ and it is sufficient to show that this induces an isomorphism in integral homology.

The sheaves $\mathcal{W}^0, h\mathcal{W}^0$ admit homotopy colimit decompositions, expressed in terms of the functor β of Definition A.13. Theorem 1.3 is deduced by applying Proposition A.17: the proof reduces to showing that the morphism j_π^2 induces a homology equivalence between fibres over the same point. This follows from Vassiliev’s main theorem by identifying the morphism between the fibres explicitly.

⁽¹¹⁾A connected space is simple if it has abelian fundamental group which acts trivially on the higher homotopy groups.

6. ELEMENTARY MORSE SINGULARITIES

The proof of Theorem 1.5 requires an analysis of standard models for the elements of \mathcal{W} in terms of certain multi-elementary Morse singularities. The constructions introduced here are exploited in Section 7.

6.1. Morse vector spaces and vector bundles

DEFINITION 6.1. —

1. A Morse vector space is a pair (V, ρ) , where V is a finite dimensional real vector space equipped with an inner product and $\rho : V \rightarrow V$ is a linear, isometric involution. The involution ρ induces a decomposition $V \cong V^\rho \oplus V^{-\rho}$, $u \mapsto (u_+, u_-)$, where ρ acts trivially on V^ρ and by multiplication by -1 on $V^{-\rho}$.
2. The Morse index of (V, ρ) is the dimension of $V^{-\rho}$.
3. The Morse function of (V, ρ) is the smooth function $f_V : V \rightarrow \mathbb{R}$ given by

$$f_V(u) := \langle u, \rho u \rangle \cong \|u_+\|^2 - \|u_-\|^2.$$

The Morse vector spaces provide a good local model for elementary Morse functions, namely those with a single critical point.

DEFINITION 6.2. — For (V, ρ) a Morse vector space, the saddle of (V, ρ) is the smooth manifold with boundary: $\text{Saddle}(V, \rho) := \{u \in V \mid \|u_+\|^2 \|u_-\|^2 \leq 1\}$.

Example 6.3. — Let (V, ρ) be a Morse vector space of dimension three and of Morse index one; the above construction yields $f_V : \text{Saddle}(V, \rho) \rightarrow \mathbb{R}$. The function f_V is fibrewise singular; there is an isolated critical point in the fibre above $0 \in \mathbb{R}$ and the morphism f_V restricts to a bundle of closed surfaces above $\mathbb{R} \setminus \{0\}$. The fibre above a point of $(0, \infty)$ is a hyperboloid of two sheets and above $(-\infty, 0)$ is a hyperboloid of one sheet; the singular fibre above $\{0\}$ is a cone.

LEMMA 6.4. — There is a smooth embedding of codimension zero:

$$\begin{aligned} \text{Saddle}(V, \rho) \setminus V^\rho &\hookrightarrow D(V^\rho) \times S(V^{-\rho}) \times \mathbb{R} \\ u &\mapsto (\|u_-\|u_+, \|u_-\|^{-1}u_-, f_V(u)) \end{aligned}$$

with complement $0 \times S(V^{-\rho}) \times [0, \infty)$. In particular, there is a diffeomorphism of the boundary of $\text{Saddle}(V, \rho)$ with $S(V^\rho) \times S(V^{-\rho}) \times \mathbb{R}$.

There is an analogous smooth embedding $\text{Saddle}(V, \rho) \setminus V^{-\rho} \hookrightarrow S(V^\rho) \times D(V^{-\rho}) \times \mathbb{R}$ with complement $S(V^\rho) \times 0 \times (-\infty, 0]$.

Example 6.5. — Let (V, ρ) be a Morse vector space as in Example 6.3 and let M be a smooth, oriented, closed surface with boundary, with an embedding $D^2 \times S^0 \hookrightarrow \text{Int}M$ into the interior of M . There is a smooth 3-manifold constructed by gluing:

$$W(M, (V, \rho)) \cong (M^\circ \times \mathbb{R}) \cup_{S^1 \times S^0 \times \mathbb{R}} \text{Saddle}(V, \rho),$$

where M° is obtained by removing the interior of the embedded disks $D^2 \times S^0$, such that f_V extends to a smooth function $\widetilde{f}_V : W(M, (V, \rho)) \rightarrow \mathbb{R}$. This defines an element of $\mathcal{W}(\ast)$.

For example, if $M = S^2$ is the sphere with two disjoint embedded disks then the bundle of surfaces over $\mathbb{R} \setminus \{0\}$ has fibre above $(0, \infty)$ a sphere and, above $(-\infty, 0)$, a torus. The singular fibre above $\{0\}$ is topologically a sphere with two disjoint points identified.

The notion of a Morse vector space extends to that of a Morse vector bundle.

DEFINITION 6.6. — *A Morse vector bundle over X is a triple (V, p, ρ) , where $p : V \rightarrow X$ is a smooth Riemannian vector bundle and $\rho : V \rightarrow V$ is a fibrewise linear isometric involution over X .*

The saddle of a Morse vector bundle (V, p, ρ) is the smooth manifold over X obtained by applying the saddle construction fibrewise.

Examples 6.3 and 6.5 generalize to the parametrized situation and give examples of sections of $\mathcal{W}(X)$ not in $\mathcal{V}(X)$.

6.2. Regularization of elementary Morse singularities

There are two standard ways of regularizing the Morse singularity of the function f_V , by removing the embedded subspace V^ρ (respectively $V^{-\rho}$).

LEMMA 6.7. — *Let (V, ρ) be a Morse vector space. There exists a proper, regular function:*

$$f_V^+ : \text{Saddle}(V, \rho) \setminus V^\rho \rightarrow \mathbb{R}$$

such that f_V^+ agrees with f_V on an open subset of $\text{Saddle}(V, \rho)$ which contains the boundary and on the subset $\{u \in \text{Saddle}(V, \rho) \mid f_V(u) \leq -1\}$.

There exists an analogous construction of a proper, regular function: $f_V^- : \text{Saddle}(V, \rho) \setminus V^{-\rho} \rightarrow \mathbb{R}$.

Remark 6.8. — The hypothesis on the open subset of agreement of f_V^+ is necessary for two reasons: to ensure that smooth gluing is possible (using the neighbourhood of the boundary) and to ensure that there is an explicit form of the restriction of f_V^+ to the submanifold $f_V(u) = -1$.

LEMMA 6.9. — *The functions f_V^+, f_V^- induce diffeomorphisms*

$$\begin{aligned} \text{Saddle}(V, \rho) \setminus V^\rho &\hookrightarrow D(V^\rho) \times S(V^{-\rho}) \times \mathbb{R}, \\ \text{Saddle}(V, \rho) \setminus V^{-\rho} &\hookrightarrow S(V^\rho) \times D(V^{-\rho}) \times \mathbb{R}, \end{aligned}$$

by $u \mapsto (\|u_-\|u_+, \|u_-\|^{-1}u_-, f_V^+(u))$ and $u \mapsto (\|u_+\|^{-1}u_+, \|u_+\|u_-, f_V^-(u))$ respectively.

The following Lemma ensures that elementary Morse functions are modelled by saddles; moreover, there are standard ways to reparametrize Morse functions upon saddles.

LEMMA 6.10 ([12]). — *Let N be a smooth manifold equipped with an elementary Morse function $f : N \rightarrow \mathbb{R}$, with unique critical value 0 ; then there exist a Morse vector space (V, ρ) and a codimension zero embedding $\lambda : \text{Saddle}(V, \rho) \rightarrow N \setminus \partial N$, such that $f\lambda = f_V$.*

DEFINITION 6.11. — *For N, f as above, the positive regularization is the pair $(N_+^{\text{rg}}, f_+^{\text{rg}})$, where $N_+^{\text{rg}} := N \setminus \lambda(V^\rho)$ and f_+^{rg} is given by patching f and f_V^+ . If N is oriented, N_+^{rg} is given the induced orientation.*

The negative regularization $(N_-^{\text{rg}}, f_-^{\text{rg}})$ is defined in the analogous way.

LEMMA 6.12. — *For $(N_+^{\text{rg}}, f_+^{\text{rg}})$ the positive regularization as above, the function f_+^{rg} is smooth and proper.*

6.3. Surgery and the long trace construction

The following surgery construction is related to Example 6.5; there is an evident parametrized version of the construction.

DEFINITION 6.13. — *Let M be a smooth, compact manifold, equipped with a codimension zero embedding, $e : D(V^\rho) \times S(V^{-\rho}) \rightarrow M \setminus \partial M$, for a Morse vector space (V, ρ) with $\dim V = \dim M + 1$.*

The long trace of e is the smooth manifold $\text{trace}(e)$ which is obtained as the pushout of the codimension zero embeddings

$$\begin{array}{ccc} \text{Saddle}(V, \rho) \setminus V^\rho & \xrightarrow{\quad\quad\quad} & \text{Saddle}(V, \rho) \\ \downarrow & & \\ (M \times \mathbb{R}) \setminus e(0 \times S(V^{-\rho})) \times [0, \infty). & & \end{array}$$

The long trace is equipped with the elementary Morse function which is the smooth height function on the complement of $\text{Saddle}(V, \rho)$ and identifies with the function f_V on the copy of $\text{Saddle}(V, \rho)$.

The regularization constructions of the previous section give two ways in which to regularize the function f_V , which correspond to removing the subspaces V^ρ (respectively $V^{-\rho}$) from the embedded copy of $\text{Saddle}(V, \rho)$.

There is a related surgery construction which corresponds to changing the choice of regularization:

DEFINITION 6.14. — *For $e : D(V^\rho) \times_X S(V^{-\rho}) \hookrightarrow M$ a smooth embedding, where $q : M \rightarrow X$ is a smooth bundle of d -manifolds with an orientation of the vertical tangent bundle and (V, ρ) is a Morse vector bundle over X , let $q^b : M^b \rightarrow X$ be the bundle of d -manifolds which is obtained by fibrewise surgery, by removing the interior of $D(V^\rho) \times_X S(V^{-\rho})$ and gluing in $S(V^\rho) \times_X D(V^{-\rho})$.*

Example 6.15. — Let $M = S^2$ be the sphere with an embedding $e : D^2 \times S^0 \hookrightarrow M$; the manifold M^b is diffeomorphic to the torus. The construction can be reversed by inverting the rôle of the Morse index.

7. STRATIFICATION OF THE SHEAVES \mathcal{W} , \mathcal{W}_{loc}

The proof of the generalized Mumford conjecture requires an analysis of the homotopy fibre of the map of representing spaces, $|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|$. This involves the formation of homotopy colimit decompositions of the sheaves \mathcal{W} , \mathcal{W}_T via stratifications (see Corollary 7.4). The strategy of the construction is resumed in the following sequence of results, which require the introduction of certain auxiliary sheaves.

The first step replaces the sheaves \mathcal{W} , \mathcal{W}_{loc} by sheaves \mathcal{L} , \mathcal{L}_{loc} , in which there are standard forms for neighbourhoods of the critical points.

PROPOSITION 7.1. — *There is a commutative diagram of sheaves,*

$$\begin{array}{ccccc} \mathcal{L} & \xrightarrow{\simeq} & \mathcal{W}^\mu & \xrightarrow{\simeq} & \mathcal{W} \\ \downarrow & & \downarrow & & \downarrow \\ \mathcal{L}_{\text{loc}} & \xleftarrow{\simeq} & \mathcal{W}_{\text{loc}}^\mu & \xrightarrow{\simeq} & \mathcal{W}_{\text{loc}}, \end{array}$$

in which the horizontal morphisms are weak equivalences.

There are compatible stratifications of the sheaves \mathcal{L} , \mathcal{L}_{loc} , which give rise to diagrams $\mathcal{L}_T \rightarrow \mathcal{L}$, $\mathcal{L}_{\text{loc},T} \rightarrow \mathcal{L}_{\text{loc}}$, indexed over the small category \mathcal{K}^{op} . These provide homotopy decompositions of \mathcal{L} , \mathcal{L}_{loc} respectively, by the following result, using the techniques of Appendix A.6, where the definition of the homotopy colimit (as a sheaf) of a diagram of sheaves is given.

PROPOSITION 7.2. — *There is a commutative diagram of sheaves*

$$\begin{array}{ccc} \text{hocolim}_{T \in \mathcal{K}} \mathcal{L}_T & \xrightarrow{\simeq} & \mathcal{L} \\ \downarrow & & \downarrow \\ \text{hocolim}_{T \in \mathcal{K}} \mathcal{L}_{\text{loc},T} & \xrightarrow{\simeq} & \mathcal{L}_{\text{loc}}, \end{array}$$

in which the horizontal morphisms are weak equivalences.

Corollary 7.4 is expressed in terms of diagrams of sheaves \mathcal{W}_T , $\mathcal{W}_{\text{loc},T}$, which are indexed over \mathcal{K}^{op} .

PROPOSITION 7.3. — *There are natural commutative diagrams of sheaves, for $T \in \mathcal{K}$:*

$$\begin{array}{ccc} \mathcal{L}_T & \xrightarrow{\simeq} & \mathcal{W}_T \\ \downarrow & & \downarrow \\ \mathcal{L}_{\text{loc},T} & \xrightarrow{\simeq} & \mathcal{W}_{\text{loc},T}, \end{array}$$

in which the horizontal morphisms are weak equivalences.

Propositions 7.1, 7.2, 7.3 and homotopy invariance of homotopy colimits imply the following:

COROLLARY 7.4. — *The homotopy fibre of the morphism $|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|$ is weakly equivalent to the homotopy fibre of the morphism $\text{hocolim}_{T \in \mathcal{K}} |\mathcal{W}_T| \rightarrow \text{hocolim}_{T \in \mathcal{K}} |\mathcal{W}_{\text{loc}, T}|$.*

7.1. The sheaves \mathcal{L} and \mathcal{L}_{loc}

The passage to the sheaves \mathcal{L} , \mathcal{L}_{loc} corresponds to requiring standard forms for neighbourhoods of the critical points arising in \mathcal{W} and \mathcal{W}_{loc} .

DEFINITION 7.5. — *Let \mathcal{L}_{loc} denote the sheaf which has sections over X the set of triples (p, g, V) , where:*

1. p is an étale map $Y \rightarrow X$;
2. g is a smooth function $Y \rightarrow \mathbb{R}$ such that the morphism $(p, g) : Y \rightarrow X \times \mathbb{R}$ is proper;
3. $\omega : V \rightarrow Y$ is a three-dimensional oriented Morse vector bundle.

The above definition implies that $Y \rightarrow X$ is a finite étale covering.

DEFINITION 7.6. — *Let \mathcal{L} denote the sheaf which has sections over X the set of tuples $(p, g, V, \pi, f, \lambda)$ such that*

1. $(p, g, V) \in \mathcal{L}_{\text{loc}}(X)$;
2. $(\pi, f) \in \mathcal{W}(X)$;
3. $\lambda : \text{Saddle}(V, \rho) \rightarrow E$ is a smooth embedding such that $\Sigma(\pi, f) \subset \text{image}(\lambda)$ and λ respects the orientation along the fibrewise singularity set;
4. f is a function such that $f\lambda(u) = f_V(u) + g(\omega(u))$, so that $f\lambda$ has the same fibrewise singularity set as f_V and g corresponds to the critical value function.

Tuples as above satisfy $Y = \Sigma(\pi, f)$ and the vector bundle $V \rightarrow Y$ is isomorphic to the restriction of $T^\pi E$ to $\Sigma(\pi, f)$.

The definitions yield canonical morphisms of sheaves $\mathcal{W} \longleftarrow \mathcal{L} \longrightarrow \mathcal{L}_{\text{loc}}$. The proof of Proposition 7.1 requires the introduction of the auxiliary sheaves \mathcal{W}^μ and $\mathcal{W}_{\text{loc}}^\mu$; the relative surjectivity criterion of Proposition A.7 is used to show that the morphisms $\mathcal{L} \rightarrow \mathcal{W}^\mu$ and $\mathcal{W}_{\text{loc}}^\mu \rightarrow \mathcal{L}_{\text{loc}}$ are weak equivalences. The remaining comparisons are straightforward.

7.2. The category \mathcal{K}

The Morse index of a Morse vector space, (V, ρ) , is an integer in the interval $[0, \dim V]$. When considering the regularization of elementary Morse singularities, one can attribute an integer in $\{1, -1\}$ to a critical point according to the way in which the function is regularized, corresponding to sending the critical value either to ∞ or $-\infty$. This motivates the following definition:

DEFINITION 7.7. — *Let \mathcal{K} denote the category with objects finite sets over $[\mathbf{3}] := \{0, 1, 2, 3\}$ and with morphisms injective maps over $[\mathbf{3}]$, $\iota : S \hookrightarrow T$, together with a function $\epsilon : T \setminus \iota S \rightarrow \{\pm 1\}$. The composition of morphisms is given in the obvious way.*

A morphism in \mathcal{K} is the composite of ‘elementary’ morphisms of the form $S \hookrightarrow S \cup \{a\}$, monomorphism of sets over $[\mathbf{3}]$, together with the value $\epsilon(a) \in \{\pm 1\}$.

7.3. Stratifying - the sheaves \mathcal{L}_T and $\mathcal{L}_{\text{loc},T}$

The sheaves \mathcal{L} and \mathcal{L}_{loc} are stratified by taking into account the Morse index and the possible regularizations of the fibrewise singularity sets. The following definition is justified by the observation that, for any element (p, g, V) of $\mathcal{L}_{\text{loc}}(X)$, the function g is locally either bounded above or bounded below (the statement is made precise in [12]); this follows from the properness of the morphism (p, g) .

DEFINITION 7.8. — *For T an object of \mathcal{K} , let $\mathcal{L}_{\text{loc},T}$ denote the sheaf with sections over X the set of tuples (p, g, V, δ, h) where:*

1. $(p, g, V) \in \mathcal{L}_{\text{loc}}(X)$;
2. $\delta : Y \rightarrow \{-1, 0, +1\}$ is a continuous function;
3. $h : T \times X \rightarrow \delta^{-1}(0) \subset Y$ is a diffeomorphism over $[\mathbf{3}] \times X$, where the structure morphism $Y \rightarrow [\mathbf{3}]$ is induced by the Morse index;

such that, for each $x \in X$, there exists a neighbourhood U of x such that g is bounded below on $p^{-1}(U) \cap \delta^{-1}(+1)$ and bounded above on $p^{-1}(U) \cap \delta^{-1}(-1)$.

Remark 7.9. —

1. The function δ is locally constant; the existence of h implies that δ is constant on each sheet of $\delta^{-1}(0)$.
2. The data h, δ are introduced since a choice of regularization is allowed only where the function g is locally bounded.

LEMMA 7.10. — *The association $T \mapsto \mathcal{L}_{\text{loc},T}$ is contravariantly functorial, where for a generating morphism $S \hookrightarrow S \cup \{a\}$, $(p, g, V, \delta, h) \in \mathcal{L}_{\text{loc},S \cup \{a\}}(X)$ has image (p, g, V, δ', h') , where the subset $(\delta')^{-1}(0)$ corresponds to $S \times X \hookrightarrow (S \cup \{a\})(X)$ via h , the morphism h' is the induced morphism and the value of δ' on $\{a\} \times X$ is $\epsilon(a)$.*

DEFINITION 7.11. — *For T an object of \mathcal{K} ,*

1. let \mathcal{L}'_T denote the pullback of the diagram $\mathcal{L} \rightarrow \mathcal{L}_{\text{loc}} \leftarrow \mathcal{L}_{\text{loc},T}$;
2. let \mathcal{L}_T denote the subsheaf of \mathcal{L}'_T which is given by elements for which g is identically zero on $\delta^{-1}(0)$.

These constructions yield the diagram of morphisms of sheaves

$$\begin{array}{ccc} \mathcal{L}_T & \longrightarrow & \mathcal{L} \\ \downarrow & & \downarrow \\ \mathcal{L}_{\text{loc},T} & \longrightarrow & \mathcal{L}_{\text{loc}} \end{array}$$

natural in T .

Proposition 7.2 asserts that the morphisms $\text{hocolim}_{T \in \mathcal{K}} \mathcal{L}_T \rightarrow \mathcal{L}$ and $\text{hocolim}_{T \in \mathcal{K}} \mathcal{L}_{\text{loc},T} \rightarrow \mathcal{L}_{\text{loc}}$ are weak equivalences. The sheaf theoretic model for the

homotopy colimit introduced in Definition A.13 is used, together with certain auxiliary sheaves; the proofs appeal to the relative surjectivity criterion of Proposition A.7. The reader is referred to [12] for the details.

7.4. The sheaves \mathcal{W}_T , $\mathcal{W}_{\text{loc},T}$

The sheaves \mathcal{W}_T , $\mathcal{W}_{\text{loc},T}$ introduced in this section play a fundamental rôle in the proof of the generalized Mumford conjecture.

DEFINITION 7.12. — *For T an object of \mathcal{K} , let $\mathcal{W}_{\text{loc},T}$ denote the sheaf with sections over X the set of three-dimensional smooth, oriented Riemannian vector bundles, $\omega : V \rightarrow T \times X$, equipped with a fibrewise linear isometric involution $\rho : V \rightarrow V$ such that the fibres above $t \times X \subset T \times X$ have Morse index given by the image of t in [3]. The sheaf structure is given by the pull-back of vector bundles.*

LEMMA 7.13. — *The association $T \mapsto \mathcal{W}_{\text{loc},T}$ is contravariantly functorial in \mathcal{K} .*

DEFINITION 7.14. — *For T an object of \mathcal{K} , let \mathcal{W}_T denote the sheaf with sections over X the set of data $((V, \rho), q : M \rightarrow X, e)$ satisfying*

1. (V, ρ) is an element of $\mathcal{W}_{\text{loc},T}$;
2. $q : M \rightarrow X$ is a smooth bundle of closed surfaces, with oriented vertical tangent bundle;
3. $e : D(V^\rho) \times_{T \times X} S(V^{-\rho}) \rightarrow M$ is a smooth embedding over X , respecting the vertical orientations.

The functoriality of \mathcal{W}_- with respect to \mathcal{K} corresponds to the alternative choices of regularizations. This uses the surgery construction of Definition 6.14.

LEMMA 7.15. — *The association $T \mapsto \mathcal{W}_T$ is contravariantly functorial in \mathcal{K} , with respect to the following structure for elementary morphisms in \mathcal{K} of the form $S \hookrightarrow S \cup \{a\}$. An element $((V, \rho), q, e)$ maps to $((V', \rho'), q', e')$, where*

1. $(M', q') := (M^b, q^b)$ if $\epsilon(a) = -1$, and $(M', q') = (M, q)$ otherwise;
2. (V', ρ') is given by functoriality of $\mathcal{W}_{\text{loc},T}$;
3. e' is the induced embedding.

By construction, there is a forgetful map $\mathcal{W}_T \rightarrow \mathcal{W}_{\text{loc},T}$, which is natural in T .

7.5. Relating $\mathcal{L}_{\text{loc},T}$ and $\mathcal{W}_{\text{loc},T}$

There is a morphism of sheaves $\mathcal{L}_{\text{loc},T} \rightarrow \mathcal{W}_{\text{loc},T}$ which is given by $(p, g, V, \delta, h) \mapsto h^*(V)$. Similarly, there is a map $\mathcal{W}_{\text{loc},T} \rightarrow \mathcal{L}_{\text{loc},T}$ which is induced by taking δ and g to be identically zero.

The relative surjectivity criterion of Proposition A.7 is used to show that the morphism $\mathcal{W}_{\text{loc},T} \rightarrow \mathcal{L}_{\text{loc},T}$ is a weak equivalence; it is straightforward to deduce that $\mathcal{L}_{\text{loc},T} \rightarrow \mathcal{W}_{\text{loc},T}$ is a weak equivalence.

7.6. Regularization and the natural transformation $\mathcal{L}_T \rightarrow \mathcal{W}_T$

A natural morphism of sheaves $\mathcal{L}_T \rightarrow \mathcal{W}_T$ is constructed by using the regularization construction below.

DEFINITION 7.16. — For $(\pi, f, p, g, V, \delta, h, \lambda)$ representing a section of $\mathcal{L}_T(X)$, write V_+ for $V|_{\delta^{-1}(1)}$, V_- for $V|_{\delta^{-1}(-1)}$ and V_0 for $V|_{\delta^{-1}(0)}$.

Let $(E^{\text{rg}}, \pi^{\text{rg}}, f^{\text{rg}})$ denote the structure:

1. $E^{\text{rg}} := E \setminus \lambda(V_+^\rho \cup V_0^\rho \cup V_-^{-\rho})$;
2. $\pi^{\text{rg}} := \mu|_{E^{\text{rg}}}$;
3. $f^{\text{rg}} : E^{\text{rg}} \rightarrow \mathbb{R}$ the extension of $f|_{E \setminus \text{Im}(\lambda)}$ by f_V^+ on the image of $V_+ \cup V_0$ and f_V^- on the image of V_- .

LEMMA 7.17. — Let $(\pi, f, p, g, V, \delta, h, \lambda)$ represent a section of $\mathcal{L}_T(X)$, then E^{rg} is an open subset of E , $\pi^{\text{rg}} : E^{\text{rg}} \rightarrow X$ is a smooth submersion and f^{rg} is regular on each fibre of π^{rg} . Moreover, $(\pi^{\text{rg}}, f^{\text{rg}}) : E^{\text{rg}} \rightarrow X \times \mathbb{R}$ is a smooth, proper submersion.

Ehresmann’s fibration lemma implies that the map $(\pi^{\text{rg}}, f^{\text{rg}}) : E^{\text{rg}} \rightarrow X \times \mathbb{R}$ is a smooth bundle of closed, oriented surfaces.

DEFINITION 7.18. — For $(\pi, f, p, g, V, \delta, h, \lambda)$ representing a section of $\mathcal{L}_T(X)$, let $((V, \rho), q : M \rightarrow X, e)$ denote the element of $\mathcal{W}_T(X)$ defined by:

1. $q : M \rightarrow X$ is the fibre of E^{rg} above $X \times \{-1\}$;
2. $e : \{v \in \text{Saddle}(V_0, \rho) \mid f_V(v) = -1\} \rightarrow M$ is the induced embedding, which has source $D(V_0^\rho) \times_{T \times X} S(V_0^{-\rho})$;
3. (V, ρ) is the object $h^*(V_0)$, considered as an object of $\mathcal{W}_{\text{loc}, T}(X)$.

PROPOSITION 7.19. — The construction of Definition 7.18 defines a morphism of sheaves $\mathcal{L}_T \rightarrow \mathcal{W}_T$, which is natural in T .

Proof. — The proof of the naturality with respect to T is a verification that the functorial behaviour of \mathcal{W}_T defined in Lemma 7.15 is compatible with the functorial behaviour of \mathcal{L}_T . □

The natural morphism fits into a commutative diagram

$$\begin{array}{ccc} \mathcal{L}_T & \longrightarrow & \mathcal{W}_T \\ \downarrow & & \downarrow \\ \mathcal{L}_{\text{loc}, T} & \longrightarrow & \mathcal{W}_{\text{loc}, T} \end{array}$$

hence the proof of Proposition 7.3 is completed by showing that the morphism $\mathcal{L}_T \rightarrow \mathcal{W}_T$ is a weak equivalence.

The long trace construction (Definition 6.13) induces a morphism (not natural with respect to T) of sheaves, $\mathcal{W}_T \rightarrow \mathcal{L}_T$. Proposition 7.3 can be deduced from the following result (for the proof, the reader is referred to [12]).

PROPOSITION 7.20. — The morphism $\mathcal{W}_T \rightarrow \mathcal{L}_T$ is a weak equivalence.

8. REFINEMENTS OF THE HOMOTOPY COLIMIT DECOMPOSITIONS

The stratifications of the previous section require to be refined in two ways. Firstly, it is necessary to add controlled boundaries to the 3-manifolds which are considered, so as to allow concatenation. The second refinement reduces to the case in which the fibres of the bundles are connected; this condition is required in Lemma 9.1.

8.1. Controlled boundaries

Let C be a non-empty closed, smooth, oriented 1-manifold. Any such manifold is null-bordant: there exists a smooth, compact oriented 2-manifold K with boundary $\partial K = C$ as oriented 1-manifolds.

Example 8.1. — For the application to the stable moduli space of surfaces, take C to be the manifold $S^1 \amalg -S^1$.

All of the sheaves considered in the previous section can be modified by allowing a constant boundary derived from C . For example, consider the sheaf \mathcal{W} .

DEFINITION 8.2. — *Let $\partial\mathcal{W}$ denote the sheaf with sections over X the set of pairs (π, f) , where $\pi : E \rightarrow X$ is a smooth submersion of fibre dimension three with oriented fibrewise tangent bundle, $f : E \rightarrow \mathbb{R}$ is a smooth map and ∂E over $X \times \mathbb{R}$ has a neighbourhood which is diffeomorphic to a collar on $X \times C \times \mathbb{R}$, respecting the orientation. The pair (π, f) should satisfy the properness and fibrewise Morse conditions of Definition 3.9.*

There is a natural morphism $\partial\mathcal{W} \rightarrow \mathcal{W}$ of sheaves, which is induced by the null-bordism of C . The reader is referred to [12] for a proof of the following result.

PROPOSITION 8.3. — *The morphism $\partial\mathcal{W} \rightarrow \mathcal{W}$ is a weak equivalence of sheaves.*

The definitions of all the sheaves of Section 7 generalize to the case where boundaries are permitted, as above. The analogue of Proposition 8.3 holds in each case: the null-bordism induces a natural morphism which is a weak equivalence. Henceforth, all such sheaves are taken with boundary, without addition of the decoration ∂ .

The definition of $\mathcal{W}_T(X)$ becomes:

DEFINITION 8.4. — *For T an object of \mathcal{K} , let \mathcal{W}_T denote the sheaf with sections over X given by tuples $((V, \rho), q : M \rightarrow X, e)$ such that :*

1. $(V, \rho) \in \mathcal{W}_{\text{loc}, T}$, of rank three;
2. $q : M \rightarrow X$ is a smooth bundle of closed 2-manifolds with boundary, with an orientation of the vertical tangent bundle;
3. there exists a neighbourhood of ∂M in M which identifies over X with a collar on $X \times C$, respecting the orientations;

4. $e : D(V^\rho) \times_{T \times X} S(V^{-\rho}) \rightarrow M$ is a smooth embedding over X , respecting orientations, with image disjoint from the boundary ∂M .

8.2. Concordance lifting and fibres

Proposition 8.6 below identifies the homotopy fibre of the morphism $|\mathcal{W}_T| \rightarrow |\mathcal{W}_{\text{loc},T}|$, by Corollary A.12, using the following concordance lifting property (see Definition A.9).

LEMMA 8.5. — *For T an object of \mathcal{K} , the forgetful morphism $\mathcal{W}_T \rightarrow \mathcal{W}_{\text{loc},T}$ has the concordance lifting property.*

Fix an object (V, ρ) of $\mathcal{W}_{\text{loc},T}(\ast)$; this is a three-dimensional oriented Morse vector bundle on the discrete space T . The canonical morphism $X \rightarrow \ast$ for a smooth manifold X , induces an associated object $(V, \rho)_X$ of $\mathcal{W}_{\text{loc},T}(X)$, such that the structure is trivial with respect to X .

PROPOSITION 8.6. — *For T an object of \mathcal{K} , the fibre over (V, ρ) of the forgetful morphism $\mathcal{W}_T \rightarrow \mathcal{W}_{\text{loc},T}$ is weakly equivalent to the sheaf which has sections over X the set of smooth bundles $(q : M \rightarrow X)$ of vertically tangentially oriented compact surfaces with collared boundary, where the boundary bundle $\partial M \rightarrow X$ identifies with*

$$-(C \amalg S(V^\rho) \times_T S(V^{-\rho})) \times X \rightarrow X.$$

Proof. — By cutting the interior of the embedded thickened spheres from the fibre bundle. □

8.3. The connectivity theorem

DEFINITION 8.7. — *For T an object of \mathcal{K} , let $\mathcal{W}_{c,T}$ denote the subsheaf of \mathcal{W}_T which has sections over X the set of tuples $((V, \rho), q : M \rightarrow X, e)$ such that the fibres of $M \setminus \text{Im}(e) \rightarrow X$ are connected.*

The construction is not functorial in T , since the extreme values $\{0, 3\}$ of the Morse index can lead to the introduction of non-connected fibres. This motivates the consideration of the following decomposition of the category \mathcal{K} .

DEFINITION 8.8. — *Let \mathcal{K}' denote the full subcategory of \mathcal{K} with objects such that the structure map $T \rightarrow \mathbf{[3]}$ has image contained in $\{0, 3\}$ and let \mathcal{K}'' be the full subcategory with objects such that the structure map $T \rightarrow \mathbf{[3]}$ has image contained in $\{1, 2\}$.*

The disjoint union of finite sets induces an equivalence of categories, $\mathcal{K}' \times \mathcal{K}'' \rightarrow \mathcal{K}$.

LEMMA 8.9. — *Let Q be an arbitrary object of \mathcal{K} , then the association $S \mapsto \mathcal{W}_{c,Q|S}$ is contravariantly functorial in $S \in \mathcal{K}''$.*

The product decomposition of the category \mathcal{K} implies the following result:

LEMMA 8.10. — *Let $T \mapsto \mathcal{F}_T$ be a functor from \mathcal{K}^{op} to sheaves on \mathfrak{X} ; then there is a weak equivalence $\text{hocolim}_{T \in \mathcal{K}} |\mathcal{F}_T| \simeq \text{hocolim}_{Q \in \mathcal{K}'} \text{hocolim}_{S \in \mathcal{K}''} |\mathcal{F}_{Q|S}|$.*

The following technical result⁽¹²⁾ avoids the necessity of considering the homotopy fibre of the homotopy colimit over \mathcal{K}'' of $\mathcal{W}_{c,Q\Pi S} \rightarrow \mathcal{W}_{Q\Pi S}$, for non-trivial Q . The reader is referred to [12] for the proof.

LEMMA 8.11. — *For any morphism $P \rightarrow Q$ in \mathcal{K}' , the following commutative square is homotopy cartesian:*

$$\begin{array}{ccc} \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{Q\Pi S}| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{P\Pi S}| \\ \downarrow & & \downarrow \\ \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},Q\Pi S}| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},P\Pi S}|. \end{array}$$

The connectivity theorem reads as follows:

THEOREM 8.12. — *The inclusion $\mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{c,S}| \rightarrow \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_S|$ is a weak equivalence.*

8.4. Indications on the proof of the connectivity theorem

The proof of Theorem 8.12 is by the construction of a homotopy inverse. Fibrewise surgery is used to make the fibres connected manifolds; this is achieved by cutting out pairs of disks and gluing in tubes. Such surgeries can be described in terms of the functoriality of the sheaves \mathcal{W}_T with respect to $\mathcal{K}^{\mathrm{op}}$.

The proof uses methods of homotopical algebra to compare the surgery techniques which correspond to the choice of the regularization parameter $\{\pm 1\}$ for morphisms in the category \mathcal{K} . This reduces the proof to showing that the category which parametrizes the multiple surgeries which make a surface connected is contractible.

The reader is referred to [12] for the proof that the classifying space is contractible, which uses the criterion, derived from Proposition A.7, for a sheaf to be contractible.

9. THE PROOF OF THE MAIN THEOREM

Throughout this section, let C denote the manifold $S^1 \amalg -S^1$. The proof of Theorem 1.2 is completed by proving that the homotopy fibre of $|\mathcal{W}| \rightarrow |\mathcal{W}_{\mathrm{loc}}|$ is homotopy equivalent to $\mathbb{Z} \times B\Gamma_{\infty,2}^+$.

⁽¹²⁾A homotopy cartesian diagram corresponds to the homotopy limit, and is a derived version of the fibred product.

9.1. Reduction to integral homology

The Whitehead theorem implies that it is sufficient to construct a morphism

$$\mathbb{Z} \times B\Gamma_{\infty,2} \rightarrow \text{hfib}\{|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|\}$$

which induces an isomorphism in integral homology. Indeed, the morphism $|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|$ is homotopic to a morphism of infinite loop spaces and therefore factorizes up to homotopy across Quillen’s plus construction: $\mathbb{Z} \times B\Gamma_{\infty,2}^+ \rightarrow \text{hfib}\{|\mathcal{W}| \rightarrow |\mathcal{W}_{\text{loc}}|\}$. This is a morphism between infinite loop spaces (not *a priori* an infinite loop map), which induces an isomorphism in integral homology; the Whitehead theorem implies that it is a weak equivalence.

9.2. Homotopy fibres for connected strata

The following result is fundamental; it follows from Proposition 8.6, restricted to the connected case.

LEMMA 9.1. — *Let T be an object of \mathcal{K} , then the homotopy fibre over any basepoint of the localization map $|\mathcal{W}_{c,T}| \rightarrow |\mathcal{W}_{\text{loc},T}|$ is homotopy equivalent to $\amalg_g B\Gamma_{g,2+2|T|}$.*

9.3. Stabilization via the genus

The collared boundary of the total space E of an object of $\mathcal{W}_T(X)$ allows for concatenation; this is a parametrized version of the concatenation used in the Harer-Ivanov cohomological stability theorems.

DEFINITION 9.2. — *Let $z \in \mathcal{W}_{c,\emptyset}(\ast)$ be an oriented surface of genus one, with two boundary components. For T an object of \mathcal{K} :*

1. *let $.z : \mathcal{W}_T \rightarrow \mathcal{W}_T$ and $.z : \mathcal{W}_{c,T} \rightarrow \mathcal{W}_{c,T}$ denote the morphisms which are induced by concatenation with z ;*
2. *let $z^{-1}\mathcal{W}_T$ (respectively $z^{-1}\mathcal{W}_{c,T}$) denote the colimit of the induced direct system of sheaves.*

The spaces represented by the colimits are weakly equivalent to the colimits of the individual spaces, by the following result:

LEMMA 9.3 ([12]). — *There are natural weak equivalences $|z^{-1}\mathcal{W}_T| \simeq z^{-1}|\mathcal{W}_T|$, $|z^{-1}\mathcal{W}_{c,T}| \simeq z^{-1}|\mathcal{W}_{c,T}|$.*

Lemma 9.1 stabilizes to yield the following result:

COROLLARY 9.4. — *Let T be an object of \mathcal{K} ; then the homotopy fibre of the localization map $|z^{-1}\mathcal{W}_{c,T}| \rightarrow |\mathcal{W}_{\text{loc},T}|$ is homotopy equivalent to $\mathbb{Z} \times B\Gamma_{\infty,2+2|T|}$ over any basepoint.*

For Q a fixed object of \mathcal{K} , the associations $S \mapsto z^{-1}\mathcal{W}_S$, $S \mapsto z^{-1}\mathcal{W}_{c,Q \amalg S}$ are contravariantly functorial in S in \mathcal{K}'' . Theorem 8.12 implies the following

COROLLARY 9.5. — *The map $\text{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{c,S}| \rightarrow \text{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_S|$ is a weak equivalence.*

The realization spaces of \mathcal{W} and $z^{-1}\mathcal{W}$ by the following proposition, which follows from the fact that $|\mathcal{W}|$ is an infinite loop space (by Theorems 1.3 and 1.4).

PROPOSITION 9.6. — *There are weak equivalences $|\mathcal{W}| \simeq |z^{-1}\mathcal{W}| \simeq \operatorname{hocolim}_{T \in \mathcal{K}} |z^{-1}\mathcal{W}_T|$.*

9.4. Morphisms and homology equivalences

For any small category \mathcal{C} and any functor F_- from \mathcal{C} to a suitable category of topological spaces, Spaces, there exists a canonical morphism $F_C \rightarrow \operatorname{hocolim} F_-$, for any object C of \mathcal{C} .

LEMMA 9.7. — *Let Q be an object of \mathcal{K} . There is a canonical morphism*

$$\mathbb{Z} \times B\Gamma_{\infty, 2+2|Q|} \rightarrow \operatorname{hfib}\{\operatorname{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{c, Q \amalg S}| \rightarrow \operatorname{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\operatorname{loc}, Q \amalg S}|\}.$$

Proof. — In the case $Q = \emptyset$, it suffices to observe that $\emptyset \in \mathcal{K}''$ and that $|z^{-1}\mathcal{W}_{c, \emptyset}| \simeq \mathbb{Z} \times B\Gamma_{\infty, 2}$ and $|\mathcal{W}_{\operatorname{loc}, \emptyset}| \simeq *$. The canonical morphism is the required morphism. The general case is a straightforward modification of this argument. \square

The following technical result relates the homology of the homotopy fibre of a map between spaces obtained by gluing to the homology of the homotopy fibres of the maps between individual terms.

PROPOSITION 9.8 ([12, 13]). — *Let \mathcal{C} be a small category and let $u : \mathcal{G}_1 \rightarrow \mathcal{G}_2$ be a natural transformation between functors $\mathcal{G}_i : \mathcal{C} \rightarrow \text{Spaces}$.*

Suppose that, for any morphism $f : a \rightarrow b$ of \mathcal{C} , the map f_ from any homotopy fibre of u_a to the corresponding homotopy fibre of u_b induces an isomorphism in integral homology. Then, for any object a of \mathcal{C} , the inclusion of any homotopy fibre of u_a in the corresponding homotopy fibre of $u_* : \operatorname{hocolim} \mathcal{G}_1 \rightarrow \operatorname{hocolim} \mathcal{G}_2$ induces an isomorphism in integer homology.*

PROPOSITION 9.9. — *The canonical morphism*

$$\mathbb{Z} \times B\Gamma_{\infty, 2} \rightarrow \operatorname{hfib}\{\operatorname{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{c, S}| \rightarrow \operatorname{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\operatorname{loc}, S}|\}$$

induces an isomorphism in integral homology.

Proof. — There is a commutative diagram of homotopy fibre sequences, for $S \rightarrow T$ a morphism of \mathcal{K}'' :

$$\begin{array}{ccccc} \mathbb{Z} \times B\Gamma_{\infty, 2+2|T|} & \longrightarrow & |z^{-1}\mathcal{W}_{c, T}| & \longrightarrow & |\mathcal{W}_{\operatorname{loc}, T}| \\ \downarrow & & \downarrow & & \downarrow \\ \mathbb{Z} \times B\Gamma_{\infty, 2+2|S|} & \longrightarrow & |z^{-1}\mathcal{W}_{c, S}| & \longrightarrow & |\mathcal{W}_{\operatorname{loc}, S}| \end{array}$$

where the map of homotopy fibres corresponds geometrically to attaching tubes $D^1 \times S^1$ or pairs of disks $D^2 \times S^0$, according to the regularization index. The Harer-Ivanov stability theorems imply that these morphisms induce an isomorphism in homology. The result follows by applying Proposition 9.8. \square

9.5. Applying the homotopy cartesian square

Lemma 8.11 stabilizes to give the following result.

LEMMA 9.10. — *For any morphism $P \rightarrow Q$ in \mathcal{K}' , the following commutative square is homotopy cartesian:*

$$\begin{array}{ccc} \mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{Q\mathrm{II}S}| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{P\mathrm{II}S}| \\ \downarrow & & \downarrow \\ \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},Q\mathrm{II}S}| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},P\mathrm{II}S}| \end{array}$$

This allows the deduction of the following result.

LEMMA 9.11. — *For any object Q of \mathcal{K} , the canonical morphism*

$$\mathbb{Z} \times B\Gamma_{\infty,2+2|Q|} \rightarrow \mathrm{hfib}\{\mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{Q\mathrm{II}S}| \rightarrow \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},Q\mathrm{II}S}|\}$$

induces an isomorphism in integral homology.

Proof. — There is a commutative diagram

$$\begin{array}{ccccccc} \mathbb{Z} \times B\Gamma_{\infty,2+2|Q|} & \longrightarrow & F & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{Q\mathrm{II}S}| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},Q\mathrm{II}S}| \\ \downarrow & & \simeq \downarrow & & \downarrow & & \downarrow \\ \mathbb{Z} \times B\Gamma_{\infty,2} & \longrightarrow & F' & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_S| & \longrightarrow & \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},S}| \end{array}$$

in which F, F' denote the respective homotopy fibres, which are weakly equivalent, since the right hand square is homotopy cartesian.

The left hand vertical morphism is an integral homology isomorphism, by the Harer-Ivanov stability theorems and the bottom left hand morphism induces an isomorphism in integral homology, by the result of the previous section and Corollary 9.4, which implies that the term $\mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_S|$ is equivalent to $\mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{c,S}|$.

The result follows. \square

9.6. Proof of the main theorem

The proof of the Theorem 1.5 is completed by repeating the above argument, using Proposition 9.8 together with the Harer-Ivanov stability theorem, applied to the homotopy fibre of the morphism:

$$\mathrm{hocolim}_{Q \in \mathcal{K}'} \mathrm{hocolim}_{S \in \mathcal{K}''} |z^{-1}\mathcal{W}_{Q\mathrm{II}S}| \rightarrow \mathrm{hocolim}_{Q \in \mathcal{K}'} \mathrm{hocolim}_{S \in \mathcal{K}''} |\mathcal{W}_{\mathrm{loc},Q\mathrm{II}S}|.$$

APPENDIX A

HOMOTOPY THEORY OF SHEAVES

Throughout this section, all sheaves are defined on the category \mathfrak{X} of smooth manifolds.

A.1. Concordance for sheaves

DEFINITION A.1. — Let \mathcal{F} be a sheaf and let X be a smooth manifold. Two sections $s_0, s_1 \in \mathcal{F}(X)$ are concordant if there exists a section $s \in \mathcal{F}(X \times \mathbb{R})$ such that $s = p_X^* s_0$ on an open neighbourhood of $X \times (-\infty, 0]$ and $s = p_X^* s_1$ on an open neighbourhood of $X \times [1, \infty)$, where $p_X : X \times \mathbb{R} \rightarrow X$ is the projection. The concordance s is said to start at s_0 .

Concordance defines an equivalence relation on the set $\mathcal{F}(X)$; the set of equivalence classes modulo concordance is written $\mathcal{F}[X]$.

There is a relative version of concordance with respect to a closed subset $A \subset X$ (not necessarily a manifold) of a smooth manifold X .

DEFINITION A.2. — For $A \subset X$ a closed subset of a smooth manifold X in \mathfrak{X} ,

1. let \mathcal{F}_A denote the colimit $\mathcal{F}_A := \operatorname{colim}_U \mathcal{F}(U)$, where U ranges over the category of open neighbourhoods of A in X ;
2. for $s \in \mathcal{F}_A$, let $\mathcal{F}(X, A; s)$ denote the fibre of the canonical morphism $\mathcal{F}(X) \rightarrow \mathcal{F}_A$ above s .

DEFINITION A.3. — For \mathcal{F}, X, A, s as above, elements $t_0, t_1 \in \mathcal{F}(X, A; s)$ are concordant relative to A if there exists an element $t \in \mathcal{F}(X \times \mathbb{R}, A \times \mathbb{R}; q^* s)$ which defines a concordance for t_0, t_1 regarded as elements of $\mathcal{F}(X)$, where $q^* s \in \mathcal{F}_{A \times \mathbb{R}}$ is induced by pullback of s via the projection $q : A \times \mathbb{R} \rightarrow A$.

Let $\mathcal{F}[X, A; s]$ denote the set $\mathcal{F}(X, A; s)$ modulo the equivalence relation given by concordance relative to A .

A.2. The representing space $|\mathcal{F}|$

Let $\Delta^n \subset \mathbb{R}^{n+1}$ denote the affine plane $\sum x_i = 1$, so that $n \mapsto \Delta^n$ defines a cosimplicial smooth manifold.

DEFINITION A.4. — For \mathcal{F} a sheaf of sets, let $|\mathcal{F}|$ denote the geometric realization of the simplicial set $\mathcal{F}(\Delta^*)$.

A point $z \in \mathcal{F}(*)$ induces a point in $|\mathcal{F}|$. The following representability statement is fundamental.

PROPOSITION A.5 ([12]). — For any point $z \in \mathcal{F}(*)$, smooth manifold X and closed subset $A \subset X$, there is a natural isomorphism $\mathcal{F}[X, A; z] \cong [(X, A), (|\mathcal{F}|, z)]$, where the right hand side denotes homotopy classes of maps of pairs.

The following formal properties are basic.

1. The functor $\mathcal{F} \mapsto |\mathcal{F}|$ takes pullback squares of sheaves to pullbacks of compactly-generated Hausdorff spaces.
2. $|\mathcal{F}_1 \amalg \mathcal{F}_2| \cong |\mathcal{F}_1| \amalg |\mathcal{F}_2|$.

DEFINITION A.6. — *A morphism of sheaves of sets $v : \mathcal{E} \rightarrow \mathcal{F}$ is a weak equivalence if $|v| : |\mathcal{E}| \rightarrow |\mathcal{F}|$ is a weak equivalence of topological spaces. This is equivalent to the induced morphism $\pi_n(|\mathcal{E}|, z) \cong \mathcal{E}[S^n, e; z] \rightarrow \pi_n(|\mathcal{F}|, z) \cong \mathcal{F}[S^n, e; z]$ being an isomorphism, for all n, z .*

A.3. Special weak equivalences

The following special weak equivalences resemble trivial fibrations in homotopical algebra.

PROPOSITION A.7 ([12]). — *Let $v : \mathcal{E} \rightarrow \mathcal{F}$ be a morphism of sheaves of sets. Suppose that v induces a surjection $\mathcal{E}[X, A; s] \rightarrow \mathcal{F}[X, A; v(s)]$ for all smooth X and all $s \in \mathcal{E}_A$; then v is a weak equivalence.*

Example A.8. — Let \mathcal{F} be the terminal sheaf which is induced by the constant presheaf with value a singleton set; for a sheaf of sets, \mathcal{E} , there is a canonical morphism of sheaves, $\mathcal{E} \rightarrow \mathcal{F}$. The morphism induces a surjection on relative concordance classes, as in the proposition above, if and only if any section $s \in \mathcal{E}_A$ extends to an element of $\mathcal{E}[X, A; s]$. This yields a criterion for a sheaf to be homotopically trivial.

A.4. The concordance lifting property

The concordance lifting property plays the rôle of the fibration hypothesis in homotopical algebra; it is important in considering the homotopy fibre of a morphism of sheaves (see Corollary A.12 below).

DEFINITION A.9. — *A morphism $u : \mathcal{E} \rightarrow \mathcal{F}$ of sheaves has the concordance lifting property if, for any section $s \in \mathcal{E}(X)$ and concordance $h \in \mathcal{F}(X \times \mathbb{R})$ starting at $u(s)$, there exists a concordance $H \in \mathcal{E}(X \times \mathbb{R})$, starting at s , which lifts h .*

PROPOSITION A.10 ([12]). — *Let $\mathcal{E} \times_{\mathcal{G}} \mathcal{F}$ be the pullback of the diagram of sheaves, $\mathcal{E} \rightarrow \mathcal{G} \xleftarrow{u} \mathcal{F}$. If u has the concordance lifting property, then so does $\mathcal{E} \times_{\mathcal{G}} \mathcal{F} \rightarrow \mathcal{E}$ and the associated diagram of spaces,*

$$\begin{array}{ccc} |\mathcal{E} \times_{\mathcal{G}} \mathcal{F}| & \longrightarrow & |\mathcal{F}| \\ \downarrow & & \downarrow |u| \\ |\mathcal{E}| & \longrightarrow & |\mathcal{G}|, \end{array}$$

is homotopy cartesian.

DEFINITION A.11. — *For $u : \mathcal{E} \rightarrow \mathcal{F}$ a morphism of sheaves and $a \in \mathcal{F}(*)$, let \mathcal{E}_a denote the fibre of u over a , which is the sheaf defined by $\mathcal{E}_a(X) := \{s \in \mathcal{E}(X) | u(s) = a\}$.*

COROLLARY A.12 ([12]). — *Let $u : \mathcal{E} \rightarrow \mathcal{F}$ be a morphism of sheaves which has the concordance lifting property, then the induced diagram of spaces $|\mathcal{E}_a| \rightarrow |\mathcal{E}| \rightarrow |\mathcal{F}|$ is a homotopy fibre sequence.*

A.5. Functors to small categories

Let \mathcal{F} be a sheaf with values in the category of small categories. The nerve construction [19] $N_\bullet\mathcal{F}$ defines a sheaf with values in the category of simplicial sets; hence $N_\bullet\mathcal{F}(\Delta^*)$ has the structure of a bisimplicial set. This can be realized as a topological space in a number of ways; in particular, the topological realization of the diagonal simplicial set is weakly equivalent to the classifying space $B|\mathcal{F}|$ of the topological category $|\mathcal{F}|$. (Here $|\mathcal{F}|$ is the topological category with object space $|N_0\mathcal{F}|$ and morphism space $|N_1\mathcal{F}|$.)

DEFINITION A.13 ([12]). — For \mathcal{F} as above, let $\beta\mathcal{F}$ denote the sheaf of sets which has sections over X the set of pairs $(\mathfrak{Y}, \phi_{\bullet\bullet})$, where $\mathfrak{Y} := \{Y_j | j \in \mathcal{J}\}$ is a locally finite open cover of X , indexed by a fixed index set \mathcal{J} , and, for nonempty finite subsets $R \subset S$, $\phi_{RS} \in N_1\mathcal{F}(Y_S)$ is a morphism subject to cocycle conditions, where Y_S denotes the open set $Y_S := \bigcap_{j \in S} Y_j$.

The following theorem is proved in [12] and gives a sheaf theoretic model for the classifying space $B|\mathcal{F}|$.

THEOREM A.14. — The spaces $|\beta\mathcal{F}|$ and $B|\mathcal{F}|$ are homotopy equivalent.

A.6. Homotopy colimits

Homotopy colimits are used for gluing; they behave as the derived functor of the colimit and, in particular, satisfy a homotopy invariance property.

DEFINITION A.15. — Let \mathcal{F}_C be a sheaf of sets, functorial in objects C of a small, discrete category \mathcal{C} .

1. Let $\mathcal{C} \int \mathcal{F}$ denote the sheaf with values in small categories which associates to X the small category with objects pairs (C, w) , where C is an object of \mathcal{C} and $w \in \mathcal{F}_C(X)$. The morphisms are defined in the usual way.
2. Let $\text{hocolim}_{C \in \mathcal{C}} |\mathcal{F}_C|$ be the space

$$B|\mathcal{C} \int \mathcal{F}_\bullet| \simeq |\beta(\mathcal{C} \int \mathcal{F}_\bullet)|.$$

For \mathcal{F}, \mathcal{C} as above, it makes sense to define the homotopy colimit as a sheaf:

$$\text{hocolim}_{C \in \mathcal{C}} \mathcal{F}_C := \beta(\mathcal{C} \int \mathcal{F}_\bullet).$$

The structure of this object is made explicit in [12].

DEFINITION A.16. — Let $g : \mathcal{E} \rightarrow \mathcal{F}$ be a natural transformation of sheaves of small categories. The natural transformation, g , is a transport projection if the square

$$\begin{array}{ccc} N_1\mathcal{E} & \longrightarrow & N_0\mathcal{E} \\ \downarrow & & \downarrow \\ N_1\mathcal{F} & \longrightarrow & N_0\mathcal{F} \end{array}$$

is a cartesian square of sheaves of sets.

The following gluing result for homotopy colimits allows the passage from local to global.

PROPOSITION A.17 ([12]). — *Let $g : \mathcal{E} \rightarrow \mathcal{F}$, $g' : \mathcal{E}' \rightarrow \mathcal{F}$ be transport projections between sheaves of small categories and let $u : \mathcal{E} \rightarrow \mathcal{E}'$ be a morphism of sheaves of small categories above \mathcal{F} . Suppose that*

1. *the morphisms $N_0\mathcal{E} \rightarrow N_0\mathcal{F}$, $N_0\mathcal{E}' \rightarrow N_0\mathcal{F}$ have the concordance lifting property;*
2. *for all $a \in \mathcal{F}(*)$, the morphism $N_0\mathcal{E}_a \rightarrow N_0\mathcal{E}'_a$ induces a weak equivalence (respectively an integral homology equivalence).*

Then the morphism $\beta u : \beta\mathcal{E} \rightarrow \beta\mathcal{E}'$ is a weak equivalence (respectively induces an integral homology equivalence).

REFERENCES

- [1] T. BRÖCKER & K. JÄNICH – *Introduction to differential topology*, Cambridge University Press, Cambridge, 1982, Translated from the German by C. B. Thomas and M. J. Thomas.
- [2] C. J. EARLE & J. EELLS – “A fibre bundle description of Teichmüller theory”, *J. Differential Geometry* **3** (1969), p. 19–43.
- [3] C. J. EARLE & A. SCHATZ – “Teichmüller theory for surfaces with boundary”, *J. Differential Geometry* **4** (1970), p. 169–185.
- [4] S. GALATIUS – “Mod p homology of the stable mapping class group”, *Topology* **43** (2004), no. 5, p. 1105–1132.
- [5] R. HAIN & E. LOOIJENGA – “Mapping class groups and moduli spaces of curves”, in *Algebraic geometry—Santa Cruz 1995* (Providence, RI), Proc. Sympos. Pure Math., vol. 62, Amer. Math. Soc., 1997, p. 97–142.
- [6] J. HARER – “The second homology group of the mapping class group of an orientable surface”, *Invent. Math.* **72** (1983), no. 2, p. 221–239.
- [7] ———, “Stability of the homology of the mapping class groups of orientable surfaces”, *Ann. of Math. (2)* **121** (1985), no. 2, p. 215–249.
- [8] J. HARRIS & I. MORRISON – *Moduli of curves*, Graduate Texts in Mathematics, vol. 187, Springer-Verlag, New York, 1998.
- [9] N. V. IVANOV – “Stabilization of the homology of Teichmüller modular groups”, *Algebra i Analiz* **1** (1989), no. 3, p. 110–126.
- [10] I. MADSEN & R. J. MILGRAM – *The classifying spaces for surgery and cobordism of manifolds*, Annals of Mathematics Studies, vol. 92, Princeton University Press, Princeton, N.J., 1979.
- [11] I. MADSEN & U. TILLMANN – “The stable mapping class group and $Q(\mathbb{C}P_+^\infty)$ ”, *Invent. Math.* **145** (2001), no. 3, p. 509–544.
- [12] I. MADSEN & M. WEISS – “The stable moduli space of Riemann surfaces: Mumford’s conjecture”, *Preprint* (2004), p. 97.

- [13] D. MCDUFF & G. SEGAL – “Homology fibrations and the “group-completion” theorem”, *Invent. Math.* **31** (1975/76), no. 3, p. 279–284.
- [14] E. Y. MILLER – “The homology of the mapping class group”, *J. Differential Geom.* **24** (1986), no. 1, p. 1–14.
- [15] S. MORITA – “Characteristic classes of surface bundles”, *Invent. Math.* **90** (1987), no. 3, p. 551–577.
- [16] ———, “Structure of the mapping class groups of surfaces: a survey and a prospect”, in *Proceedings of the Kirbyfest (Berkeley, CA, 1998)*, Geom. Topol. Monogr., vol. 2, Geom. Topol. Publ., Coventry, 1999, p. 349–406 (electronic).
- [17] D. MUMFORD – “Towards an enumerative geometry of the moduli space of curves”, in *Arithmetic and geometry, Vol. II*, Progr. Math., vol. 36, Birkhäuser Boston, Boston, MA, 1983, p. 271–328.
- [18] D. QUILLEN – “Elementary proofs of some results of cobordism theory using Steenrod operations”, *Advances in Math.* **7** (1971), p. 29–56 (1971).
- [19] G. SEGAL – “Categories and cohomology theories”, *Topology* **13** (1974), p. 293–312.
- [20] R. E. STONG – *Notes on cobordism theory*, Mathematical notes, Princeton University Press, Princeton, N.J., 1968.
- [21] U. TILLMANN – “On the homotopy of the stable mapping class group”, *Invent. Math.* **130** (1997), no. 2, p. 257–275.
- [22] V. A. VASSILIEV – “Topology of spaces of functions without complicated singularities”, *Funktsional Anal. i Prilozhen* **93** (1989), p. 24–36.
- [23] ———, *Complements of discriminants of smooth maps: topology and applications*, Translations of Mathematical Monographs, vol. 98, American Mathematical Society, Providence, RI, 1992, Translated from the Russian by B. Goldfarb.

Geoffrey POWELL

Université Paris XIII

Institut Galilée

LAGA – Département de Mathématiques

UFR 7539 du CNRS

99, avenue J.-B. Clément

F-93430 VILLETANEUSE

E-mail : powell@math.univ-paris13.fr

CATÉGORIES DÉRIVÉES ET GÉOMÉTRIE BIRATIONNELLE
[d'après Bondal, Orlov, Bridgeland, Kawamata...]

par Raphaël ROUQUIER

1. INTRODUCTION

L'objet principal de cet exposé est la catégorie dérivée $D^b(X)$ des faisceaux cohérents sur une variété X . La catégorie dérivée organise l'information homologique (groupes d'extensions entre faisceaux cohérents) et numérique (K -théorie). Nous allons étudier son comportement au cours des opérations de « chirurgie algébrique » (« flips » et « flops »).

La catégorie dérivée d'un espace projectif se décrit à partir d'une algèbre de dimension finie (Beilinson, 1978) et ceci a placé dans un cadre approprié les descriptions de fibrés vectoriels en terme d'algèbre linéaire. À la suite de ce résultat, des descriptions analogues (décomposition semi-orthogonale de la catégorie dérivée) ont été recherchées pour d'autres variétés. De telles décompositions devraient apparaître en présence d'un « flip », étape cruciale du programme de Mori de modèles minimaux (MMP) pour la classification des variétés projectives lisses, et cela a amené en particulier la question de l'invariance de la catégorie dérivée par « flop » (Bondal-Orlov, 1995). D'un autre côté, la conjecture homologique de symétrie miroir (Kontsevich, 1994) a elle aussi posé le problème de l'invariance birationnelle de la catégorie dérivée, pour des variétés de Calabi-Yau. Indépendamment, la construction d'une équivalence dérivée entre une variété abélienne et sa duale (Mukai, 1981) a montré la relation entre la réalisation d'une variété X comme un espace de modules d'objets sur Y et la construction d'une équivalence (ou d'un foncteur pleinement fidèle) entre la catégorie dérivée de X et celle de Y .

Commençons par poser des problèmes sur les catégories dérivées de variétés, en suivant trois points importants du MMP.

On note K_X le diviseur canonique d'une variété lisse X et on considère l'équivalence linéaire entre diviseurs. On se donne un diagramme où f et g sont des morphismes

birationnels entre variétés projectives lisses complexes

(1)

$$\begin{array}{ccc} & Z & \\ f \swarrow & & \searrow g \\ X & & Y \end{array}$$

On a une première conjecture sur les flops généralisés (cf. [8, Conjecture 4.4] et [29, Conjecture 1.2]) :

CONJECTURE 1.1 (Bondal-Orlov). — *Si $f^*K_X \sim g^*K_Y$, alors $D^b(X) \simeq D^b(Y)$.*

On sait que dans cette situation les nombres de Hodge coïncident (cf. remarque 3.9). La conjecture a une réponse positive en dimension 3 (corollaire 4.11), pour des variétés symplectiques de dimension 4 (corollaire 4.7) et dans le cadre torique (théorème 4.15). Pour des variétés de Calabi-Yau (ω trivial), on s'attend donc à ce que birationalité et équivalence dérivée coïncident, comme le prédit la conjecture de Kontsevich de symétrie miroir [35]. La « réciproque » de la conjecture 1.1 n'est pas vraie (remarque 3.15).

On a une conjecture sur les flips généralisés [8, Conjecture 4.4] :

CONJECTURE 1.2 (Bondal-Orlov). — *Si $f^*K_X - g^*K_Y$ est équivalent à un diviseur effectif, alors il existe un foncteur pleinement fidèle $D^b(Y) \rightarrow D^b(X)$.*

La minimisation d'une variété dans le MMP devrait alors s'interpréter comme une minimisation de la catégorie dérivée, un modèle minimal pour une variété X devant être construit comme un espace de module d'objets de la catégorie dérivée de X . Dans la conjecture 1.2, il serait aussi souhaitable de savoir décrire l'orthogonal de l'image de $D^b(Y)$ dans $D^b(X)$.

On a enfin une conjecture de finitude [29, Conjecture 1.5] :

CONJECTURE 1.3 (Kawamata). — *Soit X une variété projective lisse. Alors, il n'existe qu'un nombre fini de classes d'isomorphisme de variétés projectives lisses Y telles que $D^b(X) \simeq D^b(Y)$.*

La réponse est positive en dimension ≤ 2 (cf. [14, Corollary 1.2] et [29, Theorem 1.6]) et pour X, Y des variétés abéliennes [47, Corollary 2.8].

Dans la première partie de cet exposé, nous montrons, dans la situation extrême où le fibré canonique est ample ou anti-ample, comment reconstruire la variété à partir de la catégorie dérivée (théorème 2.1). Nous expliquons ensuite le mécanisme de dévissage de catégories dérivées (§2.2). Dans le §3, nous présentons la construction de transformations à noyau et les invariants transportés par les équivalences, puis nous exposons le cas des variétés abéliennes et des surfaces, où la théorie est presque complète. La partie §4 présente plusieurs cas de réponse positive aux conjectures 1.1 et 1.2.

Outre le livre en préparation [26], le lecteur pourra consulter [8, 18, 25, 48] pour des exposés généraux.

Je remercie Arnaud Beauville, Tom Bridgeland, Olivier Debarre et Alastair King pour leurs remarques sur une version préliminaire de ce texte, et Daniel Huybrechts, pour de très nombreuses discussions.

2. PROPRIÉTÉS INTERNES

2.1. Reconstruction

2.1.1. Terminologie. — Une *variété* sera pour nous un schéma quasi-projectif X sur \mathbf{C} . La plupart du temps, il s'agira de variétés projectives lisses. Les points considérés seront toujours des points fermés.

La catégorie dérivée $D^b(X)$ est définie comme la localisation de la catégorie des complexes bornés de faisceaux cohérents en la classe des quasi-isomorphismes (=morphisme de complexes qui induisent un isomorphisme entre faisceaux de cohomologie). Ses objets sont donc les complexes bornés de faisceaux cohérents sur X . Les flèches sont obtenues à partir de morphismes de complexes auxquels les inverses des quasi-isomorphismes ont été ajoutés. La catégorie $D^b(X)$ n'est pas abélienne en général, mais elle possède la structure de catégorie triangulée : le rôle des suites exactes courtes est joué par les triangles distingués.

Tous les foncteurs considérés entre catégories triangulées seront triangulés. Une sous-catégorie *épaisse* d'une catégorie triangulée est une sous-catégorie triangulée pleine close par facteurs directs. La sous-catégorie d'une catégorie triangulée *engendrée* (resp. *faiblement engendrée*) par une famille d'objets est la plus petite sous-catégorie pleine triangulée (resp. la plus petite sous-catégorie épaisse) contenant cette famille.

2.1.2. Catégories abéliennes. — Soit X une variété lisse. On sait, depuis Gabriel [23], que la catégorie des faisceaux cohérents X -coh sur X détermine X :

- l'application qui à un fermé associe les faisceaux supportés par ce fermé induit une bijection $Z \mapsto \mathcal{I}_Z$ de l'ensemble des fermés de X vers l'ensemble des sous-catégories de Serre de X -coh engendrées par un objet
- la catégorie quotient X -coh/ \mathcal{I}_Z est équivalente à $(X - Z)$ -coh et son centre s'identifie à $\mathcal{O}_X(X - Z)$.

Suivant Thomason et Balmer [3] (voir aussi [51, Theorem 3.11]), ce principe de reconstruction s'étend à la catégorie $D^b(X)$, si on munit celle-ci de sa structure tensorielle, en plus de sa structure triangulée :

- l'application qui à un fermé Z de X associe la sous-catégorie pleine $D_Z^b(X)$ de $D^b(X)$ des complexes dont les faisceaux de cohomologie sont supportés par Z est injective, d'image l'ensemble des sous-catégories épaisses faiblement engendrées par un élément et \otimes -idéales (*i.e.*, stables par $- \otimes L$ pour tout $L \in D^b(X)$) [58, Theorem 3.15]
- la catégorie $D^b(X - Z)$ s'identifie à $D^b(X)/D_Z^b(X)$ et, si $X - Z$ est un ouvert affine, le centre de $D^b(X - Z)$ s'identifie à $\mathcal{O}_X(X - Z)$.

Par conséquent, $X\text{-coh}$, vue comme catégorie abélienne, et $D^b(X)$, vue comme catégorie triangulée tensorielle, ne sont pas des invariants intéressants de X !

2.1.3. Fibré canonique (anti-)ample. — La catégorie $D^b(X)$, munie de sa seule structure de catégorie triangulée, ne détermine pas la variété X , mais elle apparaît comme un invariant intéressant. Dans la suite, les catégories dérivées seront considérées avec leur seule structure triangulée (voir à ce sujet §3.1.4). Le premier exemple (Mukai) est celui de l'équivalence dérivée entre une variété abélienne et sa duale (théorème 3.11).

Bondal et Orlov démontrent que lorsque le fibré canonique ω_X est ample ou anti-ample, alors la variété est déterminée par sa catégorie dérivée [7, Theorem 2.5].

THÉORÈME 2.1 (Bondal-Orlov). — *Soit X une variété projective lisse telle que ω_X ou ω_X^{-1} est ample. Si Y est une variété projective lisse et si on a une équivalence de catégories triangulées $D^b(X\text{-coh}) \simeq D^b(Y\text{-coh})$, alors $X \simeq Y$.*

Un point crucial est joué dans la preuve par la notion de foncteur de Serre [5, §3]. Soit \mathcal{T} une catégorie \mathbf{C} -linéaire telle que $\dim \text{Hom}(M, N) < \infty$ pour tous $M, N \in \mathcal{T}$. Un foncteur de Serre est la donnée d'une auto-équivalence $S : \mathcal{T} \xrightarrow{\sim} \mathcal{T}$ et d'isomorphismes bifonctoriels pour tous $M, N \in \mathcal{T}$:

$$\text{Hom}(M, N) \xrightarrow{\sim} \text{Hom}(N, SM)^*.$$

Si on voit \mathcal{T} comme une « algèbre avec plusieurs objets », alors ceci correspond à la notion d'algèbre de Frobenius.

Un foncteur de Serre, s'il existe, est unique à isomorphisme unique près. Par conséquent, une équivalence de catégories commute avec les foncteurs de Serre. En outre, si \mathcal{T} est une catégorie triangulée, alors un foncteur de Serre est automatiquement triangulé.

La définition est motivée par la dualité de Serre :

THÉORÈME 2.2. — *Si X est une variété projective lisse purement de dimension n , alors $S = \omega_X[n] \otimes -$ est un foncteur de Serre pour $D^b(X)$.*

PREUVES DU THÉORÈME 2.1 (esquisses) —

On se ramène facilement au cas où X et Y sont connexes (cf. proposition 2.4) et on fixe une équivalence $F : D^b(X) \xrightarrow{\sim} D^b(Y)$. On suppose ω_X ample (preuve identique dans l'autre cas). Le théorème résulterait immédiatement de l'invariance des algèbres canoniques, si on savait que ω_Y était ample (cf. §3.2.2).

• 1ère approche [7]. Sur une variété projective lisse connexe Z , pour tout point z et tout $i \in \mathbf{Z}$, les $C = \mathcal{O}_z[i] \in D^b(Z)^{(1)}$ vérifient

$$(2) \quad S(C) \simeq C[\dim Z], \quad \text{End}_{D^b(Z)}(C) = \mathbf{C} \quad \text{et} \quad \text{Hom}(C, C[i]) = 0 \quad \text{pour} \quad i < 0.$$

Sur la variété X où ω_X est ample, les conditions (2) caractérisent les objets $\mathcal{O}_x[i]$ dans $D^b(X)$. On en déduit que l'ensemble $\{F(\mathcal{O}_x)[i]\}_{i,x}$ contient les $\mathcal{O}_y[j]$ pour $y \in Y$

⁽¹⁾ \mathcal{O}_z est le faisceau gratte-ciel en z

et $j \in \mathbf{Z}$. Si $F(\mathcal{O}_x)$ n'est pas de cette forme, il est orthogonal aux $\mathcal{O}_y[j]$, donc il est nul. On déduit alors que F envoie tout $\mathcal{O}_x[i]$ sur un $\mathcal{O}_y[j]$ et ceci induit une bijection entre points de X et de Y . On caractérise ensuite les faisceaux inversibles décalés sur une variété lisse Z comme les $C \in D^b(Z)$ tels que pour tout $z \in Z$, il existe $n \in \mathbf{Z}$ tel que

$$\mathrm{Hom}(L, \mathcal{O}_z[n]) \simeq \mathbf{C} \text{ et } \mathrm{Hom}(L, \mathcal{O}_z[i]) = 0 \text{ pour } i \neq n.$$

On en déduit que F envoie un faisceau inversible sur un faisceau inversible décalé. Soit $L \in \mathrm{Pic}(X)$. Quitte à décaler F , alors on peut supposer $F(L) \in \mathrm{Pic}(Y)$. L'algèbre $\bigoplus_{i \geq 0} \mathrm{Hom}(L, S^i(L)[-i \dim X])$ est isomorphe à l'algèbre canonique de X et les ouverts définis par ses éléments forment une base de la topologie de X . Cette algèbre est isomorphe à l'algèbre définie de la même façon pour Y et elle donne donc une base de la topologie de Y . Ceci montre que ω_Y est ample et que les algèbres canoniques de X et Y sont isomorphes.

- 2ème approche [25, §4]. On commence comme ci-dessus par vérifier que les $\mathcal{O}_x[i]$ s'envoient sur des $\mathcal{O}_y[j]$. La suite de la preuve n'utilise plus que ω_X est ample. On utilise le théorème 3.7 plus bas qui affirme qu'il existe $K \in D^b(Y \times X)$ tel que $F = \Phi_K$. Alors, le lemme 3.1 plus bas montre que $Y \simeq X$.

- 3ème approche [51, §3.2.4]. Soit \mathcal{I} une sous-catégorie épaisse de $D^b(X)$. Si \mathcal{I} est stable par $L^{-1} \otimes -$ pour un faisceau ample L , alors elle est \otimes -idéale. Cette propriété est donc équivalente à la stabilité sous S^{-1} . Par conséquent, l'ensemble des fermés de X se retrouve à partir de $D^b(X)$ (à partir de sa seule structure triangulée). Pour tout fermé Z de Y , il existe donc un fermé Z' de X tel que $F(D_{Z'}^b(X)) = D_Z^b(Y)$. On montre que cette injection de l'ensemble des fermés de Y vers ceux de X se restreint en une bijection $Y \rightarrow X$ d'inverse continu. On identifie enfin les faisceaux d'anneaux. \square

Bondal et Orlov [7, Theorem 3.1] déterminent le groupe $\mathrm{Aut}(D^b(X))$ des classes d'isomorphisme d'auto-équivalences de $D^b(X)$ lorsque ω_X^\pm est ample (ceci est par exemple fourni par la deuxième preuve du théorème 2.1) :

$$\mathrm{Aut}(D^b(X)) = \mathrm{Pic}(X) \rtimes \mathrm{Aut}(X) \times \mathbf{Z}.$$

2.2. Décompositions semi-orthogonales

2.2.1. *Décompositions partielles.* — Considérons la forme d'Euler sur la K -théorie $K_0(X)$.

$$\langle [\mathcal{F}], [\mathcal{G}] \rangle = \sum_{i \geq 0} (-1)^i \dim \mathrm{Ext}^i(\mathcal{F}, \mathcal{G}).$$

Nous allons décrire l'analogie, pour la catégorie dérivée, d'une base triangulaire pour cette forme, ou plus généralement d'une décomposition semi-orthogonale de $K_0(X)$.

Soit \mathcal{I} une sous-catégorie épaisse d'une catégorie triangulée \mathcal{T} . On pose ${}^\perp \mathcal{I} = \{C \in \mathcal{T} \mid \mathrm{Hom}(C, I) = 0 \text{ pour tout } I \in \mathcal{I}\}$ et $\mathcal{I}^\perp = \{C \in \mathcal{T} \mid \mathrm{Hom}(I, C) = 0 \text{ pour tout } I \in \mathcal{I}\}$. On dit que $\langle \mathcal{I}^\perp, \mathcal{I} \rangle$ est une *décomposition semi-orthogonale* de \mathcal{T} lorsque pour tout objet C de \mathcal{T} , il existe un triangle distingué $C_1 \rightarrow C \rightarrow C_2 \rightsquigarrow$ avec $C_1 \in \mathcal{I}$ et $C_2 \in \mathcal{I}^\perp$. Ceci

revient à demander que le foncteur canonique $\mathcal{I}^\perp \rightarrow \mathcal{T}/\mathcal{I}$ soit une équivalence ou à demander que le foncteur d'inclusion $\mathcal{I} \rightarrow \mathcal{T}$ ait un adjoint à droite.

Lorsque $\mathcal{I}^\perp = \langle \mathcal{K}, \mathcal{J} \rangle$, on écrit $\mathcal{T} = \langle \mathcal{K}, \mathcal{J}, \mathcal{I} \rangle$ et on généralise aux décompositions $\mathcal{T} = \langle \mathcal{I}_1, \dots, \mathcal{I}_m \rangle$.

L'existence d'un « générateur fort » pour $D^b(X)$ fournit un théorème de représentabilité à la Brown pour les foncteurs cohomologiques sur $D^b(X)$ (cf. [5] et [9]) et on obtient un théorème général d'existence de décompositions :

THÉORÈME 2.3 (Bondal, Kapranov, Van den Bergh). — *Soient X une variété projective lisse et $\mathcal{I} = D^b(X)$ une sous-catégorie triangulée pleine d'une catégorie triangulée \mathcal{T} . Alors, le foncteur d'inclusion $\mathcal{I} \rightarrow \mathcal{T}$ a des adjoints à gauche et à droite, i.e., on a des décompositions semi-orthogonales $\mathcal{T} = \langle \mathcal{I}^\perp, \mathcal{I} \rangle$ et $\mathcal{T} = \langle \mathcal{I}, {}^\perp\mathcal{I} \rangle$.*

Une décomposition orthogonale de $D^b(X)$ correspond à une décomposition de X en union de composantes connexes :

PROPOSITION 2.4. — *Soit X une variété connexe. Soient \mathcal{I}_1 et \mathcal{I}_2 deux sous-catégories épaisses de $D^b(X)$ telles que $D^b(X) = \mathcal{I}_1 \oplus \mathcal{I}_2$ (i.e., $D^b(X) = \langle \mathcal{I}_1, \mathcal{I}_2 \rangle = \langle \mathcal{I}_2, \mathcal{I}_1 \rangle$). Alors, $\mathcal{I}_1 = 0$ ou $\mathcal{I}_2 = 0$.*

PREUVE — Un objet indécomposable de $D^b(X)$ est dans \mathcal{I}_1 ou dans \mathcal{I}_2 . Soient r, s tels que $\mathcal{O}_X \in \mathcal{I}_r$ et $\{r, s\} = \{1, 2\}$. Soit $X_i = \{x \in X \mid \mathcal{O}_x \in \mathcal{I}_i\}$. Si $x \in X_s$, alors $\text{Hom}(\mathcal{O}_X, \mathcal{O}_x) = 0$, ce qui est impossible, donc $X_s = \emptyset$. Si $C \in \mathcal{I}_s$, alors $\text{Hom}(C, \mathcal{O}_x[i]) = 0$ pour tout $x \in X$ et tout $i \in \mathbf{Z}$, donc $C = 0$. \square

Remarque 2.5. — Soit \mathcal{T} une catégorie triangulée avec un foncteur de Serre S et $\mathcal{T} = \langle \mathcal{I}, \mathcal{I}^\perp \rangle$ une décomposition semi-orthogonale. Alors, $\mathcal{T} = \langle {}^\perp\mathcal{I}, \mathcal{I} \rangle$ et ${}^\perp\mathcal{I} = S^{-1}(\mathcal{I}^\perp)$.

Considérons en particulier $\mathcal{T} = D^b(X)$ où X est une variété projective lisse connexe de Calabi-Yau. Alors, il n'y a pas de décomposition semi-orthogonale non triviale de $D^b(X)$, car une telle décomposition serait une décomposition orthogonale.

2.2.2. Décompositions complètes. — Voyons le cas particulier d'une suite exceptionnelle d'objets. C'est une suite (C_1, \dots, C_m) d'objets de \mathcal{T} telle que

- $\text{Hom}(C_i, C_j[r]) = 0$ pour $r \in \mathbf{Z}$ et $i > j$
- $\text{Hom}(C_i, C_i[r]) = 0$ pour $r \neq 0$
- $\text{End}(C_i) = \mathbf{C}$.

On dit que la suite est *complète* si elle engendre \mathcal{T} .

Soit (C_1, \dots, C_m) une suite exceptionnelle complète. Notons \mathcal{I}_i la sous-catégorie triangulée de \mathcal{T} engendrée par C_i . Nous noterons $D^b(A)$ la catégorie dérivée bornée des modules de type fini sur une algèbre A .

On a une équivalence $C_i \otimes_{\mathbf{C}} - : D^b(\mathbf{C}) \xrightarrow{\sim} \mathcal{I}_i$ et une décomposition semi-orthogonale $\mathcal{T} = \langle \mathcal{I}_1, \dots, \mathcal{I}_m \rangle$. Réciproquement, toute décomposition semi-orthogonale en des catégories équivalentes à $D^b(\mathbf{C})$ provient d'une suite exceptionnelle d'objets. L'ensemble $\{[C_i]\}$ forme une base de $K_0(\mathcal{T})$. Si \mathcal{T} est localement de type fini (i.e., si

$\dim \bigoplus_i \text{Hom}(M, N[i]) < \infty$ pour tous $M, N \in \mathcal{T}$, alors la matrice de la forme d'Euler dans la base $\{[C_i]\}$ est triangulaire.

Exemple 2.6. — Soit $X = \mathbf{P}^n$. Beilinson [4] montre que $(\mathcal{O}(-n), \mathcal{O}(-n+1), \dots, \mathcal{O}(0))$ est une suite exceptionnelle complète. L'orthogonalité est claire. La résolution de la diagonale $\Delta \subset \mathbf{P}^n \times \mathbf{P}^n$:

$$0 \rightarrow \mathcal{O}(-n) \boxtimes \Omega^n(n) \rightarrow \dots \rightarrow \mathcal{O}(-1) \boxtimes \Omega^1(1) \rightarrow \mathcal{O} \boxtimes \mathcal{O} \rightarrow \mathcal{O}_\Delta \rightarrow 0$$

décrit le foncteur identité de $D^b(\mathbf{P}^n)$ comme extension de foncteurs $\mathcal{O}(-i) \otimes H^*(\Omega^i(i) \otimes -)$ et ceci démontre l'engendrement. Soient $\mathcal{F} = \bigoplus_{i=0}^n \mathcal{O}(-i)$ et $A = \text{End}(\mathcal{F})$, une algèbre de dimension finie. Alors, on a en plus ici $\text{Ext}^{>0}(\mathcal{F}, \mathcal{F}) = 0$ et on déduit que le foncteur $R\text{Hom}(\mathcal{F}, -) : D^b(\mathbf{P}^n) \rightarrow D^b(A)$ est une équivalence.

Nous renvoyons le lecteur à [24] pour un article de synthèse, consacré en particulier aux espaces projectifs et aux surfaces de Del Pezzo.

2.2.3. Minimisation. — Kapranov a construit des suites exceptionnelles complètes pour les quadriques projectives lisses et les variétés de drapeaux de type A [27] (cf. [53] pour un survol des constructions pour les variétés de Fano).

King [34, Conjecture 9.3] conjecture que toute variété torique complète lisse X admet une suite (L_1, \dots, L_n) de fibrés en droites telle que $\text{Ext}^{>0}(L_i, L_j) = 0$ pour tous i, j et les L_i engendrent $D^b(X)$ (alors, $D^b(X)$ est équivalente à $D^b(A)$, où $A = \text{End}(\bigoplus_i L_i)$). Kawamata [33] démontre l'existence d'une suite exceptionnelle complète de faisceaux pour toute variété torique projective lisse.

Dès que $K_0(X)$ n'est pas de type fini, $D^b(X)$ ne peut être équivalente à la catégorie dérivée d'une algèbre de dimension finie. On montre par contre que pour toute variété X , il existe une dg-algèbre (=algèbre différentielle graduée) A dont la catégorie des complexes parfaits est équivalente à $D^b(X)$ (cf. Keller, Thomason, Neeman, Kontsevich, Bondal-Van den Bergh, [52, Proposition 3.14 et Theorem 7.39]).

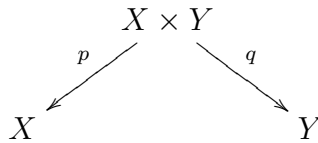
La recherche de modèles minimaux est reliée à la minimisation de la catégorie dérivée. On cherche une suite exceptionnelle L_1, \dots, L_n avec n maximal. Soit \mathcal{I} la sous-catégorie triangulée de $D^b(X)$ engendrée par les L_i . Alors, $D^b(X) = \langle \mathcal{I}, \mathcal{I}^\perp \rangle$ et la géométrie de X devrait être en partie contrôlée par la catégorie triangulée $\mathcal{T} = \mathcal{I}^\perp$. Il serait intéressant d'étudier l'indépendance de \mathcal{T} du choix de L_1, \dots, L_n et même son indépendance bi-rationnelle (cf. [37]). La catégorie \mathcal{T} apparaît parfois comme la catégorie dérivée d'une variété X' , de dimension inférieure ou égale (cf. [38] pour un exemple de variété de Fano X de dimension 3 où X' est une courbe de genre 7). L'exemple le plus simple est celui d'un éclatement de centre un espace projectif (cf. théorème 4.2).

3. COMPARAISONS

3.1. Transformations à noyau

3.1.1. *Définition.* — L'idée des transformations à noyau est la suivante : on se donne une fonction $\phi : X \times Y \rightarrow \mathbf{C}$. On a alors une application des fonctions sur Y vers les fonctions sur X donnée par $f \mapsto (x \mapsto \int_Y f(y)\phi(x, y)dy)$.

Cette construction a un analogue pour les faisceaux cohérents (les mêmes constructions pour les faisceaux constructibles ou les \mathcal{D} -modules sont classiques). Soient X et Y deux variétés projectives lisses et $p : X \times Y \rightarrow X$, $q : X \times Y \rightarrow Y$ les deux projections.

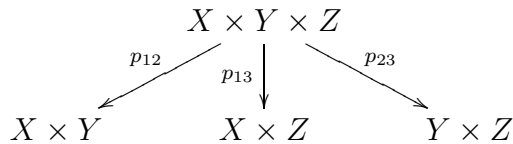


Soit $K \in D^b(X \times Y)$. On définit alors le foncteur (dit de « Fourier-Mukai ») $\Phi_K : D^b(Y) \rightarrow D^b(X)$ par

$$\Phi_K(C) = \mathbf{R}p_*(K \otimes^{\mathbf{L}} q^*C).$$

Soit $K^\vee = \mathbf{R}\mathcal{H}om(K, \mathcal{O}_{X \times Y}) \in D^b(Y \times X)$. Si X et Y sont de dimension pure, alors les foncteurs $\Phi_{K^\vee \otimes p^*\omega_X[\dim X]}$ et $\Phi_{K^\vee \otimes q^*\omega_Y[\dim Y]}$ sont respectivement adjoints à gauche et à droite de Φ_K .

Soient Z une autre variété projective lisse et $L \in D^b(Y \times Z)$.



On pose $K \circ L = \mathbf{R}p_{13*}(p_{12}^*K \otimes^{\mathbf{L}} p_{23}^*L)$. On a alors un isomorphisme canonique $\Phi_K \circ \Phi_L \xrightarrow{\sim} \Phi_{K \circ L}$.

Le lemme classique suivant permet de reconnaître quand K provient d'un isomorphisme de variétés (cf. [25, Corollary 4.3]).

LEMME 3.1. — *On suppose que Y est connexe et que pour tout $y \in Y$, il existe $x \in X$ et $n \in \mathbf{Z}$ tels que $\Phi_K(\mathcal{O}_y) \simeq \mathcal{O}_x[n]$. Alors, il existe un morphisme $\sigma : Y \rightarrow X$ de graphe Γ_σ et il existe $L \in \text{Pic}(Y)$ et $m \in \mathbf{Z}$ tels que $K \simeq \mathcal{O}_{\Gamma_\sigma} \otimes q^*L[m]$.*

Si Φ_K est une équivalence, alors σ est un isomorphisme.

PREUVE (esquisse) — Soit $y \in Y$ d'anneau local $\mathcal{O}_{\mathfrak{m}_y}$ et soit $n \in \mathbf{Z}$ tel que $K \otimes_{\mathcal{O}_Y}^{\mathbf{L}} \mathcal{O}_y$ est concentré en degré $-n$. Le lemme de Nakayama montre que $q_*(K \otimes_{\mathcal{O}_Y} \mathcal{O}_{\mathfrak{m}_y}) \simeq \mathcal{O}_{\mathfrak{m}_y}[n]$. Il existe alors un voisinage ouvert U de y tel que $q_*(K \otimes_{\mathcal{O}_Y} \mathcal{O}_U) \simeq \mathcal{O}_U[n]$ et on obtient un morphisme $U \rightarrow X$. Ceux-ci se recollent en $\sigma : Y \rightarrow X$ avec les propriétés voulues. Si Φ_K est une équivalence, alors $K^\vee \otimes p^*\omega_X[\dim X]$ définit un morphisme $X \rightarrow Y$ inverse de σ . \square

3.1.2. *Pleine fidélité.* — La pleine fidélité du foncteur Φ_K peut se tester sur une famille d'objets appropriée (cf. [11, Theorems 5.1 and 5.4]). Nous utilisons ici les faisceaux gratte-ciel [6, Theorem 1.1], le point-clef étant qu'un objet orthogonal (à gauche ou à droite) aux faisceaux gratte-ciel et à leurs décalés est nul.

PROPOSITION 3.2 (Bondal-Orlov). — *Soit $K \in D^b(X \times Y)$. Le foncteur Φ_K est pleinement fidèle si et seulement si pour tous $y, y' \in Y$, on a*

$$\mathrm{Hom}(\Phi_K(\mathcal{O}_y), \Phi_K(\mathcal{O}_{y'})[i]) = \begin{cases} 0 & \text{sauf si } y = y' \text{ et } 0 \leq i \leq \dim Y \\ k & \text{si } y = y' \text{ et } i = 0. \end{cases}$$

C'est une équivalence si en plus $\Phi_K(\mathcal{O}_y) \otimes \omega_X \simeq \Phi_K(\mathcal{O}_y)$ pour tout $y \in Y$.

Cette proposition montre que lorsque X et Y ont des fermés stricts Z et Z' en dehors desquels K est le faisceau de structure du graphe d'un isomorphisme $Y - Z' \xrightarrow{\sim} X - Z$, alors le critère précédent peut se vérifier en remplaçant X et Y par leurs complétés formels le long de Z et Z' . Ceci permet de substituer à X et Y des modèles préférés, à condition de garder les mêmes complétés formels (cf. la preuve du théorème 4.6).

Un cas particulier de pleine fidélité est fourni par le résultat suivant, qui se déduit immédiatement de la formule de projection.

PROPOSITION 3.3. — *Soit $f : V \rightarrow W$ un morphisme entre variétés projectives lisses. Si le morphisme canonique $\mathcal{O}_W \rightarrow \mathbf{R}f_*\mathcal{O}_V$ est un isomorphisme, alors $\mathbf{L}f^* : D^b(W) \rightarrow D^b(V)$ est pleinement fidèle.*

3.1.3. *Familles.* — Soient X', Y' deux variétés projectives lisses et $K' \in D^b(X' \times Y')$. Alors, on a le résultat classique (cf. [48, Proposition 2.1.7]) :

PROPOSITION 3.4. — *Si $\Phi_K : D^b(Y) \rightarrow D^b(X)$ et $\Phi_{K'} : D^b(Y') \rightarrow D^b(X')$ sont pleinement fidèles (resp. sont des équivalences), alors $\Phi_{K \boxtimes K'} : D^b(Y \times Y') \rightarrow D^b(X \times X')$ est pleinement fidèle (resp. est une équivalence).*

Soient $p : X \rightarrow S$ et $q : Y \rightarrow S$ des morphismes projectifs lisses entre variétés projectives lisses. Soient $s_0 \in S$, $X_0 = p^{-1}(s_0)$ et $Y_0 = q^{-1}(s_0)$. Soient $i : X_0 \rightarrow X$, $j : Y_0 \rightarrow Y$ et $k : X \times_S Y \rightarrow X \times Y$ les immersions fermées. Les propriétés d'une famille de noyaux se spécialisent [19, Proposition 6.2] :

PROPOSITION 3.5 (Chen). — *Soit $K \in D^b(X \times_S Y)$ tel que $\Phi_{k_*K} : D^b(Y) \rightarrow D^b(X)$ soit pleinement fidèle (resp. soit une équivalence). Alors, $\Phi_{\mathbf{L}(i \times j)^*K} : D^b(Y_0) \rightarrow D^b(X_0)$ est pleinement fidèle (resp. est une équivalence).*

Ce résultat permet de vérifier dans certains cas qu'un noyau donne une équivalence par déformation (cf. par exemple §4.4).

Remarque 3.6. — Soit $\Phi_K : D^b(Y) \rightarrow D^b(X)$ un foncteur pleinement fidèle. Alors, on doit penser à Y comme l'espace de modules fin de $\{F(\mathcal{O}_y)\}_{y \in Y}$ et à K comme l'objet universel associé.

3.1.4. *Représentabilité.* — L'imperfection des axiomes des catégories triangulées rend la preuve du résultat suivant délicate (cf. [46], [48, Theorem 3.2.1] et [9, Theorem 1.1] qui assure l'existence d'adjoints ; cf. [31, Theorem 1.1] pour une extension aux champs de Deligne-Mumford et une autre preuve).

THÉORÈME 3.7 (Orlov). — *Soit $F : D^b(Y) \rightarrow D^b(X)$ un foncteur pleinement fidèle. Alors, il existe un unique $K \in D^b(X \times Y)$ tel que $F \simeq \Phi_K$.*

Une approche préférable à ce problème (et à ceux de §3.2) consiste à considérer une structure plus riche que celle de catégorie triangulée, celle de dg-catégorie (les Hom sont munis d'une structure de complexe d'espaces vectoriels) dont le « H^0 » est la catégorie dérivée de départ [59]. On dispose d'une dg-catégorie $L_{\text{coh}}(X)$ dont le « H^0 » est $D^b(X)$. Toen montre que la dg-catégorie des foncteurs de $L_{\text{coh}}(Y)$ vers $L_{\text{coh}}(X)$ est (quasi-)équivalente à $L_{\text{coh}}(X \times Y)$ [59, Theorem 8.15] et ceci fournit un analogue du théorème 3.7, pour des foncteurs non nécessairement pleinement fidèles. Les foncteurs $D^b(Y) \rightarrow D^b(X)$ obtenus sont alors tous du type Φ_K et réciproquement tout foncteur de ce type provient d'un foncteur défini au niveau des dg-catégories.

3.2. Invariants d'une équivalence

3.2.1. Soit $F : D^b(Y) \xrightarrow{\sim} D^b(X)$ une équivalence, avec X et Y projectives lisses connexes. D'après le théorème 3.7, il existe $K \in D^b(X \times Y)$ tel que $F \simeq \Phi_K$.

La commutation de F avec les foncteurs de Serre montre que $\dim X = \dim Y$ et que ω_X et ω_Y ont le même ordre [14, Lemma 2.1].

Un argument de rigidité montre que F induit un isomorphisme de groupes algébriques $\text{Pic}^0(Y) \rtimes \text{Aut}^0(Y) \xrightarrow{\sim} \text{Pic}^0(X) \rtimes \text{Aut}^0(X)$, où $\text{Aut}^0(X)$ est la composante neutre de $\text{Aut}(X)$ (cf. théorème 3.11 pour un cas où les deux facteurs sont échangés).

3.2.2. *Cohomologie.* — Passons maintenant à des invariants du type cohomologie ou algèbre canonique (cf. [42, Theorem 4.9], [16], [17], [29, Theorem 2.3] et [48, Theorem 2.1.8]).

Soit $HA_{i,k}(X) = \text{Ext}_{X \times X}^i(\mathcal{O}_{\Delta X}, i_*\omega_X^k)$ où $i : \Delta X \rightarrow X \times X$ est l'inclusion de la diagonale. Soit $HA(X) = \bigoplus_{i,k} HA_{i,k}(X)$. On munit $HA(X)$ d'une structure d'algèbre bigraduée via les isomorphismes canoniques $\text{Ext}_{X \times X}^i(i_*\omega_X^r, i_*\omega_X^s) \xrightarrow{\sim} \text{Ext}_{X \times X}^i(\mathcal{O}_{\Delta X}, i_*\omega_X^{s-r})$.

On a $HA_{i,k}(X) = \llbracket \text{Hom}(S_X, S_X^k[i-k \dim X]) \rrbracket$, où le terme de droite doit être compris comme le H^0 d'un complexe de Hom's pris au niveau des dg-catégories.

On a $HA_{i,k}(X) \simeq \bigoplus_{p+q=i} H^p(X, \Lambda^q \mathcal{T}_X \otimes \omega_X^k)$ (cf. [36] et [54, Corollary 2.6]), où \mathcal{T}_X est le fibré tangent. En particulier, $\bigoplus_{k \geq 0} HA_{0,k}(X)$ est isomorphe à l'algèbre canonique $R(X) = \bigoplus_{k \geq 0} H^0(X, \omega_X^k)$.

THÉORÈME 3.8. — *F induit un isomorphisme d'algèbres graduées $HA(Y) \xrightarrow{\sim} HA(X)$. En particulier, F induit un isomorphisme gradué entre les algèbres canoniques $R(Y) \xrightarrow{\sim} R(X)$ et un isomorphisme entre les espaces vectoriels de cohomologie $H^*(Y, \mathbf{C}) \xrightarrow{\sim} H^*(X, \mathbf{C})$.*

PREUVE (esquisse) — On utilise l'équivalence $\Phi_{K\boxtimes L} : D^b(Y \times Y) \xrightarrow{\sim} D^b(X \times X)$ où $L = K^\vee \otimes p^*\omega_X[\dim X] \simeq K^\vee \otimes q^*\omega_Y[\dim Y]$ vu dans $D^b(X \times Y)$. \square

L'isomorphisme $H^*(X, \mathbf{C}) \xrightarrow{\sim} H^*(Y, \mathbf{C})$ n'est pas compatible à la multiplication ni à la graduation classique, en général. Il est par contre compatible à la graduation donnée par ${}^nH(X, \mathbf{C}) = \bigoplus_{p-q=n} H^p(X, \Omega_X^q)$.

Remarque 3.9. — On s'attend tout de même à l'égalité des nombres de Hodge de X et Y (cf. [15, §1.3]). Dans la situation de la conjecture 1.1, la formule de changement de variable pour l'intégration motivique montre que les variétés X et Y ont les mêmes nombres de Hodge (Kontsevich, Batyrev et Denef–Loeser), cf. [50, §4] et [39].

On déduit du théorème 3.8 l'invariance de la dimension de Kodaira. On déduit aussi que si $\kappa(X, \omega_X) = \dim X$ (i.e., X de type général) ou $\kappa(X, \omega_X^{-1}) = \dim X$, alors X et Y sont birationnelles.

Via le théorème de Grothendieck-Riemann-Roch, on obtient l'invariance de la cohomologie à coefficients rationnels.

Soit $L \in D^b(X \times Y)$. Soit $\text{ch} : K_0(X) \rightarrow H^*(X, \mathbf{Q})$ la classe de Chern et soit td_X la classe de Todd de X . On définit un morphisme

$$\phi_L : H^*(Y, \mathbf{Q}) \rightarrow H^*(X, \mathbf{Q}), \quad \xi \mapsto p_* \left(p^*(\sqrt{\text{td}_X}) \cdot \text{ch}([L]) \cdot q^*(\sqrt{\text{td}_Y}) \cdot q^*(\xi) \right).$$

Ce morphisme est gradué pour la graduation nH et on a $\phi_{L \circ L'} = \phi_L \circ \phi_{L'}$.

Lorsque $L = K$, alors ce morphisme est un isomorphisme et sa complexification est l'isomorphisme du théorème 3.8.

La transformation Φ_L induit un morphisme $[\Phi_L] : K_0(Y) \rightarrow K_0(X)$ et on a un diagramme commutatif

$$\begin{array}{ccc} K_0(Y) & \xrightarrow{[\Phi_L]} & K_0(X) \\ \text{ch}(-) \cdot \sqrt{\text{td}_Y} \downarrow & & \downarrow \text{ch}(-) \cdot \sqrt{\text{td}_X} \\ H^*(Y, \mathbf{Q}) & \xrightarrow{\phi_L} & H^*(X, \mathbf{Q}) \end{array}$$

Remarque 3.10. — Hille et Van den Bergh [25, Remark 3.4] mentionnent l'invariance de la K -théorie topologique par équivalence dérivée et en déduisent l'invariance de $H^*(X, \mathbf{Z})$ dans $H^*(X, \mathbf{Q})$ pour les courbes, les surfaces $K3$ et les variétés abéliennes.

3.3. Variétés abéliennes

Le résultat suivant de Mukai [41, Theorem 2.2] est le point de départ des travaux sur les équivalences entre catégories dérivées de faisceaux cohérents.

Soient A une variété abélienne et $\hat{A} = \text{Pic}^0(A)$ sa variété abélienne duale. On note \mathcal{P} le fibré de Poincaré sur $A \times \hat{A}$.

THÉORÈME 3.11 (Mukai). — *Le foncteur $\Phi_{\mathcal{P}} : D^b(\hat{A}\text{-coh}) \rightarrow D^b(A\text{-coh})$ est une équivalence.*

PREUVE — On vérifie les conditions de la proposition 3.2. Pour $x \in \hat{A}$, alors $\Phi_K(\mathcal{O}_x)$ est un fibré en droites L_x de degré 0 sur A . Puisque $H^*(L) = 0$ pour tout $L \in \text{Pic}^0(A)$ non trivial, on déduit que $\Phi_{\mathcal{P}}$ est pleinement fidèle, et donc une équivalence, car $\omega_A \simeq \mathcal{O}_A$. \square

On décrit maintenant toutes les équivalences dérivées entre variétés abéliennes (cf. [47], [48, §5] et [49, §11 et §15]).

Soit B une variété abélienne. Soit

$$f = \begin{pmatrix} x & y \\ z & t \end{pmatrix} : B \times \hat{B} \rightarrow A \times \hat{A}.$$

On pose

$$\tilde{f} = \begin{pmatrix} \hat{t} & -\hat{y} \\ -\hat{z} & \hat{x} \end{pmatrix} : A \times \hat{A} \rightarrow B \times \hat{B}.$$

On note $U(B \times \hat{B}, A \times \hat{A})$ l'ensemble des isomorphismes $f : B \times \hat{B} \xrightarrow{\sim} A \times \hat{A}$ tels que $f^{-1} = \tilde{f}$.

Soient $a \in A$ et $\alpha \in \hat{A}$. On note $m_a : A \rightarrow A$, $b \mapsto a + b$. On pose $\Phi_{a,\alpha} = L_\alpha \otimes m_{a*}(-) : D^b(A) \xrightarrow{\sim} D^b(A)$.

THÉORÈME 3.12 (Polishchuk, Orlov). — Soit $F : D^b(B) \xrightarrow{\sim} D^b(A)$ une équivalence. Alors, il existe $\sigma \in U(B \times \hat{B}, A \times \hat{A})$ tel que

$$(3) \quad \Phi_{\sigma(b,\beta)} = F \circ \Phi_{b,\beta} \circ F^{-1} \text{ pour tous } b \in B \text{ et } \beta \in \hat{B}.$$

Réciproquement, soit $\sigma \in U(B \times \hat{B}, A \times \hat{A})$. Alors, il existe une équivalence $F : D^b(B) \xrightarrow{\sim} D^b(A)$ vérifiant (3).

PREUVE (éléments) — L'invariance de $\text{Aut}^0 \times \text{Pic}^0$ (cf. §3.2) fournit un isomorphisme $B \times \hat{B} \xrightarrow{\sim} A \times \hat{A}$, dont on vérifie qu'il a la propriété voulue. La réciproque requiert la construction d'un fibré simple semi-homogène de pente donnée. \square

On obtient alors une description explicite du groupe des auto-équivalences. On a une suite exacte de groupes

$$0 \rightarrow (A \times \hat{A})(\mathbf{C}) \times \mathbf{Z} \rightarrow \text{Aut}(D^b(A)) \rightarrow U(A \times \hat{A}, A \times \hat{A}) \rightarrow 1.$$

Remarque 3.13. — On peut conjecturer qu'une variété projective lisse dérivé-équivalente à une variété abélienne est une variété abélienne.

3.4. Surfaces

Notons tout d'abord que le cas des courbes n'est pas intéressant ! Deux courbes projectives lisses sont dérivé-équivalentes si et seulement si elles sont isomorphes : en genre $\neq 1$, le théorème 2.1 donne le résultat. Pour les courbes elliptiques, on le déduit par exemple de l'invariance de la structure de Hodge entière.

Nous décrivons brièvement la situation pour les surfaces. Les démonstrations demandent une analyse minutieuse suivant la classification des surfaces minimales.

Soient X et Y deux surfaces projectives lisses connexes non isomorphes. On suppose que X est minimale, *i.e.*, ne contient pas de \mathbf{P}^1 avec auto-intersection -1 . On a une description précise des cas d'équivalences dérivées [15] (cf. [42, 46] pour les $K3$).

THÉORÈME 3.14 (Bridgeland-Maciocia). — *On a $D^b(X) \simeq D^b(Y)$ si et seulement si une des assertions suivantes est vraie*

- X et Y sont toutes deux abéliennes (ou toutes deux des $K3$) et il existe une isométrie entre leurs réseaux transcendants compatible avec les structures de Hodge
- X et Y sont des surfaces elliptiques et Y est un schéma de Picard relatif de la fibration elliptique de X [10].

Remarque 3.15. — La K -équivalence entre deux variétés projectives lisses X et Y (= existence d'un diagramme (1) avec $f^*K_X \sim g^*K_Y$) implique que les variétés sont isomorphes en codimension 1. Si deux surfaces sont K -équivalentes, elles sont donc isomorphes. Uehara [60] construit des exemples de surfaces elliptiques birationnelles et dérivé-équivalentes mais qui ne sont pas isomorphes, donc pas K -équivalentes non plus.

4. FLIPS ET FLOPS

4.1. Introduction

Un cas particulier de la conjecture 1.1 est celui d'un flop. Ce cas est important car une des conjectures du MMP est que deux modèles minimaux birationnels sont connectés par une suite de flops.

Un flop est un diagramme

$$\begin{array}{ccc} X & & X^+ \\ & \searrow f & \swarrow f^+ \\ & \bar{X} & \end{array}$$

où

- \bar{X} est une variété projective de Gorenstein,
- f et f^+ sont des résolutions crépantes (*i.e.*, $f^*\omega_{\bar{X}} \simeq \omega_X$ et $(f^+)^*\omega_{\bar{X}} \simeq \omega_{X^+}$) dont le lieu exceptionnel est de codimension ≥ 2 et
- il existe un diviseur D sur X tel que $-D$ est relativement f -ample et le transformé strict de D est relativement f^+ -ample.

Le morphisme f détermine uniquement f^+ (car $X^+ = \text{Proj} \bigoplus_{m \geq 0} \mathcal{O}_{\bar{X}}(f_*(mD))$, indépendant de D). La conjecture 1.1 prédit que $D^b(X) \simeq D^b(X^+)$ et donc que X^+ peut se construire comme un espace de module d'objets de $D^b(X)$ (cf. §4.5 pour la dimension 3). Cette approche pourrait aussi s'appliquer pour les flips.

Nous allons voir des exemples (§4.3, §4.4.1 et §4.5) où la transformation de noyau $\mathcal{O}_{X \times_{\bar{X}} X^+}$ est une équivalence $D^b(X^+) \xrightarrow{\sim} D^b(X)$. Néanmoins, nous verrons dans §4.4.2 une situation où cette transformation n'est pas une équivalence.

4.2. Éclatements

4.2.1. *Fibrés projectifs.* — Soit E un fibré vectoriel de rang $r \geq 1$ sur une variété projective lisse Y et soit $q : \mathbf{P}(E) \rightarrow Y$ le fibré projectif associé.

La proposition suivante [45, Theorem 2.6] fournit une version relative de l'exemple 2.6.

PROPOSITION 4.1 (Orlov). — *On a une décomposition semi-orthogonale*

$$D^b(\mathbf{P}(E)) = \langle D^b(Y)_{-r+1}, D^b(Y)_{-r+2}, \dots, D^b(Y)_0 \rangle,$$

où $D^b(Y)_d$ est l'image de $D^b(Y)$ par le foncteur pleinement fidèle $\mathcal{O}_q(d) \otimes q^*(-)$.

PREUVE (esquisse) — La pleine fidélité est fournie par la proposition 3.3. Soient $C, D \in D^b(Y)$ et $L \in D^b(\mathbf{P}(E))$. On a

$$\mathrm{Hom}(q^*C, L \otimes^{\mathbf{L}} q^*D) \simeq \mathrm{Hom}(C, \mathbf{R}q_*(L \otimes^{\mathbf{L}} q^*D)) \simeq \mathrm{Hom}(C, (\mathbf{R}q_*L) \otimes^{\mathbf{L}} D).$$

La semi-orthogonalité résulte alors de l'annulation de $\mathbf{R}q_*\mathcal{O}_q(i)$ pour $-1 \geq i \geq -r + 1$.

Soit \mathcal{T} la sous-catégorie triangulée de $D^b(\mathbf{P}(E))$ engendrée par $D^b(Y)_{-r+1}, \dots, D^b(Y)_0$. Soient $y \in Y$ et $F = q^{-1}(y)$. Les faisceaux $\mathcal{O}_F(-r + 1), \dots, \mathcal{O}_F(0)$ engendrent $D^b(F)$ (exemple 2.6), donc, vus comme faisceaux sur $\mathbf{P}(E)$, ils engendrent $D^b_F(\mathbf{P}(E))$. Par conséquent, $D^b_F(\mathbf{P}(E)) \subset \mathcal{T}$ et en particulier $\mathcal{O}_x \in \mathcal{T}$ pour $x \in F$. On en déduit que ${}^{\perp}\mathcal{T} = 0$, d'où $\mathcal{T} = D^b(\mathbf{P}(E))$ par le théorème 2.3.□

4.2.2. *Décomposition pour un éclatement.* — Soient maintenant X une variété projective lisse et Y une sous-variété fermée lisse de X purement de codimension $r \geq 2$. Soit $\mathcal{N}_{Y/X}$ le fibré normal de Y dans X . Soient \tilde{X} l'éclatée de X le long de Y et $\tilde{Y} \simeq \mathbf{P}(\mathcal{N}_{Y/X})$ le diviseur exceptionnel, image inverse de Y dans \tilde{X} . On a un diagramme commutatif

$$\begin{array}{ccc} \tilde{Y} & \xrightarrow{j} & \tilde{X} \\ q \downarrow & & \downarrow p \\ Y & \xrightarrow{i} & X \end{array}$$

Le théorème suivant décrit comment la catégorie dérivée grossit par éclatement [45, Theorem 4.3].

THÉORÈME 4.2 (Orlov). — *On a une décomposition semi-orthogonale*

$$D^b(\tilde{X}) = \langle D^b(Y)_{-r+1}, D^b(Y)_{-r+2}, \dots, D^b(Y)_{-1}, D^b(X)_0 \rangle$$

où $D^b(Y)_d = j_*(\mathcal{O}_q(d) \otimes q^*D^b(Y))$ et $D^b(X)_0 = \mathbf{L}p^*D^b(X)$. Les foncteurs canoniques induisent des équivalences $D^b(X) \xrightarrow{\sim} D^b(X)_0$ et $D^b(Y) \xrightarrow{\sim} D^b(Y)_d$ pour tout d .

PREUVE (esquisse) — Un point utile est l’existence, pour tout $C \in D^b(\tilde{Y})$, d’un triangle distingué

$$(4) \quad C \otimes \mathcal{O}_q(1) \rightarrow \mathbf{L}j^*j_*C \rightarrow C \rightsquigarrow .$$

La proposition 3.3 montre que $\mathbf{L}p^* : D^b(X) \rightarrow D^b(\tilde{X})$ est pleinement fidèle. Pour l’équivalence $D^b(Y) \xrightarrow{\sim} D^b(Y)_d$, la proposition 4.1 ramène le problème à montrer que $\text{Hom}(\mathbf{L}j^*j_*\mathcal{O}_F, \mathcal{O}_F[i]) = 0$ pour toute fibre F de q et tout $i \neq 0$. Le triangle distingué (4) ramène l’annulation recherchée à celles de $H^*(\mathbf{P}^{r-1}, \mathcal{O}(-1))$ et $H^{>0}(\mathbf{P}^{r-1}, \mathcal{O})$.

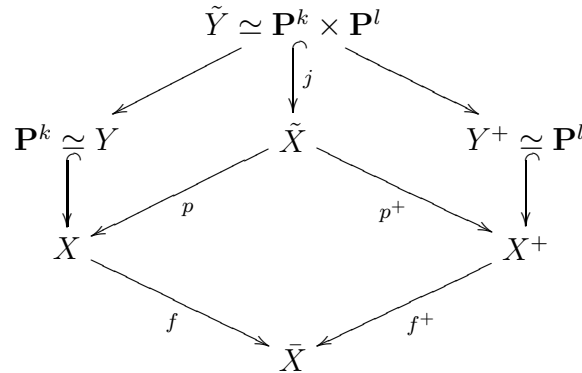
La semi-orthogonalité de $\mathcal{T} = \langle D^b(Y)_{-r+1}, D^b(Y)_{-r+2}, \dots, D^b(Y)_{-1}, D^b(X)_0 \rangle$ s’établit par la même technique que dans la preuve de la proposition 4.1.

Il reste à établir que ${}^\perp\mathcal{T}$ est nul. Soit $y \in Y$. On a $\mathbf{L}^d p^* \mathcal{O}_y \simeq \Omega_{\mathbf{P}^{r-1}(y)}^d(d)$ pour $0 \leq d \leq r-1$ et $\mathbf{L}^d p^* \mathcal{O}_y = 0$ pour $d \geq r$. On a $H^*(\mathbf{P}^{r-1}, \Omega^d(d)) = 0$ pour $1 \leq d \leq r-1$ et la résolution de la diagonale dans \mathbf{P}^{r-1} (cf. l’exemple 2.6) montre que, pour $1 \leq d \leq r-1$, $\Omega^d(d)$ est dans la sous-catégorie de $D^b(\mathbf{P}^{r-1})$ engendrée par $\mathcal{O}(-r+1), \dots, \mathcal{O}(-1)$. Par conséquent, le cône du morphisme canonique $\mathbf{L}p^* \mathcal{O}_y \rightarrow \mathcal{O}_{p^{-1}(y)}$ est dans \mathcal{T} , donc finalement $\mathcal{O}_{p^{-1}(y)}$ est dans \mathcal{T} et on termine comme pour la proposition 4.1. \square

Remarque 4.3. — Les arguments essentiels pour les deux résultats précédents apparaissent aussi dans les travaux de Thomason [56, 57] qui s’intéresse à la K -théorie supérieure.

4.3. Flips et flops standard

Soient X une variété projective lisse et Y une sous-variété fermée isomorphe à \mathbf{P}^k telle que $\mathcal{N}_{Y/X} \simeq \mathcal{O}_Y(-1)^{l+1}$ avec $1 \leq l \leq k$. Soit $p : \tilde{X} \rightarrow X$ l’éclatée de X le long de Y . Le diviseur exceptionnel \tilde{Y} est isomorphe à $\mathbf{P}^k \times \mathbf{P}^l$ et on a $\mathcal{N}_{\tilde{Y}/\tilde{X}} \simeq \mathcal{O}(-1, -1)$. Soit $p^+ : \tilde{X} \rightarrow X^+$ la contraction de \tilde{Y} sur \mathbf{P}^l . On suppose que X^+ est une variété projective. Soient $f : X \rightarrow \bar{X}$ et $f^+ : X^+ \rightarrow \bar{X}$ les contractions de \mathbf{P}^k et \mathbf{P}^l sur un point. Ils fournissent l’exemple le plus simple de flip (et de flop lorsque $k = l$). On a $\tilde{X} \simeq X \times_{\bar{X}} X^+$.



Les conjectures 1.1 et 1.2 sont connues dans ce cas (cf. [6, Theorem 3.6] et [48, Theorem 2.2.9]) :

THÉORÈME 4.4 (Bondal-Orlov). — *Le foncteur $\Phi_{\mathcal{O}_{\tilde{X}}} = \mathbf{R}p_*\mathbf{L}(p^+)^* : D^b(X^+) \rightarrow D^b(X)$ est pleinement fidèle et c'est une équivalence si $l = k$.*

PREUVE (esquisse) — Soit \mathcal{A} (resp. \mathcal{B}) la sous-catégorie triangulée pleine de $D^b(\tilde{X})$ engendrée par les $j_*\mathcal{O}(r, s)$ avec $r = 0$ et $-l \leq s \leq -1$ (resp. $-k \leq r < 0$ et $-l \leq s \leq -1$). Les décompositions du théorème 4.2 et de l'exemple 2.6 donnent une décomposition semi-orthogonale $(\mathbf{L}p^*D(X))^{\perp} = \langle \mathcal{B}, \mathcal{A} \rangle$. On a $\mathcal{B} \subset (\mathbf{L}(p^+)^*D^b(X^+))^{\perp}$. Puisque $(\omega_{\tilde{X}})_{|\tilde{Y}} \simeq \mathcal{O}(-k, -l)$, on a aussi $\mathcal{A} \otimes \omega_{\tilde{X}} \subset (\mathbf{L}(p^+)^*D^b(X^+))^{\perp}$, donc $\mathcal{A} \subset {}^{\perp}(\mathbf{L}(p^+)^*D^b(X^+))$.

Soit $C \in D^b(X^+)$ et soit \bar{C} le cône du morphisme canonique $\mathbf{L}p^*\mathbf{R}p_*(\mathbf{L}(p^+)^*C) \rightarrow \mathbf{L}(p^+)^*C$. On a $\bar{C} \in (\mathbf{L}p^*D(X))^{\perp} \cap {}^{\perp}\mathcal{B} = \mathcal{A}$. Par conséquent, le morphisme canonique

$$\mathrm{Hom}(C, D) \rightarrow \mathrm{Hom}(\mathbf{R}p_*\mathbf{L}(p^+)^*C, \mathbf{R}p_*\mathbf{L}(p^+)^*D)$$

est un isomorphisme pour tout $D \in D^b(X^+)$.

Supposons maintenant $k = l$. Soit $p^! = \mathbf{L}p^*(-) \otimes \omega_{\tilde{X}/X}[k]$. Le foncteur $\mathbf{R}p_*^+p^!$ est adjoint à droite de $\mathbf{R}p_*\mathbf{L}(p^+)^*$. On montre, comme ci-dessus, que $\mathbf{R}p_*^+(\omega_{\tilde{X}/X}[k] \otimes \mathbf{L}p^*(-))$ est pleinement fidèle et finalement $\mathbf{R}p_*\mathbf{L}(p^+)^*$ est une équivalence. \square

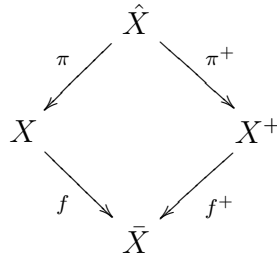
Remarque 4.5. — Il devrait en fait être vrai que $D^b(X)/\mathbf{R}p_*\mathbf{L}(p^+)^*D^b(X^+)$ a une suite exceptionnelle complète de longueur $k - l$.

4.4. Flop de Mukai

4.4.1. Soit X une variété projective lisse de dimension $2n \geq 4$. Soit $i : Y = \mathbf{P}^n \hookrightarrow X$ une immersion fermée. On suppose $\mathcal{N}_{Y/X} \simeq \Omega_Y^1$. On considère comme précédemment l'éclatement $p : \tilde{X} \rightarrow X$. Soit Y^{\vee} l'espace projectif dual. Alors, le diviseur exceptionnel \tilde{Y} s'identifie à la variété d'incidence dans $Y \times Y^{\vee}$ et on a $\mathcal{N}_{\tilde{Y}/\tilde{X}} \simeq \mathcal{O}(-1, -1)_{|\tilde{Y}}$. On a alors une contraction $p^+ : \tilde{X} \rightarrow X^+$ de \tilde{Y} sur Y^{\vee} et on suppose X^+ projective.

Le foncteur $\mathbf{R}p_*\mathbf{L}(p^+)^* : D^b(X^+) \rightarrow D^b(X)$ n'est pas une équivalence (cf. [29, Proposition 5.12] et [43, Corollary 2.2]). Par contre, un autre noyau fournit une équivalence (cf. [29, Corollary 5.7] et [43, Theorem 4.4]).

Soient $f : X \rightarrow \bar{X}$ et $f^+ : X^+ \rightarrow \bar{X}$ les contractions de Y et Y^+ sur un point (flop de Mukai). Soit $\hat{X} = X \times_{\bar{X}} X^+$. Alors, $\hat{X} = \tilde{X} \cup (Y \times Y^+) \subset X \times X^+$ et l'intersection est à croisements normaux.



THÉORÈME 4.6 (Kawamata, Namikawa). — *Le foncteur $\Phi_{\mathcal{O}_{\hat{X}}} = \mathbf{R}\pi_*\mathbf{L}(\pi^+)^* : D^b(X^+) \rightarrow D^b(X)$ est une équivalence.*

PREUVE (esquisse) — Il suffit de traiter le cas où $X = \text{Spec } \Omega_Y^1$ et $i : Y \rightarrow X$ est la section nulle, car cela ne change pas le complété formel de X le long de Y (cf. §3.1.2). La non-projectivité de X ne pose pas de problème nouveau. Soit $\mathcal{X} = \text{Spec } \mathcal{O}_Y(-1)^{n+1}$. La suite exacte $0 \rightarrow \Omega_Y^1 \rightarrow \mathcal{O}_Y(-1)^{n+1} \rightarrow \mathcal{O}_Y \rightarrow 0$ fournit un morphisme lisse $\mathcal{X} \rightarrow \mathbf{A}^1$ et la fibre de 0 s'identifie à X .

On a $\mathcal{N}_{Y/\mathcal{X}} \simeq \mathcal{O}_Y(-1)^{n+1}$ et on a alors un flop standard $\tilde{\mathcal{X}} \rightarrow \mathcal{X}^+$ (cf. §4.3). Le foncteur $\Phi_{\mathcal{O}_{\tilde{\mathcal{X}}}} : D^b(\mathcal{X}^+) \rightarrow D^b(\mathcal{X})$ est une équivalence et il résulte de la proposition 3.5 que $\Phi_{\mathcal{O}_{\tilde{\mathcal{X}}}}$ est une équivalence. \square

Deux variétés projectives symplectiques birationnelles de dimension 4 sont connectées par une suite de flops de Mukai ([63, Theorem 1.2], [20, §8] ; il faut en fait permettre la contraction simultanée de plusieurs \mathbf{P}^2 disjoints pour que les variétés intermédiaires restent projectives [62]). On déduit alors ([43, Corollary 4.5] et [29, Remark 5.13]) :

COROLLAIRE 4.7 (Kawamata, Namikawa). — *Deux variétés projectives symplectiques birationnelles de dimension 4 sont dérivé-équivalentes.*

4.4.2. *Flop de Mukai stratifié.* — Markman [40] a étudié une généralisation du flop de Mukai. Nous suivons ici la présentation de [44] et ne donnons que la version « linéarisée ».

Soit V un espace vectoriel de dimension n et soit $G(V, r)$ la grassmannienne des sous-espaces de dimension r de V , où r est un entier $\leq n/2$. Soient $X = T^*G(V, r)$ et \bar{X} la sous-variété fermée de $\text{End}_{\mathbf{C}}(V)$ des endomorphismes a tels que $a^2 = 0$ et $\text{rang}(a) \leq r$. On identifie $T^*G(V, r)$ à la variété des paires (W, ϕ) où W est un sous-espace de dimension r de V et $\phi \in \text{Hom}(V/W, W)$. On dispose d'une application moment $f : X \rightarrow \bar{X}$ qui envoie (W, ϕ) sur la composition $a : V \twoheadrightarrow V/W \xrightarrow{\phi} W \hookrightarrow V$. L'application moment est un isomorphisme au-dessus de l'ouvert des endomorphismes de rang r . On dispose de même d'une application moment $f^+ : X^+ = T^*G(V, n-r) \rightarrow \bar{X}$.

Le noyau « évident » ne fournit pas une équivalence dérivée, bien que sa classe dans K_0 soit adéquate [44, Theorem 2.7 et Observation 4.9] :

THÉORÈME 4.8 (Namikawa). — *L'application $[\Phi_{\mathcal{O}_{X \times_{\bar{X}} X^+}}] : K_0(X^+) \rightarrow K_0(X)$ est un isomorphisme pour tous n, r mais $\Phi_{\mathcal{O}_{X \times_{\bar{X}} X^+}} : D^b(X^+) \rightarrow D^b(X)$ n'est pas une équivalence pour $n = 4$ et $r = 2$.*

Dans le cas $n = 4$ et $r = 2$, Kawamata construit un noyau de la forme $\mathcal{O}_{X \times_{\bar{X}} X^+} \otimes \mathcal{L}$ où \mathcal{L} est un fibré en droites, qui induit une équivalence [32].

Voyons maintenant une construction pour n, r généraux d'un noyau supporté par $X \times_{\bar{X}} X^+$ et induisant une équivalence dérivée.

Soit U la sous-variété ouverte de $G(V, r) \times G(V, n-r)$ des (W, W') tels que $W \cap W' = 0$ et soit $\iota : U \rightarrow G(V, r) \times G(V, n-r)$ l'immersion ouverte.

$$\begin{array}{ccc}
 & U & \\
 & \downarrow \iota & \\
 & G(V, r) \times G(V, n-r) & \\
 \alpha \swarrow & & \searrow \beta \\
 G(V, r) & & G(V, n-r)
 \end{array}$$

Il est classique [28, Exercice III.15] que la transformation de noyau $\iota_! \mathbf{C}_U$ induit une équivalence entre catégories dérivées de faisceaux constructibles de \mathbf{C} -espaces vectoriels :

$$\mathbf{R}\alpha_!(\mathbf{C}_U \otimes \beta^{-1}(-)) : D_{\text{cons}}^b(G(V, n-r)) \xrightarrow{\sim} D_{\text{cons}}^b(G(V, r)).$$

Soit \mathcal{K} le module de Hodge mixte correspondant à $\iota_! \mathbf{C}_U$. C'est un noyau inversible pour les transformations entre catégories dérivées de modules de Hodge mixtes. Soit $K = \text{Gr}(\mathcal{K})$, un faisceau cohérent \mathbf{G}_m -équivariant sur $T^*(G(V, r) \times G(V, n-r)) = X \times X^+$. Alors, l'inversibilité du noyau \mathcal{K} montre que $\Phi_K : D^b(X^+) \rightarrow D^b(X)$ est une équivalence (c'est bien sûr aussi une équivalence pour les catégories \mathbf{G}_m -équivariantes).

Il serait intéressant de décrire explicitement le faisceau K . Kashiwara a proposé une description pour $n = 4$ et $r = 2$, qu'il faudrait comparer avec le noyau de [32].

Cette construction se généralise à des variétés de drapeaux paraboliques pour des groupes semi-simples complexes arbitraires. Dans le cas des variétés de drapeaux complets, une telle construction au niveau de la K -théorie a été effectuée par Tanisaki [55].

On peut aussi construire une équivalence $D^b(X^+) \xrightarrow{\sim} D^b(X)$ en regroupant les $D^b(T^*G(V, r))$, $0 \leq r \leq \dim V$, et en appliquant les méthodes de [21].

4.5. Dimension 3

4.5.1. On expose ici la construction de Bridgeland [12] (voir aussi [19, 29, 61]).

Considérons un flop entre variétés projectives lisses de dimension 3 :

$$\begin{array}{ccc}
 X & & X^+ \\
 & \searrow f & \swarrow f^+ \\
 & \bar{X} &
 \end{array}$$

Remarque 4.9. — Lorsque f ne contracte qu'une courbe irréductible C , alors $C \simeq \mathbf{P}^1$ et le fibré normal $\mathcal{N}_{C/X}$ est $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$, $\mathcal{O} \oplus \mathcal{O}(-2)$ ou $\mathcal{O}(1) \oplus \mathcal{O}(-3)$ (cf. [22, Corollary 16.3]). Le premier cas est un flop standard (§4.3) et le second cas peut se traiter par des méthodes similaires [6, Theorem 3.9].

Considérons le diagramme

$$\begin{array}{ccc} & X \times_{\bar{X}} X^+ & \\ \pi \swarrow & & \searrow \pi^+ \\ X & & X^+ \end{array}$$

THÉORÈME 4.10 (Bridgeland). — *Le foncteur $\mathbf{R}\pi_*\mathbf{L}(\pi^+)^* : D^b(X^+) \rightarrow D^b(X)$ est une équivalence.*

Bridgeland construit le flop f^+ à partir de f : la variété X^+ apparaît comme un espace de modules fin de certains objets de $D^b(X)$ (« faisceaux pervers ponctuels ») et le noyau de l'équivalence est le fibré universel. La détermination de ce fibré, et donc la forme précise du théorème 4.10, sont dues à Chen.

Un flop généralisé entre variétés projectives lisses de dimension 3 se décompose en une suite de flops [29, Theorem 4.6] et on en déduit :

COROLLAIRE 4.11. — *La conjecture 1.1 est vraie en dimension 3.*

4.5.2. *Construction du flop et équivalence.* — Soit \bar{X} une variété projective connexe de Gorenstein de dimension 3. Soit $f : X \rightarrow \bar{X}$ une résolution crépante dont le lieu exceptionnel est union d'un nombre fini de courbes et soit D un diviseur sur X tel que $-D$ est relativement f -ample.

Le théorème d'annulation de Grauert-Riemenschneider montre, via la formule de projection, que $\mathbf{R}^{>0}f_*\mathcal{O}_X = 0$. Le foncteur $\mathbf{L}f^* : D(\bar{X}) \rightarrow D(X)$ est donc pleinement fidèle (variante de la proposition 3.3 pour les catégories dérivées non bornées que requiert la non-lissité de \bar{X}). Soit \mathcal{B} son image. On a une décomposition semi-orthogonale $D(X) = \langle \mathcal{B}^\perp, \mathcal{B} \rangle$, où $\mathcal{B}^\perp = \{C \in D(X) \mid \mathbf{R}f_*C = 0\}$.

On construit une nouvelle t-structure sur $D(X)$ en recollant la t-structure standard de \mathcal{B} avec celle de \mathcal{B}^\perp décalée de 1 vers la gauche. Son cœur $\text{Per}_{X/\bar{X}}$ (une catégorie abélienne) consiste en les $C \in D(X)$ tels que

- $\mathcal{H}^i C = 0$ pour $i \neq 0, 1$
- $f_*\mathcal{H}^{-1}C = 0$
- $\mathbf{R}^1f_*\mathcal{H}^0C = 0$ et $\text{Hom}(\mathcal{H}^0C, M) = 0$ pour tout $M \in \mathcal{B}^\perp \cap X\text{-coh}$.

On a $\mathcal{O}_X \in \text{Per}_{X/\bar{X}}$ et on dit que $E \in \text{Per}_{X/\bar{X}}$ est un « faisceau pervers ponctuel » si

- E est un quotient de \mathcal{O}_X (dans $\text{Per}_{X/\bar{X}}$)
- la classe de Chern de E est égale à celle d'un faisceau gratte-ciel.

Soit S une variété. Une famille de faisceaux pervers ponctuels paramétrée par S est un objet $\mathcal{E} \in D^b(S \times X)$ tel que $\mathbf{L}j_s^*\mathcal{F} \in \text{Per}_{X/\bar{X}}$ pour tout point (fermé) $s \in S$. Ici, $j_s : s \times X \rightarrow S \times X$ est l'inclusion. Deux familles qui diffèrent par le produit par un fibré inversible de S sont dites équivalentes.

On considère le foncteur qui à une variété S associe l'ensemble des classes d'équivalence de familles de faisceaux pervers ponctuels paramétrées par S . Bridgeland

[12, Theorem 3.8] démontre l'existence d'un espace de modules fin : le foncteur est représentable par une variété projective $M(X/\bar{X})$.

Soit U l'ouvert de \bar{X} des points au-dessus desquels f est un isomorphisme. Alors, $\mathcal{O}_{f^{-1}(u)}$ est un faisceau pervers ponctuel pour tout $u \in U$. Ceci définit une immersion ouverte $U \rightarrow M(X/\bar{X})$ et on note X^+ l'adhérence de l'image (on montre en fait que $X^+ = M(X/\bar{X})$). On note $f^+ : X^+ \rightarrow \bar{X}$ le morphisme canonique. Chen [19, Proposition 4.2] montre que $\mathcal{O}_{X \times_{\bar{X}} X^+}$ est un fibré universel.

Le théorème 4.10 admet la version plus précise suivante.

THÉORÈME 4.12 (Bridgeland). — *Le foncteur $\mathbf{R}\pi_*\mathbf{L}(\pi^+)^* : D^b(X^+) \rightarrow D^b(X)$ est une équivalence. En outre, $f^+ : X^+ \rightarrow \bar{X}$ est un flop : f^+ est une résolution crépante et le transformé strict de D est relativement f^+ -ample.*

La preuve est essentiellement la même que pour la correspondance de McKay [13]. L'outil-clef est un résultat d'algèbre commutative que nous rappelons maintenant (dans la version de [15, §5]).

Soient Z une variété irréductible et $C \in D^b(Z)$ non nul. Soit $\text{Supp}(C)$ l'union des supports des $\mathcal{H}^i(C)$. Soit $\text{ampl}(C)$ l'ensemble des $i \in \mathbf{Z}$ tels qu'il existe $z \in Z$ avec $\text{Hom}(C, \mathcal{O}_z[-i]) \neq 0$. La dimension homologique de C est $\text{hd}(C) = \sup(\text{ampl}(C)) - \inf(\text{ampl}(C))$.

THÉORÈME 4.13 (« Nouveau théorème d'intersection »). — *On a $\text{codim Supp}(C) \leq \text{hd}(C)$.*

Soit $z \in Z$ tel que $\text{Supp}(C) = \{z\}$, $\mathcal{H}^0(C) \simeq \mathcal{O}_z$ et $\text{ampl}(C) \subseteq [-\dim Z, 0]$. Alors, z est un point lisse de Z et $C \simeq \mathcal{O}_z$.

PREUVE DU THÉORÈME 4.12 (esquisse) — Soit $p_X : X \times X^+ \rightarrow X$ la première projection. Soient $\mathcal{P} = \mathcal{O}_{X \times_{\bar{X}} X^+}$ et $\mathcal{P}_w = \Phi_{\mathcal{P}}(\mathcal{O}_w)$ pour $w \in X^+$. Soient $\mathcal{P}' = \mathcal{P}^\vee \otimes p_X^* \omega_X[3]$ et $\mathcal{Q} = \mathcal{P}' \circ \mathcal{P} \in D^b(X^+ \times X^+)$. On vérifie que $\text{Hom}(\mathcal{P}_w, \mathcal{P}_{w'}[i]) = 0$ pour tout i , lorsque $f^+(w) \neq f^+(w')$ et on montre que $\text{Hom}(\mathcal{P}_w, \mathcal{P}_{w'}) = \delta_{w,w'} \mathbf{C}$, d'où $\text{Hom}(\mathcal{P}_{w'}, \mathcal{P}_w[3]) \simeq \delta_{w,w'} \mathbf{C}$ par dualité de Serre.

Soit $\mathcal{Q}' = \mathcal{Q}|_{X^+ \times_{\bar{X}} X^+ - \Delta X^+}$. Le support de \mathcal{Q}' est contenu dans $X^+ \times_{\bar{X}} X^+ - \Delta X^+$, donc est de codimension ≥ 2 , s'il est non vide. On a $\text{Hom}(\mathcal{Q}', \mathcal{O}_{w,w'}[i]) \simeq \text{Hom}(\mathcal{P}_w, \mathcal{P}_{w'}[i])$, donc $\text{hd}(\mathcal{Q}') \leq 1$. Le théorème 4.13 montre alors que $\mathcal{Q}' = 0$.

Le morphisme canonique $\text{Hom}(\mathcal{O}_w, \mathcal{O}_w[1]) \rightarrow \text{Hom}(\mathcal{P}_w, \mathcal{P}_w[1])$ est injectif (il s'identifie à l'application de Kodaira-Spencer). On en déduit que $\mathcal{H}^0(\Phi_{\mathcal{Q}}(\mathcal{O}_w)) \simeq \mathcal{O}_w$. On a $\text{ampl}(\Phi_{\mathcal{Q}}(\mathcal{O}_w)) = [-3, 0]$. Le théorème 4.13 montre que $\Phi_{\mathcal{Q}}(\mathcal{O}_w) \simeq \mathcal{O}_w$ et que X^+ est lisse en w . La crépance de f permet alors de conclure que $\Phi_{\mathcal{P}}$ est une équivalence, via la proposition 3.2.

La crépance de f^+ s'obtient en utilisant la trivialité du foncteur de Serre de $D_F^b(X^+)$ pour toute fibre F de f^+ [13, Lemma 3.1]. On montre que $\Phi_{\mathcal{P}}[-1]$ se restreint en une équivalence entre faisceaux cohérents sur X^+ tués par $\mathbf{R}f_*^+$ et faisceaux cohérents sur X tués par $\mathbf{R}f_*$ et l'amplitude relative du transformé strict de D s'en déduit [12, §4.6]. \square

Remarque 4.14. — Il serait intéressant de voir si la même méthode fournit une construction de flops (cf. [1] pour des résultats dans cette direction, en dimension quelconque).

4.6. Variétés singulières

Les constructions pour les variétés projectives lisses ont des généralisations à une classe plus large, dans le cadre du MMP [2, 19, 29, 30, 31, 61]. Ceci est utile pour construire des équivalences entre variétés projectives lisses, car il existe des décompositions de flops généralisés en flops qui font intervenir des variétés singulières. Kawamata démontre en particulier un théorème d'équivalence dérivée pour des flops particuliers entre champs de Deligne-Mumford toriques. Il en déduit qu'un flop torique généralisé donne lieu à une équivalence dérivée [31, Corollary 4.5] :

THÉORÈME 4.15 (Kawamata). — *Soient $f : Z \rightarrow X$ et $g : Z \rightarrow Y$ des morphismes toriques birationnels entre variétés toriques projectives lisses tels que $f^*\omega_X \simeq g^*\omega_Y$. Alors, $D^b(Y) \simeq D^b(X)$.*

Mentionnons pour terminer la nécessité de considérer des variétés analytiques et d'effectuer les constructions dans le cadre de la log-géométrie.

RÉFÉRENCES

- [1] D. Abramovich and J.C. Chen, *Computations with moduli of perverse point sheaves*, preprint math.AG/0304353.
- [2] D. Abramovich and J.C. Chen, *Flops, flips and perverse point sheaves on threefold stacks*, preprint math.AG/0304354.
- [3] P. Balmer, *Presheaves of triangulated categories and reconstruction of schemes*, Math. Ann. **324** (2002), 557–580.
- [4] A.A. Beilinson, *The derived category of coherent sheaves on P^n* , Selecta Math. Soviet. **3** (1983/84), 233–237 ou (en russe) Funktsional. Anal. i Prilozhen. **12** (1978), 68–69.
- [5] A. Bondal et M.M. Kapranov, *Representable functors, Serre functors, and mutations*, Math. USSR-Izv. **35** (1990), 519–541.
- [6] A. Bondal et D. Orlov, *Semiorthogonal decompositions for algebraic varieties*, preprint alg-geom/9506012.
- [7] A. Bondal et D. Orlov, *Reconstruction of a variety from the derived category and groups of autoequivalences*, Compositio Math. **125** (2001), 327–344.
- [8] A. Bondal et D. Orlov, *Derived categories of coherent sheaves*, Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), 47–56, Higher Ed. Press, 2002.
- [9] A. Bondal et M. Van den Bergh, *Generators and representability of functors in commutative and noncommutative geometry*, Moscow Math. J. **3** (2003), 1–36.

- [10] T. Bridgeland, *Fourier-Mukai transforms for elliptic surfaces*, J. reine angew. Math. **498** (1998), 115–133.
- [11] T. Bridgeland, *Equivalences of triangulated categories and Fourier-Mukai transforms*, Bull. London Math. Soc. **31** (1999), 25–34.
- [12] T. Bridgeland, *Flops and derived categories*, Invent. Math. **147** (2002), 613–632.
- [13] T. Bridgeland, A. King et M. Reid, *The McKay correspondence as an equivalence of derived categories*, J. Amer. Math. Soc. **14** (2001), 535–554.
- [14] T. Bridgeland et A. Maciocia, *Complex surfaces with equivalent derived categories*, Math. Z. **236** (2001), 677–697.
- [15] T. Bridgeland et A. Maciocia, *Fourier-Mukai transforms for K3 and elliptic fibrations*, J. Algebraic Geom. **11** (2002), 629–657.
- [16] A. Căldăraru, *The Mukai pairing, I: the Hochschild structure*, preprint math.AG/0308079(v2).
- [17] A. Căldăraru, *The Mukai pairing, II: the Hochschild-Kostant-Rosenberg isomorphism*, preprint math.AG/0308080(v3).
- [18] A. Căldăraru, *Derived categories of sheaves: a skimming*, preprint math.AG/0501094.
- [19] J.-C. Chen, *Flops and equivalences of derived categories for threefolds with only terminal Gorenstein singularities*, J. Differential Geom. **61** (2002), 227–261.
- [20] K. Cho, Y. Miyaoka et N.I. Shepherd-Barron *Characterizations of projective space and applications to complex symplectic manifolds*, in “Higher dimensional birational geometry (Kyoto, 1997)”, 1–88, Math. Soc. Japan, 2002.
- [21] J. Chuang et R. Rouquier, *Derived equivalences for symmetric groups and \mathfrak{sl}_2 -categorification*, preprint math.RT/0407205.
- [22] H. Clemens, J. Kollár et S. Mori, *Higher-dimensional complex geometry*, Astérisque **166** (1988).
- [23] P. Gabriel, *Des catégories abéliennes*, Bull. Soc. Math. France **90** (1962), 323–448.
- [24] A.L. Gorodentsev et S.A. Kuleshov, *Helix theory*, Mosc. Math. J. **4** (2004), 377–440, 535.
- [25] L. Hille et M. Van den Bergh, *Fourier-Mukai transforms*, preprint math.AG/0402043(v2).
- [26] D. Huybrechts, “Fourier-Mukai transforms in algebraic geometry”, livre en préparation.
- [27] M.M. Kapranov, *On the derived categories of coherent sheaves on some homogeneous spaces*, Invent. Math. **92** (1988), 479–508.
- [28] M. Kashiwara et P. Schapira, “Sheaves on manifolds”, Springer-Verlag, 1990.
- [29] Y. Kawamata, *D-equivalence and K-equivalence*, J. Differential Geom. **61** (2002), 147–171.
- [30] Y. Kawamata, *Francia’s flip and derived categories*, in “Algebraic geometry”, 197–215, de Gruyter, 2002.

- [31] Y. Kawamata, *Equivalences of derived categories of sheaves on smooth stacks*, Amer. J. Math. **126** (2004), 1057–1083.
- [32] Y. Kawamata, *Derived equivalence for stratified Mukai flop on $G(2, 4)$* , preprint math.AG/0503101.
- [33] Y. Kawamata, *Derived Categories of Toric Varieties*, preprint math.AG/0503102.
- [34] A. King, *Tilting bundles on some rational surfaces*, preprint (1997), <http://www.maths.bath.ac.uk/~masadk/papers/tilt.ps>
- [35] M. Kontsevich, *Homological algebra of mirror symmetry*, Proceedings of the International Congress of Mathematicians, Vol. 1 (Zürich, 1994), 120–139, Birkhäuser, 1995.
- [36] M. Kontsevich, *Deformation quantization of Poisson manifolds*, Lett. Math. Phys. **66** (2003), 157–216.
- [37] A. Kuznetsov, *Derived category of a cubic threefold and the variety V_{14}* , Tr. Mat. Inst. Steklova 246 (2004), 183–207.
- [38] A. Kuznetsov, *Derived category of V_{12} Fano threefolds*, preprint math.AG/0310008.
- [39] E. Looijenga, *Motivic measures*, Séminaire Bourbaki, exposé 874, Astérisque **276** (2002), 267–297.
- [40] E. Markman, *Brill-Noether duality for moduli spaces of sheaves on K3 surfaces*, J. Algebraic Geom. **10** (2001), 623–694.
- [41] S. Mukai, *Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves*, Nagoya Math. J. **81** (1981), 153–175.
- [42] S. Mukai, *On the moduli space of bundles on K3 surfaces. I*, in “Vector bundles on algebraic varieties (Bombay, 1984)”, 341–413, Tata Inst. Fund. Res., 1987.
- [43] Y. Namikawa, *Mukai flops and derived categories*, J. reine angew. Math. **560** (2003), 65–76.
- [44] Y. Namikawa, *Mukai flops and derived categories. II*, in “Algebraic structures and moduli spaces”, 149–175, Amer. Math. Soc., 2004.
- [45] D. Orlov, *Projective bundles, monoidal transformations, and derived categories of coherent sheaves*, Russian Acad. Sci. Izv. Math. **41** (1993), 133–141.
- [46] D. Orlov, *Equivalences of derived categories and K3 surfaces*, J. Math. Sci. (New York) **84** (1997), 1361–1381.
- [47] D. Orlov, *Derived categories of coherent sheaves on abelian varieties and equivalences between them*, Izv. Math. **66** (2002), 569–594.
- [48] D. Orlov, *Derived categories of coherent sheaves and equivalences between them*, Russian Math. Surveys 58 (2003), no. 3, 511–591.
- [49] A. Polishchuk, “Abelian varieties, theta functions and the Fourier transform”, Cambridge University Press, 2003.
- [50] M. Reid, *La correspondance de McKay*, Séminaire Bourbaki, exposé 867, Astérisque **276** (2002), 53–72.

- [51] R. Rouquier, *Catégories dérivées et géométrie algébrique*, notes d'exposés, janvier 2004, <http://www.math.jussieu.fr/~rouquier/preprints/luminy.dvi>
- [52] R. Rouquier, *Dimensions of triangulated categories*, preprint [math.CT/0310134\(v3\)](http://math.CT/0310134(v3)).
- [53] A. Rudakov, *Rigid and exceptional vector bundles and sheaves on a Fano variety*, Proceedings of the International Congress of Mathematicians, Vol. 1 (Zürich, 1994), 697–705, Birkhäuser 1995.
- [54] R. Swan, *Hochschild cohomology of quasiprojective schemes*, J. Pure Appl. Algebra **110** (1996), 57–80.
- [55] T. Tanisaki, *Hodge modules, equivariant K-theory and Hecke algebras*, Publ. Res. Inst. Math. Sci. **23** (1987), 841–879.
- [56] R.W. Thomason, *Les K-groupes d'un fibré projectif*, in “Algebraic K-theory and algebraic topology (Lake Louise, 1991)”, 243–248, Kluwer, 1993.
- [57] R.W. Thomason, *Les K-groupes d'un schéma éclaté et une formule d'intersection excédentaire*, Invent. Math. **112** (1993), 195–215.
- [58] R.W. Thomason, *The classification of triangulated subcategories*, Compositio Math. **105** (1997), 1–27.
- [59] B. Toen, *The homotopy category of dg-categories and derived Morita theory*, preprint [math.AG/0408337\(v5\)](http://math.AG/0408337(v5)).
- [60] H. Uehara, *An example of Fourier-Mukai partners of minimal elliptic surfaces*, Math. Res. Lett. **11** (2004), 371–375.
- [61] M. Van den Bergh, *Three-dimensional flops and noncommutative rings*, Duke Math. J. **122** (2004), 423–455.
- [62] J. Wierzba, *Birational geometry of symplectic 4-folds*, preprint, <http://www.dpmms.cam.ac.uk/~jw227/publications.html>.
- [63] J. Wierzba et J.A. Wiśniewski, *Small contractions of symplectic 4-folds*, Duke Math. J. **120** (2003), 65–95.

Raphaël ROUQUIER

Institut de Mathématiques de Jussieu
 UMR 7586 du CNRS
 UFR de Mathématiques, Université Paris 7
 Case 7012
 2 place Jussieu
 F-75251 PARIS Cédex 05
E-mail : rouquier@math.jussieu.fr

**PREUVE DE LA CONJECTURE DE POINCARÉ EN DÉFORMANT
LA MÉTRIQUE PAR LA COURBURE DE RICCI,
D'APRÈS G. PEREL'MAN**

par **Gérard BESSON**

INTRODUCTION

Dans un article célèbre de 1904 ([48]), H. Poincaré pose la question qu'en termes actuels nous énonçons sous la forme de la conjecture suivante :

CONJECTURE 0.1. — Si M est une variété compacte sans bord simplement connexe de dimension 3, alors M est homéomorphe à la sphère.

En dimension 3, la conclusion équivalente est que M est difféomorphe à la sphère. De nombreux outils topologiques ont été élaborés afin de résoudre ce problème ; un historique de ces développements est décrit dans l'article [42]. Une preuve de la conjecture qui suit permettrait de compléter la compréhension des variétés de dimension 3, compactes, connexes de groupe fondamental fini :

CONJECTURE 0.2. — Un groupe fini de difféomorphismes qui agit librement sur S^3 est conjugué à un sous-groupe du groupe d'isométries de la sphère canonique.

En 1982, W. Thurston a replacé ces questions dans un cadre plus général, inspiré par la classification des surfaces. Reprenons l'énoncé de la conjecture dite de géométrisation tel qu'il est formulé dans [57].

CONJECTURE 0.3. — L'intérieur de toute variété compacte de dimension 3 admet une décomposition canonique en pièces qui portent une structure géométrique.

Dans cet énoncé la variété peut avoir un bord. Dans la suite de ce rapport nous appellerons compacte une variété compacte sans bord (le terme fermée serait plus usuel en topologie) et toutes les variétés seront supposées orientées. La décomposition à laquelle il est fait allusion procède en deux étapes :

1) celle qui provient du théorème de Kneser dans lequel la variété est décomposée en une somme connexe d'un nombre fini de variétés premières. On rappelle qu'une variété M est dite première si $M = P \# Q$ implique $P = S^3$ ou $Q = S^3$.

2) Celle qui provient des travaux de K. Johannson et de W. Jaco et P. Shalen et qui consiste à découper le long de tores incompressibles. La conjecture affirme que ceci peut être fait en sorte que les variétés à bords qui en résultent possèdent une géométrie, c'est-à-dire une métrique riemannienne complète localement homogène. Celles-ci sont

classifiées, il y a huit possibilités (en dimension 3). Le lecteur trouvera une intéressante discussion de cette conjecture ainsi que des références précises dans [57] et [1].

Un des avantages de la conjecture 0.3 est qu'elle fait référence à l'existence de métriques riemanniennes privilégiées, sur certaines régions de la variété étudiée, et fournit ainsi des outils supplémentaires, au-delà de la topologie. En 1982, R. Hamilton a introduit une méthode, que nous pouvons qualifier d'analytique, dans le but d'aborder ces questions. Il s'agit d'étudier une équation différentielle sur l'espace des métriques riemanniennes, nous dirons un flot, dont les solutions sont une déformation d'une métrique quelconque qui tend à la rendre de courbure constante. L'équation met en jeu la courbure de Ricci, qui est une notion de courbure de même nature que la métrique, c'est-à-dire une forme bilinéaire symétrique sur chaque espace tangent. L'équation est :

$$\frac{\partial g}{\partial t} = -2 \operatorname{Ric}_{g(t)},$$

où $g(t)$ désigne la métrique riemannienne qui évolue et $\operatorname{Ric}_{g(t)}$ sa courbure de Ricci. L'équation ci-dessus est inspirée par la tentative de minimiser l'intégrale de la courbure scalaire dont ce n'est toutefois pas le flot de l'opposé du gradient ; elle est également liée à l'équation d'évolution associée aux applications harmoniques. Le signe négatif montre que la courbure positive est contractée et la courbure négative est dilatée, comme on peut s'en convaincre sur les exemples de courbure constante. Remarquons que, si la donnée initiale admet des isométries, c'est le cas pour la solution, pour tout temps de son intervalle de vie. En d'autres termes, le groupe d'isométries ne peut que croître.

Dans une série d'articles ([24], [25], [31], [29], [19], [26], [27], [28] et [30]), R. Hamilton développe les outils d'analyse nécessaires à l'utilisation de sa méthode et obtient de remarquables résultats géométriques, dont certains sont des réponses aux questions posées ci-dessus, dans des cas particuliers. De nombreuses difficultés persistent toutefois, en particulier celles liées à l'étude des singularités qui peuvent apparaître lors de l'évolution.

Récemment, G. Perelman a déposé sur la toile trois articles, [44], [46] et [45] dans lesquels une solution complète de la conjecture 0.3 est proposée. Ils apportent des idées novatrices et puissantes à la méthode du flot de la courbure de Ricci, et surtout à la description des régions qui deviennent singulières, c'est-à-dire où la courbure explose. Ceci permet de pratiquer une chirurgie, déjà en grande partie décrite dans [25], de manière efficace. On construit ainsi un flot défini pour tout temps mais pas de classe C^∞ , les singularités correspondant à des temps de chirurgies. Il se peut même que la variété disparaisse à un moment de l'évolution, on dit alors qu'elle s'éteint en temps fini. Les articles ne sont pas très détaillés, il s'agit plutôt d'esquisses de preuve, néanmoins assez claires. Ils font l'objet d'un intense travail de mise en place et de vérifications des détails (et d'exégèse). L'expertise n'étant pas complètement terminée, il est difficile de se prononcer pour l'instant (au moment où ces lignes sont écrites) sur la question de savoir si la conjecture 0.3 est prouvée. Toutefois, le cas de la conjecture de Poincaré est plus « simple » dans le schéma de G. Perelman ; il montre, en effet, que, pour n'importe

quelle métrique sur une variété compacte simplement connexe, le flot s'éteint en temps fini ([45]). Une conséquence est qu'on ne pratique qu'un nombre fini de chirurgies. L'auteur du présent texte est convaincu que les conjectures 0.1 et 0.2 sont prouvées.

Le but de ce rapport est de décrire les outils d'analyse et de géométrie nécessaires ainsi que le schéma de la preuve des conjectures 0.1 et 0.2. Il est conçu comme intermédiaire entre les survols [1], [39], [42] et [43] et les notes détaillées [36], [51] et [61]. L'auteur espère qu'il peut être un guide de lecture de ces documents. Les premières notes détaillées produites ont été celles de B. Kleiner et J. Lott, nous les conseillons vivement à tous ceux qui souhaitent comprendre les travaux de G. Perelman ; de même les documents [51] et [61] sont d'une grande précision sur de nombreux points de [44]. Ce texte est écrit alors que l'auteur travaille encore à améliorer sa compréhension de cet ensemble de travaux ; il contient certainement des erreurs qui seront expurgées dans des versions ultérieures, déposées sur le site de l'Institut Fourier⁽¹⁾. Signalons enfin les articles [47] et [2] qui s'adressent à un large public et sont d'une grande utilité.

Remerciements.— Je tiens à remercier chaleureusement L. Bessières pour le travail que nous avons fait en commun sur cette théorie durant deux ans et les nombreuses corrections apportées à ce texte, ainsi que B. Kleiner pour avoir répondu avec gentillesse à toutes nos questions et J.-P Bourguignon, J. Lott et S. Maillot pour leurs suggestions. Mes remerciements vont également à V. Bayle, M. Boileau, H.-D. Cao, B. Chow, É. Ghys, L. Guillou, J.-M. Iniotakis, B. Leeb (et tout le groupe de Munich), Ph. LeFloch, J. Porti, L. Rozoy, R. Souam et P. Topping pour des échanges fructueux.

1. DES MODÈLES JOUETS

Il existe deux modèles qui permettent de mieux se familiariser avec les rudiments concernant les équations d'évolution géométriques. Il s'agit du raccourcissement des courbes et du flot de la courbure sur les surfaces. Ce sont des jouets déjà assez sophistiqués que nous décrivons très brièvement.

1.1. Le raccourcissement des courbes

Soit \mathcal{C} une courbe plane fermée simple paramétrée par une fonction de classe C^∞ , $F_0 : S^1 \rightarrow \mathbf{R}^2$; on la suppose orientée positivement. Pour $T > 0$, on cherche une famille $F : S^1 \times [0, T) \rightarrow \mathbf{R}^2$ de classe C^∞ en ses deux variables vérifiant l'équation d'évolution suivante :

$$(*_c) \quad \begin{cases} \frac{\partial F}{\partial t}(x, t) = k(x, t)\vec{N}(x, t) \\ F(\cdot, 0) = F_0 \end{cases}$$

où $\vec{N}(x, t)$ (resp. $k(x, t)$) est le vecteur normal intérieur (resp. la courbure) de la courbe $F_t(\cdot) = F(\cdot, t)$ au point $F(x, t)$.

⁽¹⁾<http://www-fourier.ujf-grenoble.fr/GT/perelman/index.html>

Cette équation est non linéaire car k et N dépendent de F (voir la section 3). Il est facile de vérifier qu'il s'agit ici du flot de l'opposé du gradient de la fonctionnelle longueur ; un calcul immédiat montre que l'aire $A(t)$ enclose par la courbe F_t (l'aire de la composante connexe bornée) vérifie $A'(t) = -2\pi$. Le théorème principal est dû à M. Gage et R. Hamilton [19] pour le cas où \mathcal{C} est convexe et à M. Grayson [21] pour le cas général. Afin d'énoncer ces résultats (en un seul théorème) définissons les quantités $k_{\max}(t) = \max\{k(x, t); x \in S^1\}$ et, de même, $k_{\min}(t)$ comme minimum de la courbure au temps t , ainsi que $r_{\max}(t) =$ rayon du cercle circonscrit à F_t et, de même, $r_{\min}(t)$ comme le rayon du cercle inscrit dans F_t .

THÉORÈME 1.1 ([19] et [21]). — *Pour toute courbe C^∞ fermée simple il existe une unique solution de l'équation $(*_c)$ définie sur un intervalle de temps $[0, T)$ où $T = A(0)/2\pi$. La famille de courbe F_t converge, lorsque $t \rightarrow T$, vers un point et devient circulaire aux sens suivants :*

- i) *le quotient k_{\max}/k_{\min} tend vers 1 lorsque t tend vers T ,*
- ii) *le quotient r_{\max}/r_{\min} tend vers 1 lorsque t tend vers T .*

De plus, pour tout $n \geq 1$, la dérivée spatiale n -ième de k converge uniformément vers 0 lorsque t tend vers T . Enfin, si la courbe \mathcal{C} est convexe elle le reste tout au long du processus ([19]) et si \mathcal{C} n'est pas convexe la courbe F_t le devient en un temps $t < T$ ([21]).

Ce théorème affirme en fait que, quitte à renormaliser afin que l'aire intérieure soit constante, la famille F_t converge dans la topologie C^k , pour tout k , vers un cercle.

Questions :

1) Si la courbe de départ est convexe, alors F_t reste dans son intérieur pour tout t . En particulier le point limite est dans l'enveloppe convexe de \mathcal{C} . Il serait intéressant de déterminer sa position.

2) Ce théorème nécessite une courbe initiale régulière. Que se passe-t-il si \mathcal{C} possède des coins? Dans [19] il est fait allusion à une solution possible.

3) Une extension aux corps convexes de \mathbf{R}^n est décrite dans [33].

1.2. Le flot de la courbure sur les surfaces

On munit une surface abstraite compacte et orientable Σ d'une métrique riemannienne notée g_0 . Dans la classe conforme de g_0 il existe une métrique de courbure constante (unique à isométrie près). On souhaite, comme précédemment, obtenir cette « forme » idéale comme limite d'un flot géométrique. Si $g(t)$ est une famille C^∞ de métriques riemanniennes sur Σ , dépendant de manière C^∞ du paramètre t , on note $R(x, t)$ la courbure scalaire de $g(t)$ au point $x \in \Sigma$; avec les conventions habituelles, la courbure scalaire est le double de la courbure de Gauß. La mesure riemannienne, notée $v_{g(t)}$ (ou $d\text{vol}$), permet de définir le volume de $(\Sigma, g(t))$ et la quantité $r(t) = \frac{1}{\text{vol}(\Sigma, g(t))} \int_{\Sigma} R(x, t) dv_{g(t)}$. On considère l'équation suivante :

$$(*_s) \begin{cases} \frac{\partial g}{\partial t} = (r - R)g \\ g(0) = g_0. \end{cases}$$

Il est facile de vérifier que le volume est constant en temps pour toute solution ; c'est donc une version normalisée du flot de la courbure que nous considérons. On peut également s'assurer immédiatement que toute solution $g(t)$ est conforme à g_0 , pour tout $t \geq 0$. Le résultat ci-dessous est prouvé en combinant [26] et [11].

THÉORÈME 1.2 ([26] et [11]). — *Soit (Σ, g_0) une surface riemannienne de classe C^∞ ; le problème $(*_s)$ admet une solution unique définie pour $t \in [0, +\infty)$. La famille de métriques $g(t)$ converge dans la topologie C^k , pour tout k , lorsque t tend vers $+\infty$, vers une métrique de courbure constante conforme à g_0 .*

Ce théorème peut être vu comme une nouvelle preuve de l'uniformisation des surfaces compactes. C'est le cas pour toutes les surfaces de caractéristique d'Euler négative ou nulle. Pour la sphère une étape de la preuve utilisait la structure complexe et ce n'est que récemment que ce problème fut résolu, dans [9], conduisant ainsi à une nouvelle preuve complète de l'uniformisation en dimension 2.

Dans le cas où la courbure est strictement négative, la preuve de la convergence est immédiate. Il serait alors intéressant d'avoir un procédé ou un algorithme simple permettant de munir une surface de caractéristique d'Euler négative (strictement) d'une métrique de courbure négative. Le flot donnerait, alors, une métrique de courbure constante de manière rapide (cette question a été posée à l'auteur par É. Ghys).

Le cas des orbifolds de dimension 2 est particulièrement instructif. Il est traité dans les articles [15], [60] et [10]. Les orbifolds compactes en dimension 2 sont classées en deux catégories (voir [50] ou [56]) : les « bonnes » (good) orbifolds qui admettent une métrique de courbure constante et les « mauvaises » (bad) orbifolds qui n'en admettent pas. Pour ces dernières, on montre que le flot $(*_s)$ converge vers un soliton de Ricci (voir [59]). La preuve du théorème 1.2 contient une étape dans laquelle on montre que les seuls solitons sur S^2 sont constants (en temps) égaux à une métrique de courbure constante (voir [13], théorème 5.21).

1.3. Conclusion

Ces exemples ont la vertu de mettre en œuvre l'utilisation des outils de base, présentés dans les sections suivantes, sous une forme élémentaire ; en cela ils constituent une bonne introduction aux travaux qui suivent. Ils ne sont toutefois pas complets en ce sens où, dans toutes ces situations, la courbure de l'objet considéré explose (c'est-à-dire tend vers $+\infty$) au même moment en tous les points. La situation étudiée par R. Hamilton, en dimension 4, et G. Perelman, en dimension 3, est radicalement différente car la courbure scalaire peut exploser dans certaines régions de la variété alors qu'elle reste bornée ailleurs. Il est alors nécessaire de décrire ces régions avec précision.

2. PREMIERS RÉSULTATS EN DIMENSION 3

C'est le point de départ de l'ensemble de travaux qui est l'objet de ce texte. C'est le premier résultat dû à R. Hamilton qui jette les bases de la méthode. On considère une variété compacte connexe et orientable de dimension 3, notée M , munie d'une métrique riemannienne lisse g_0 . On suppose le lecteur familier avec la notion de courbure de Ricci. Pour une métrique riemannienne g nous noterons Ric_g sa courbure de Ricci, qui est un 2-tenseur symétrique, dv_g (ou $d\text{vol}$) la mesure riemannienne et R sa courbure scalaire ; lorsque nous considérons une famille $g(t)$, $R(x, t)$ désigne la courbure scalaire de $g(t)$ au point $x \in M$. Une bonne référence pour les bases de la géométrie riemannienne est l'ouvrage [20]. On cherche alors une famille $g(t)$ de métriques lisses, dépendant de manière C^∞ de t , et solution du problème de Cauchy suivant :

$$(*_3) \begin{cases} \frac{\partial g}{\partial t} = -2 \text{Ric}_{g(t)} \\ g(0) = g_0. \end{cases}$$

Une version normalisée peut être écrite, pour laquelle le volume de la métrique qui évolue est fixé ; il suffit, en effet, de remplacer la première ligne de $(*_3)$ par :

$$\frac{\partial g}{\partial t} = -2 \text{Ric}_{g(t)} + \left(\frac{2}{3} \frac{1}{\text{vol}(M, g(t))} \int_M R(x, t) dv_{g(t)}(x) \right) g(t).$$

Nous noterons $(*_3')$ cette version. On passe d'une solution de $(*_3)$ à une solution de $(*_3')$ par une homothétie et un changement de temps. Nous appellerons flot de Ricci un couple $(M, g(t))$ solution de $(*_3)$. Le théorème fondateur de la théorie est le suivant :

THÉORÈME 2.1 (R. Hamilton, [24]). — *Soit M une variété compacte de dimension 3 qui admet une métrique de courbure de Ricci strictement positive. Alors, M admet une métrique de courbure sectionnelle constante (strictement positive). En particulier, M est difféomorphe à un quotient fini de S^3 .*

La preuve consiste à résoudre l'équation $(*_3)$ et à montrer que la solution renormalisée correspondante, c'est-à-dire la solution de $(*_3')$, est définie pour $t \in [0, +\infty)$ et converge vers une métrique de courbure constante. En fait la solution de $(*_3)$ ne vit que pendant un temps fini T et sa courbure explose en T . C'est le phénomène déjà observé pour $(*_c)$. Comme dans les modèles jouets, pour les solutions de l'équation $(*_3)$, la courbure explose en tous les points lorsque le temps s'approche de T .

La correspondance $g \longrightarrow -2 \text{Ric}_g + \left(\frac{2}{3} \frac{1}{\text{vol}(M, g)} \int_M R dv_g \right) g(t)$ peut être interprétée comme un champ de vecteurs sur l'espace des métriques riemanniennes sur M . Une trajectoire de celui-ci est alors une solution de $(*_3')$; on peut espérer qu'une telle trajectoire converge vers un point fixe de ce champ de vecteurs, c'est-à-dire une métrique d'Einstein, qui, en dimension 3, ne peut qu'être de courbure constante. Il s'agit bien entendu d'un point de vue heuristique. Toutefois, cette approche est développée dans [3].

Dans [25], R. Hamilton obtient des résultats concernant le cas où la courbure de Ricci est supposée positive ou nulle ; dans ce cas il démontre que M est difféomorphe à un

quotient de S^3 , $S^2 \times S^1$ ou \mathbf{R}^3 par un groupe d'isométries sans point fixe dans leur métrique standard.

Pour les extensions en dimension 4 on peut se référer à [25], [30] et à l'article récent [8] qui reprend les méthodes utilisées par G. Perelman pour corriger et étendre [30]. Un résultat optimal sous hypothèse de pincement ponctuel est obtenu dans [40]. Enfin, en dimension n quelconque, G. Huisken montre dans [33] un théorème comparable à 2.1 avec une hypothèse plus forte sur la courbure.

3. LES OUTILS D'ANALYSE

À partir de maintenant, $(M, g(t))$ est une solution du flot de Ricci non normalisé (équation $(*_3)$), définie sur un intervalle de temps $[0, T)$ (T fini ou non) et M est une variété différentielle compacte, connexe, orientable et de dimension 3. Par abus de langage, nous dirons que $(M, g(t))$ est un flot de Ricci. La majorité des résultats qui suivent est valable en dimension $n \geq 2$ à l'exception notable du théorème 3.4. Notons que la quasi-totalité de ceux-ci est due à R. Hamilton.

3.1. Existence de solutions en temps petit

Une des premières difficultés que l'on rencontre en étudiant les équations précédentes est l'existence en temps petit des solutions du problème de Cauchy. En effet, ces équations ne sont que faiblement paraboliques, c'est-à-dire que le membre de gauche de chacune d'elles est donné par un opérateur qui n'est pas fortement elliptique; son symbole a un noyau qui provient de l'invariance par difféomorphismes du problème. Montrer l'existence en temps petit des solutions devient alors une tâche délicate que R. Hamilton mène à bien dans [24]. Dans [17], D. DeTurck décrit une élégante façon de transformer les équations ci-dessus afin qu'elles deviennent fortement paraboliques; dans ce cas l'existence de solutions résultent de théorèmes classiques (voir [37]). Les solutions de l'équation transformée donnent des solutions de l'équation initiale par simple transport par un difféomorphisme dépendant du paramètre t . Une interprétation en terme d'applications harmoniques est donnée dans [29] et [13].

3.2. L'équation d'évolution des courbures

Un des outils de base est le principe du maximum. Il s'applique aux équations paraboliques qui régissent l'évolution des différents type de courbure. Pour simplifier, nous ne considérerons que la courbure scalaire et l'opérateur de courbure.

Rappelons que R désigne la courbure scalaire. On a alors,

$$\frac{\partial R}{\partial t} + \Delta_{g(t)} R = 2|\text{Ric}(g(t))|_{g(t)}^2$$

où $\Delta_{g(t)}$ désigne le laplacien, agissant sur les fonctions, défini par la métrique $g(t)$. La convention adoptée est celle dite des géomètres (en dimension 1, c'est $-d^2/dx^2$). La

norme du tenseur de Ricci est prise pour la métrique $g(t)$ (pour la norme d'un tenseur, le lecteur peut se reporter à [20]).

Nous noterons Rm l'opérateur de courbure, c'est-à-dire l'endomorphisme symétrique de $\Lambda^2(T^*M)$ dans lui-même ; alors,

$$\frac{\partial \text{Rm}}{\partial t} + \Delta \text{Rm} = \text{Rm}^2 + \text{Rm}^\sharp.$$

Nous utilisons ici le laplacien brut, c'est-à-dire l'opposé de la trace de la dérivée covariante seconde du tenseur Rm . La notation Rm^2 désigne le carré de l'endomorphisme Rm et Rm^\sharp désigne une expression quadratique en Rm que nous décrirons plus loin.

3.3. Existence en grand temps

Il est difficile de préciser, a priori, l'intervalle de vie d'une solution. Des estimations existent toutefois. Les équations paraboliques ci-dessus permettent de montrer que si la solution n'est définie qu'en temps fini, c'est-à-dire sur un intervalle $[0, T)$ avec $T < +\infty$, alors la courbure explose en T . Plus précisément,

PROPOSITION 3.1 ([13], p. 201). — *Soit g_0 une métrique C^∞ sur M ; alors l'équation $(*_3)$ a une unique solution sur un intervalle maximal, $0 \leq t < T \leq +\infty$. De plus, si $T < +\infty$, alors,*

$$\lim_{t \nearrow T} \left(\sup_{x \in M} |\text{Rm}(x, t)| \right) = +\infty.$$

La norme de $\text{Rm}(x, t)$ est la norme d'endomorphisme, calculée à l'aide de la métrique. La preuve se fait en constatant qu'une borne sur la courbure correspond à une borne sur les dérivées secondes de la métrique. En conséquence, si la courbure est bornée au voisinage de T on peut trouver une limite à $g(t)$ lorsque t tend vers T , et poursuivre le flot à partir de celle-ci, ce qui contredit la définition de T . Pour les détails le lecteur est renvoyé au chapitre 6 de [13].

3.4. Principes du maximum

C'est l'outil indispensable pour l'étude des solutions de l'équation de la chaleur. Il faut en donner ici une version vectorielle, c'est-à-dire une version pour les systèmes paraboliques. C'est ce que fait R. Hamilton dans [24], [25] et [31]. Le lecteur intéressé peut également consulter [13] et [14].

Considérons une équation aux dérivées partielles du type $\partial s / \partial t + \Delta_t s = F(s)$ et l'équation différentielle ordinaire $ds/dt = F(s)$. Les E.D.P. ci-dessus sont dites de **réaction-diffusion**, le terme de diffusion est donné par le laplacien ; si $F = 0$, il s'agit d'une équation de la chaleur qui « étale » les solutions. Le terme non-linéaire $F(s)$ est le terme de réaction qui, en l'absence de laplacien, conduit (souvent) à l'explosion en temps fini des solutions (convergence de certaines normes vers $+\infty$). La question est de savoir qui de la réaction ou de la diffusion l'emportera, dans une situation donnée.

Le principe du maximum résulte d’une comparaison entre le comportement des solutions de l’équation aux dérivées partielles et la situation extrême donnée par l’équation différentielle ordinaire.

L’équation satisfaite par la courbure scalaire peut s’écrire $\partial R/\partial t + \Delta R \geq 0$. Le principe du maximum scalaire le plus simple (voir [49]) conduit alors au théorème suivant :

THÉORÈME 3.2 (voir [13], lemme 6.8). — *Soit $(M, g(t))$ une solution de l’équation $(*_3)$; alors la fonction $R_{\min}(t) = \min\{R(x, t); x \in M\}$ est croissante.*

Nous décrivons maintenant une version vectorielle de ce principe. Soit M munie d’une famille C^∞ de métriques $g(t)$, pour $t \in [0, T]$, et soit $\pi : \mathcal{E} \rightarrow M$ un fibré vectoriel muni d’une métrique **fixe** et d’une famille C^∞ de connexions compatibles, ∇_t . Ces données permettent de définir un laplacien agissant sur les sections de \mathcal{E} , qui dépend de t et que nous noterons simplement Δ . Considérons une fonction $F : \mathcal{E} \times [0, T] \rightarrow \mathcal{E}$ de classe C^∞ (pour simplifier) telle que, pour t donné, $F(\cdot, t)$ préserve les fibres. Soit \mathcal{K} un fermé de \mathcal{E} que nous supposons invariant par le transport parallèle de ∇_t , pour tout $t \in [0, T]$, et tel que $\mathcal{K}_x = \mathcal{K} \cap \pi^{-1}(x)$ soit fermé et convexe. L’hypothèse clé est une relation entre \mathcal{K} et l’équation différentielle ordinaire $\frac{d}{dt}u = F(u)$, définie dans chaque fibre \mathcal{E}_x de \mathcal{E} ; nous supposons que toute solution de celle-ci telle que $u(0) \in \mathcal{K}_x$ reste dans \mathcal{K}_x pour tout $t \in [0, T]$.

THÉORÈME 3.3 ([25], [24], [29] ou [13], théorème 4.8). — *Avec les hypothèses ci-dessus, soit $s(t)$ une solution de l’E.D.P.*

$$\frac{\partial}{\partial t}s + \Delta s = F(s)$$

telle que $s(0) \in \mathcal{K}$, alors, pour tout $t \in [0, T]$, $s(t) \in \mathcal{K}$.

Pour appliquer ceci à l’opérateur de courbure il faut noter que, bien que la métrique du fibré $\Lambda^2(T^*M)$ dépende de t , une astuce, due à K. Uhlenbeck (voir [13], section 6.1), permet de se ramener à une métrique fixe sur un fibré fixe. En chaque point $x \in M$ et $t \in [0, T]$ l’endomorphisme Rm se diagonalise et possède trois valeurs propres notées $\lambda(x, t) \geq \mu(x, t) \geq \nu(x, t)$. Avec les normalisations standards (voir [13], chapitre 6) ces nombres sont égaux au double de la courbure sectionnelle du 2-plan correspondant et $R(x, t) = \lambda(x, t) + \mu(x, t) + \nu(x, t)$. On montre que les deux expressions Rm^2 et Rm^\sharp se diagonalisent dans la même base et ont pour valeurs propres respectives $(\lambda^2, \mu^2, \nu^2)$ et $(\mu\nu, \lambda\nu, \lambda\mu)$. L’équation différentielle ordinaire prend donc une forme très simple,

$$\begin{cases} \frac{d\lambda}{dt} = \lambda^2 + \mu\nu \\ \frac{d\mu}{dt} = \mu^2 + \lambda\nu \\ \frac{d\nu}{dt} = \nu^2 + \lambda\mu. \end{cases}$$

Appelons ϕ la fonction réciproque de $x \rightarrow x \log x - x$. La fonction ϕ est croissante et va de $[-1, +\infty)$ dans $[1, +\infty)$. En appliquant 3.3 à l’ensemble adéquat évident, on obtient

THÉORÈME 3.4 ([24], [25], [29] [34]). — Soit $(M, g(t))$ une solution de $(*_3)$ telle que, pour tout $x \in M$, $R(x, 0) \geq -1$ et $\nu(x, 0) \geq -\phi(R(x, 0))$; alors, pour tout $t \in [0, T]$ et pour tout $x \in M$, on a

$$\nu(x, t) \geq -\phi(R(x, t)).$$

La courbure scalaire étant la somme des valeurs propres de Rm on en déduit immédiatement que

COROLLAIRE 3.5. — Sous les mêmes hypothèses,

$$R(x, t) + 2\phi(R(x, t)) \geq \lambda(x, t) \geq \nu(t, x) \geq -\phi(R(x, t)).$$

C'est-à-dire, la courbure scalaire contrôle le tenseur de courbure. L'hypothèse, qui est une normalisation, est toujours réalisée à homothétie sur la métrique près. Ce résultat est connu sous le nom de théorème de pincement de Hamilton-Ivey. Insistons sur le fait qu'il n'est valable qu'en dimension 3 et est dû à la forme particulière du terme de réaction dans l'équation qui gouverne l'opérateur de courbure. En dimension supérieure, 4 par exemple, la situation est sensiblement plus compliquée (voir [25], [30] et [40]). Nous verrons plus loin une conséquence géométrique importante de cet élégant résultat d'analyse, qui repose sur le fait que $\phi(x)/x \rightarrow 0$ lorsque $x \rightarrow +\infty$. D'autres théorèmes de pincement peuvent être obtenus en utilisant une version où l'ensemble \mathcal{E} dépend de t (voir [31] et [14] pour un principe du maximum adapté à cette situation); ils utilisent d'autres fonctions ϕ . Nous dirons d'une variété riemannienne vérifiant la conclusion du théorème ci-dessus, pour une fonction ϕ , qu'elle est de **courbure ϕ -presque positive**.

3.5. Estimation des dérivées de la courbure

Le caractère régularisant des équations qui décrivent l'évolution de la courbure se traduit par le résultat suivant :

THÉORÈME 3.6 ([29]). — Soit $(M, g(t))$ un flot de Ricci, alors, pour tout $\alpha > 0$ et $k \geq 1$, il existe une constante C_k (qui dépend de k et α) telle que si

$$|\text{Rm}(x, t)|_{g(t)} \leq K, \quad \text{pour tout } x \in M \text{ et } t \in [0, \frac{\alpha}{K}],$$

alors

$$|\nabla^k \text{Rm}(x, t)|_{g(t)} \leq \frac{C_k K}{t^{k/2}}, \quad \text{pour tout } x \in M \text{ et } t \in (0, \frac{\alpha}{K}].$$

Les normes utilisées sont calculées à l'aide de la métrique $g(t)$ et ∇ désigne sa connexion de Levi-Civita. En d'autres termes, si le flot vit assez longtemps il régularise suffisamment de sorte à obtenir un contrôle C^k de la courbure, pour tout k . Ce résultat est valable en toute dimension. L'équation d'évolution satisfaite par la courbure scalaire R conduit alors à un contrôle de $\partial R / \partial t$. Ce type d'inégalités se prouve en montrant que la quantité à estimer vérifie une équation (ou une inéquation) d'évolution parabolique à laquelle on applique un principe du maximum.

Dans [52] et [53] (voir aussi [29], paragraphe 13), W.-X. Shi prouve une version locale de 3.6, très importante, dans laquelle l'hypothèse et les conclusions sont vérifiées sur

un voisinage d'un point de M . Une liste exhaustive de références peut être consultée dans [13] page 201.

3.6. Inégalités de Harnack

Ce sont des inégalités qui permettent de comparer la courbure en deux points de l'espace-temps $M \times [0, T)$. Dans [27], R. Hamilton prouve une inégalité de Harnack différentielle remarquable, dans l'esprit des résultats de P. Li et S. T. Yau ([38]). Il s'agit d'une inégalité, portant sur un tenseur, dont nous donnons ici la conséquence sur sa trace, dans un cas simple. Supposons que $(M, g(t))$ est à opérateur de courbure positif ou nul, alors, pour tout $t \in [0, T)$ et pour tout champ de vecteurs X sur M de classe C^∞ , on a :

$$H(X) = \frac{\partial R}{\partial t} + \frac{R}{t} + 2 \langle X, \nabla R \rangle + 2 \operatorname{Ric}(X, X) \geq 0.$$

Nous dirons que $(M, g(t))$ est une **solution antique**, si elle est définie sur un intervalle du type $(-\infty, T)$ (elle a un passé infini). En appliquant l'inégalité ci-dessus à la solution translatée en temps, $\tilde{g}(t) = g(t + t_0)$, pour $t \in [0, T - t_0)$, et en faisant tendre t_0 vers $-\infty$, nous obtenons, pour une telle solution,

$$\frac{\partial R}{\partial t} + 2 \langle X, \nabla R \rangle + 2 \operatorname{Ric}(X, X) \geq 0.$$

En particulier, en prenant $X = 0$, on prouve que, sous ces hypothèses, $\frac{\partial R}{\partial t} \geq 0$, c'est-à-dire que la courbure scalaire est croissante en tout point d'une solution antique d'opérateur de courbure positif ou nul. Si la solution n'est pas antique on peut vérifier que la quantité tR est croissante.

Une version intégrée s'énonce de la manière suivante (voir [33]) : si $(M, g(t))$ est à opérateur de courbure positif ou nul, alors, pour tout $t_1 < t_2$ et $x_1, x_2 \in M$,

$$R(x_2, t_2) \geq \exp\left(-\frac{d_{t_1}^2(x_1, x_2)}{2(t_2 - t_1)}\right) R(x_1, t_1),$$

où d_{t_1} désigne la distance associée à la métrique $g(t_1)$.

Signalons une interprétation géométrique de ces inégalités dans [12].

4. LES OUTILS GÉOMÉTRIQUES

Hormis les outils standards de la géométrie riemannienne, les démonstrations faites par R. Hamilton et G. Perelman font appel à la notion de convergence des variétés riemanniennes et nécessitent donc l'utilisation d'un théorème de compacité; celui-ci doit être adapté afin que les limites obtenues soient munies d'un flot de Ricci. Il est dû à R. Hamilton ([28]).

4.1. Le théorème de compacité

Décrivons d'abord la notion de convergence utilisée. Nous appellerons flot de Ricci marqué un quadruplet $\mathcal{M} = (M, g(t), O, F)$ où M est une variété connexe de dimension n , $g(t)$ un flot de Ricci sur M , défini pour $t \in (\alpha, \omega)$, O est un point de M et F est un repère orthonormé de $(M, g(0))$ en O . On suppose de plus que $-\infty \leq \alpha < 0 < \omega \leq +\infty$ et que $g(t)$ est complète pour chaque t ; nous dirons que le flot est complet.

DÉFINITION 4.1. — *On dit qu'une suite $\mathcal{M}_k = (M_k, g_k(t), O_k, F_k)$ converge vers $\mathcal{M}_\infty = (M_\infty, g_\infty(t), O_\infty, F_\infty)$ s'il existe*

i) *une suite d'ouverts $U_k \subset M_\infty$, contenant O_∞ et telle que tout compact de M_∞ soit contenu dans U_k pour k assez grand,*

ii) *une suite de difféomorphismes $\Phi_k : U_k \longrightarrow V_k \subset M_k$ envoyant (O_∞, F_∞) sur (O_k, F_k) , tels que $\Phi_k^* g_k$ converge vers g_∞ uniformément sur tout compact de $M \times (\alpha, \omega)$, ainsi que toutes leurs dérivées spatiales et temporelles.*

On a alors le théorème suivant :

THÉORÈME 4.2 ([28]). — *Soit \mathcal{M}_k une suite flots de Ricci marqués complets telle que,*

i) *pour tout $r > 0$ et tout $t \in (\alpha, \omega)$, il existe $C(r, t) < +\infty$ vérifiant, pour tout k , $\sup_{B_t(O_k, r)} \{|\text{Rm}(g_k(t))|\} \leq C(r, t)$, où $B_t(O_k, r)$ désigne la boule de centre O_k et de rayon r pour la métrique $g_k(t)$.*

ii) *Le rayon d'injectivité en O_k de la métrique $g_k(0)$ est borné inférieurement par une constante strictement positive indépendamment de k .*

Alors, il existe une sous-suite de \mathcal{M}_k qui converge vers un flot de Ricci marqué complet.

Notons que le choix des repères en O_k n'a pour but que de fixer les difféomorphismes locaux Φ_k . Par la suite nous ne mentionnerons plus cette donnée.

ESQUISSE DE PREUVE

L'hypothèse sur la courbure et celle sur le rayon d'injectivité de la métrique $g_k(0)$ permettent de faire converger, à une sous-suite près, la suite $(M_k, g_k(0))$ pour la convergence de Lipschitz pointée. La limite n'est, a priori, que $C^{1,s}$ pour $0 \leq s < 1$. Pour gagner de la régularité on utilise les estimés de W.-X. Shi (3.5); en effet, l'inégalité $\alpha < 0$ implique que les dérivées spatiales de $g_k(0)$ sont uniformément bornées, car en 0 le flot a déjà vécu pendant une durée au moins égale à $\alpha/2$. On construit alors une variété C^∞ limite, M_∞ , et les difféomorphismes locaux Φ_k . Les mêmes estimés permettent de montrer que la suite $g_k(t)$ est équicontinue et bornée (pour des normes canoniques) d'où leur convergence pour une sous-suite. Comme précédemment la convergence a lieu au sens C^p , pour tout $p \in \mathbf{N}$, et conduit à un flot de Ricci $g_\infty(t)$. ■

Dès lors, la principale difficulté est de borner inférieurement le rayon d'injectivité en O_k . Dans certaines circonstances R. Hamilton y parvient mais, comme nous le verrons plus loin, G. Perelman prouve un très joli théorème général dont c'est une conséquence.

4.2. Les dilatations paraboliques

Il s'agit d'une transformation de la solution de l'équation parabolique $(*_3)$ qui permet d'étudier à la « loupe » ce qui se passe au voisinage d'un point où la courbure scalaire est grande. C'est une technique classique pour étudier certaines E.D.P.

DÉFINITION 4.3. — Soient $(M, g(t))$ un flot de Ricci défini sur $[0, T]$, $t_0 \in [0, T]$ et $Q > 0$. On appelle dilatation parabolique en t_0 de rapport Q la famille de métriques,

$$\tilde{g}(t) = Qg\left(t_0 + \frac{t}{Q}\right).$$

Elle vérifie l'équation $(*_3)$ sur l'intervalle $[-Qt_0, (T - t_0)Q]$.

Il est immédiat de vérifier que la transformation ci-dessus produit un nouveau flot de Ricci et l'expression « dilatation parabolique » est justifiée par le fait que l'équation parabolique est préservée. Il s'agit d'une dilatation de la métrique combinée avec une contraction du temps (lorsque $Q > 1$) ainsi qu'un décalage de l'origine des temps.

Donnons un exemple d'utilisation. Soit $(M, g(t))$ un flot de Ricci défini sur un intervalle $[0, T)$ où $T < +\infty$ et soient deux suites $x_k \in M$ et $t_k \in [0, T)$ telles que $Q_k = R(x_k, t_k) \rightarrow +\infty$ lorsque $k \rightarrow +\infty$ (ceci ne peut se produire que si $t_k \rightarrow T$). Nous considérons la solution $\tilde{g}_k(t) = Q_k g(t_k + \frac{t}{Q_k})$, définie sur $[-Q_k t_k, (T - t_k)Q_k]$. Si Q_k vérifie des hypothèses supplémentaires (voir la sous-section 5.2) et si nous pouvons prouver que le rayon d'injectivité en x_k de $\tilde{g}_k(0)$ est minoré, indépendamment de k , alors, par le théorème de compacité précédent, il existe une sous-suite de $(M, \tilde{g}_k(t), x_k)$ convergeant vers un flot de Ricci marqué $(M_\infty, g_\infty(t), x_\infty)$. Ce dernier flot est défini sur $(-\infty, T_\infty)$ (où $T_\infty \geq 0$), il s'agit donc d'une solution antique qui représente un modèle infinitésimal (et donc un modèle local approximatif) aux points où la courbure scalaire explose.

5. LES RÉSULTATS DE G. PEREL'MAN I

Les références pour cette section et les suivantes sont les articles de G. Perel'man [44], [46] et [45] ainsi que les notes écrites par J. Lott et B. Kleiner [36], par N. Sesum, G. Tian et X. Wang [51], par P. Topping [58] et par R. Ye [61].

5.1. Le flot comme gradient d'une fonctionnelle

Le flot de la courbure extrinsèque d'une courbe (équation $*_c$) est donné par l'opposé du gradient de la longueur et celui de la courbure moyenne d'un corps convexe provient du volume. En revanche, le flot de la courbure de Ricci n'est *a priori* pas un flot de gradient. Une des premières originalités de [44] est de présenter le flot de Ricci de sorte à le faire apparaître comme un flot de gradient. Plus précisément, soient (M, g)

une variété riemannienne et f une fonction C^∞ sur M à valeurs réelles ; on définit la fonctionnelle,

$$\mathcal{F}(g, f) = \int (R + |\nabla f|^2) e^{-f} \, \text{dvol} .$$

En se limitant à l'ensemble des g et f telles que la forme volume $e^{-f} \, \text{dvol}$ soit constante, le gradient de la fonctionnelle \mathcal{F} conduit aux équations,

$$\frac{\partial g}{\partial t} = -2(\text{Ric} + \text{Hess}(f)) \quad \text{et} \quad \frac{\partial f}{\partial t} = -R + \Delta f .$$

Toutes les expressions qui interviennent sont calculées pour la métrique $g(t)$. En transportant la famille $g(t)$ par une famille de difféomorphismes dépendant de t , on se ramène aisément aux deux équations :

$$\frac{\partial g}{\partial t} = -2 \text{Ric} \quad \text{et} \quad \frac{\partial f}{\partial t} = -R + \Delta f + |\nabla f|^2 .$$

Elles n'ont en général pas de solutions car, en effet, l'équation portant sur la fonction f est une équation de la chaleur rétrograde. L'astuce pour obtenir une solution de ce système consiste à considérer une solution du flot de Ricci sur un intervalle $[0, T]$, qui existe pour une donnée initiale C^∞ , et à résoudre l'équation sur f de manière rétrograde, en imposant une donnée finale $f(T)$. Le long d'une telle famille $(g(t), f(t))$, la fonctionnelle \mathcal{F} est croissante et même strictement croissante, sauf si $\text{Ric} + \text{Hess}(f) \equiv 0$.

G. Perel'man introduit alors la fonctionnelle suivante, définie sur une variété M munie d'une métrique g ,

$$\mathcal{W}(g, f, \tau) = \int [\tau(R + |\nabla f|^2) + f - n](4\pi\tau)^{-\frac{n}{2}} e^{-f} \, \text{dvol} ,$$

où f est une fonction C^∞ et $\tau > 0$. Les variables de cette fonctionnelle sont contraintes par la relation $\int_M \frac{e^{-f}}{(4\pi\tau)^{n/2}} \, \text{dvol} = 1$. Dans les applications au flot de Ricci, le nombre réel τ vaut $T - t$ pour un choix judicieux de T . La fonctionnelle \mathcal{W} est interprétée comme une entropie par G. Perel'man (voir [44], section 5) et a un rapport étroit avec les inégalités de Sobolev logarithmiques. Le principe de monotonie affirme que, si g , f et τ vérifient

$$\frac{\partial g}{\partial t} = -2 \text{Ric}, \quad \frac{\partial f}{\partial t} = -R + \Delta f + |\nabla f|^2 + \frac{n}{2\tau}, \quad \frac{d\tau}{dt} = -1 ,$$

alors \mathcal{W} est croissante et strictement croissante sauf si $\text{Ric} + \text{Hess}(f) - \frac{1}{2\tau}g = 0$. Ensuite G. Perel'man définit l'invariant $\mu(g, \tau) = \inf_f \mathcal{W}(g, f, \tau)$ et montre que la fonction μ est bornée inférieurement lorsque τ varie dans un intervalle fini $(0, \tau_0]$. De plus, la monotonie de \mathcal{W} montre que $\mu(g(t), \tau_0 - t)$ est croissante le long d'un flot de Ricci $g(t)$.

Les détails, principaux exemples et commentaires sont parfaitement rédigés dans les notes [36], [51] et [58] et dépassent le cadre limité du présent rapport. Insistons toutefois sur l'intéressante relation entre la fonctionnelle \mathcal{W} et les inégalités de Sobolev logarithmiques sur les variétés riemanniennes.

5.2. Le non-effondrement local

Nous présentons ici le premier résultat important, que l'on peut trouver dans la section 4 de [44].

DÉFINITION 5.1. — *Soient M une variété de dimension n et g une métrique sur M ; on dit que g est κ -non-effondrée à l'échelle ρ , si toute boule métrique B de rayon $0 < r < \rho$, qui vérifie $|\text{Rm}(x)| \leq r^{-2}$ pour tout $x \in B$, a un volume supérieur à κr^n .*

Notons que $\lambda^2 g$ est κ -non-effondrée à l'échelle $\lambda \rho$ dès que g l'est à l'échelle ρ . Il est aussi immédiat de vérifier qu'une limite (au sens fort décrit ci-dessus) de métriques κ -non-effondrées à l'échelle ρ possède également cette propriété. Le résultat clé est le théorème suivant, valable en dimension n ,

THÉORÈME 5.2 (G. Perel'man, [44] section 4). — *Soit $(M, g(t))$ un flot de Ricci sur une variété compacte M , défini sur un intervalle $[0, T)$ avec $T < +\infty$; alors, il existe $\kappa = \kappa(g(0), T)$ tel que, pour tout $t \in [0, T)$, $g(t)$ est κ -non-effondrée à l'échelle $T^{1/2}$.*

Le résultat 4.1 de [44] n'est pas exactement énoncé de cette manière, néanmoins le théorème ci-dessus en résume l'esprit. L'échelle $T^{1/2}$ pourrait être remplacée par une autre constante. Plus précisément, comme nous l'a suggéré J. Lott, le théorème ci-dessus pourrait être énoncé sous la forme suivante : pour toute échelle $\rho > 0$, il existe $\kappa(\rho, g(0), T)$ tel que $g(t)$ est κ -non-effondrée à l'échelle ρ , pour tout $t \in [0, T)$.

IDÉE DE LA PREUVE

On utilise l'invariant μ de la sous-section 5.1. On montre que, s'il existe une suite de boules $B_k = B_{g(t_k)}(x_k, r_k)$, pour $r_k^2 \leq T$ et $t_k \rightarrow T$, telles que $r_k^{-n} \text{vol}(B_k, g(t_k)) \rightarrow 0$, alors $\mu(g(t_k), r_k^2) \rightarrow -\infty$. Il suffit pour cela d'exhiber une fonction f_k qui minimise $\mathcal{W}(g(t_k), \cdot, r_k^2)$. La monotonie de μ le long des trajectoires du flot de Ricci implique que $\mu(g(0), t_k + r_k^2) \xrightarrow[k \rightarrow +\infty]{} -\infty$, ce qui contredit la propriété énoncée en 5.1 car la suite $t_k + r_k^2$ est bornée (T est fini). ■

Un théorème de J. Cheeger (voir [5]) affirme que, sur l'ensemble des variétés riemanniennes compactes (M, g) de dimension n telles que $|\text{Rm}(g)| = \sup\{|\text{Rm}(x)|; x \in M\} \leq 1$, $\text{vol}(M, g) \geq K$ et $\text{diam}(M, g) \leq 1$, le rayon d'injectivité (global) est minoré par un nombre strictement positif $\epsilon(n, K)$ ne dépendant que de n et K . On peut prouver une version locale dans laquelle la variété (M, g) est remplacée par une boule de rayon 1. Considérons alors un flot de Ricci défini sur un intervalle $[0, T)$, avec $T < +\infty$ et posons $r(t) = |\text{Rm}(g(t))|^{-1/2}$. Pour tout t la métrique $\tilde{g}(t) = \frac{1}{r(t)^2} g(t)$ vérifie

- i) $B_{g(t)}(x, r(t)) = B_{\tilde{g}(t)}(x, 1)$ pour $x \in M$,
- ii) $|\text{Rm}(\tilde{g}(t))| = r(t)^2 |\text{Rm}(g(t))| \leq 1$ sur $B_{\tilde{g}(t)}(x, 1)$,
- iii) $\text{vol}(B_{\tilde{g}(t)}(x, 1), \tilde{g}(t)) \geq \kappa$, d'après le théorème 5.2.

Le résultat de J. Cheeger énoncé ci-dessus montre que, si $\text{inj}(x, g)$ désigne le rayon d'injectivité en x d'une métrique g , alors $\text{inj}(x, g(t)) = r(t) \text{inj}(x, \tilde{g}(t)) \geq r(t) \epsilon(n, \kappa)$, c'est-à-dire,

$$|\text{Rm}(g(t))| \text{inj}^2(x, g(t)) \geq \epsilon^2(n, \kappa),$$

pour tout $x \in M$ et $t \in [0, T)$. C'est une difficulté majeure dans l'étude du flot de la courbure de Ricci qui est levée par ce résultat sur le rayon d'injectivité.

Supposons maintenant que $T < +\infty$ est le temps maximal de vie de la solution ; alors $|\text{Rm}(g(t))| \rightarrow +\infty$ lorsque $t \rightarrow T$. Choisissons des suites $x_k \in M$ et $t_k \xrightarrow[k \rightarrow +\infty]{} T$, telles que

$$Q_k = |\text{Rm}(x_k, t_k)| = \sup\{|\text{Rm}(g(t))|; t \in [0, t_k]\} \longrightarrow +\infty,$$

et considérons la suite de métriques $g_k(t)$ obtenues par dilatation parabolique en t_k et de rapport Q_k . En pointant en x_k , elle admet une sous-suite convergente vers un flot de Ricci ; on peut, en effet appliquer le théorème 4.2. La limite est un flot de Ricci complet qui est une solution antique et κ -non-effondrée à toute échelle.

Donnons maintenant une autre version de ce phénomène, en dimension 3, en considérant une suite (x_k, t_k) telle que la courbure scalaire vérifie :

$$H_k = R(x_k, t_k) = \sup\{R(x, t); x \in M \text{ et } t \in [0, t_k]\} \longrightarrow +\infty.$$

La suite $h_k(t)$ obtenue par dilatation parabolique en t_k et de rapport H_k admet une sous-suite convergente (pointée en x_k) vers un flot de Ricci $(M_\infty, h_\infty(t))$ (pointé en x_∞), défini sur $(-\infty, 0]$. En effet, la courbure scalaire de $g(t)$ est minorée pour tout t car son minimum croît ; de plus, pour $t \leq t_k$ son maximum est majoré par $R(x_k, t_k)$. On en déduit que, pour $t \leq 0$ fixé, la courbure scalaire de $h_k(t)$ est bornée. Le corollaire du théorème 3.4 montre alors que les courbures sectionnelles de $h_k(t)$ sont bornées pour $t \leq 0$. Le rayon d'injectivité étant contrôlé par l'argument ci-dessus, nous pouvons appliquer le théorème de compacité 4.2. La limite est de nouveau une solution antique et κ -non-effondrée. Elle vérifie de plus

PROPOSITION 5.3. — *Pour tout $t \leq 0$, la métrique $h_\infty(t)$ est non plate et d'opérateur de courbure positif ou nul.*

PREUVE

Soit $y_\infty \in M_\infty$, il est limite d'une suite $y_k \in M$ et $\text{Rm}_\infty(y_\infty, t) = \lim_{k \rightarrow \infty} \text{Rm}_k(y_k, t)$, où Rm_k désigne l'opérateur de courbure de h_k . Or, pour $t \leq 0$,

$$\text{Rm}_k(y_k, t) = \frac{\text{Rm}(y_k, t_k + \frac{t}{H_k})}{H_k} \geq -\frac{\phi(R(y_k, t_k + \frac{t}{H_k}))}{H_k} \geq -\frac{\phi(H_k)}{H_k} \xrightarrow[k \rightarrow +\infty]{} 0,$$

d'où la positivité de la courbure de $h_\infty(t)$. De plus, comme la courbure scalaire de $h_k(t_k)$ est égale à 1 en x_k , celle de $h_\infty(0)$ est aussi égale à 1 en x_∞ ; or si une des métriques $h_k(t)$ était plate, elles le seraient toutes. ■

Une version plus élaborée du théorème 5.2 est prouvée dans la section 8 de [44] ; elle repose sur l'importante section 7 dans laquelle la fonctionnelle suivante est considérée. Si $(M, g(t))$ est un flot de Ricci, γ une courbe C^1 sur M et $0 \leq \tau_1 \leq \tau = T - t \leq \tau_2$, on définit

$$\mathcal{L}(\gamma) = \int_{\tau_1}^{\tau_2} \sqrt{\tau} (R(\gamma(\tau), \tau) + |\dot{\gamma}(\tau)|_{g(\tau)}^2) d\tau.$$

La théorie variationnelle de cette fonctionnelle est décrite en détail : formule de variation première, seconde, champs de Jacobi et une notion de volume associée, appelée volume réduit. La monotonie du volume réduit est prouvée et celui-ci est alors utilisé en lieu et place de la fonctionnelle μ . Les notions et inégalités introduites dans cette section 7 de [44] sont utilisées de manière essentielle dans [46].

5.3. Les κ -solutions

On rappelle que les variétés considérées sont de dimension 3 et orientables. Nous considérons maintenant une famille de solutions qui sont appelées à être les modèles locaux au voisinage des points où la courbure scalaire est grande. Elles sont déjà apparues dans la section précédente comme limites de flots renormalisés et le but de cette sous-section est d'en donner une classification.

DÉFINITION 5.4. — *Soit $\kappa > 0$. Une κ -solution est une solution C^∞ du flot de Ricci, $(M, g(t))$, définie pour $-\infty < t \leq 0$, telle que*

- i) *M est une variété de dimension 3 qui peut être non compacte,*
- ii) *pour chaque t , la métrique $g(t)$ est complète, d'opérateur de courbure positif ou nul, de courbure sectionnelle bornée supérieurement et non plate,*
- iii) *pour chaque t , $g(t)$ est κ -non-effondrée à toute échelle.*

Les premiers exemples de κ -solutions sont S^3 et $S^2 \times \mathbf{R}$ munies de leur flot standard, c'est-à-dire telles que la métrique au temps 0 est leur métrique canonique. En revanche, on vérifie que le flot standard sur $S^2 \times S^1$ n'est pas une κ -solution.

Nous pouvons utiliser les inégalités de Harnack différentielles et intégrées décrites dans la sous-section 3.6. L'inégalité intégrée prouve, en particulier, que la courbure scalaire d'une κ -solution est partout strictement positive ; en effet, s'il existe (x_2, t_2) tel que $R(x_2, t_2) = 0$ alors, pour tout $x_1 \in M$ et tout $t_1 < t_2$, on a $R(x_1, t_1) = 0$ et $g(t_1)$ est plate, ce qui est exclu.

Nous allons associer à chaque κ -solution, une autre solution appelée soliton.

DÉFINITION 5.5. — *On appelle soliton de Ricci, ou plus simplement soliton, un flot de Ricci $g(t)$ tel qu'il existe une famille de difféomorphismes ψ_t , dépendant de manière C^∞ de t , et une fonction $\alpha(t)$, C^∞ et strictement positive, vérifiant :*

$$g(t) = \alpha(t)\psi_t^*g(0).$$

On dit que le soliton est contractant (resp. dilatant) si α est strictement décroissante (resp. strictement croissante). Enfin, le soliton est dit de type gradient si $\frac{d\psi_{t+s}}{ds}|_{s=0}(x) = \nabla_{g(t)}f_t(\psi_t(x))$, où f_t est une famille de fonctions dépendant de manière C^∞ des deux variables x et t .

Les solitons sont des points fixes du flot vu sur l'espace des métriques riemanniennes modulo l'action des difféomorphismes. Ils généralisent la notion de métrique d'Einstein

et sont des solutions auto-similaires. Un calcul immédiat montre qu'un soliton de type gradient vérifie l'équation (voir [58]) :

$$\operatorname{Ric}_{g(t)} + \operatorname{Hess}_{g(t)} f_t + \frac{\alpha'(t)}{2\alpha(t)} g(t) = 0.$$

Soit alors une κ -solution $(M, g(t))$; G. Perel'man prouve la proposition suivante :

PROPOSITION 5.6 ([44], 11.2). — *Il existe une suite (x_k, t_k) telle que la famille de métriques obtenue à partir de $g(t)$ par dilatation parabolique en t_k et de rapport $-1/t_k$, pointée en x_k converge, lorsque t_k tend vers $-\infty$, vers un soliton contractant de type gradient et non-plat. Nous l'appellerons soliton asymptotique de la κ -solution et le noterons $M_{-\infty}$.*

Un tel soliton n'est, a priori, pas unique, mais, par abus de langage, nous parlerons du soliton asymptotique. La preuve met en œuvre toute la théorie développée dans le chapitre 7 de [44] et, en particulier, l'outil important qu'est le volume réduit. Les nombreux détails ne peuvent être décrits dans ce rapport, le lecteur intéressé peut se reporter à [36], [51] et [61]. L'idée est maintenant de classifier les κ -solutions en classifiant leur soliton asymptotique. On montre qu'un soliton asymptotique est lui-même une κ -solution (voir [36], 40.3).

Considérons un tel soliton asymptotique.

1) Dans le cas où $M_{-\infty}$ **est non compact**, il ne peut avoir sa courbure sectionnelle constamment strictement positive d'après le théorème suivant prouvé dans [46], section 1 :

THÉORÈME 5.7. — *Il n'existe pas de soliton contractant de type gradient non compact de dimension 3, orienté, complet et de courbure sectionnelle strictement positive et bornée.*

Par ailleurs, si son tenseur de courbure a un noyau, pour une valeur de t , un théorème de R. Hamilton ([25]) montre que celui-ci est invariant par transport parallèle et en temps ; c'est un principe du maximum fort pour les équations paraboliques vectorielles. On prouve alors que le revêtement universel de la variété se scinde en un produit métrique $N \times \mathbf{R}$, où N est une κ -solution de dimension 2. G. Perel'man prouve dans [44], 11.3, que la seule κ -solution orientée de dimension 2 est la sphère munie de son flot de courbure constante (nous dirons la sphère ronde). Notons qu'il invoque un résultat de R. Hamilton dans [26], qui n'est malheureusement valable que pour le cas compact ; il faut donc se reporter à [36] ou bien [61] pour compléter la preuve. Le soliton est donc un quotient non-compact et orientable de $S^2 \times \mathbf{R}$, c'est-à-dire, $S^2 \times \mathbf{R}$ lui-même, muni de son flot canonique (que nous appellerons cylindrique), ou $S^2 \times_{\mathbf{Z}_2} \mathbf{R}$, où l'action de \mathbf{Z}_2 est donnée par la relation $(x, s) \sim (-x, -s)$, la variété est alors difféomorphe à $\mathbf{R}P^3 \setminus \bar{B}^3$ (où B^3 est difféomorphe à une boule euclidienne ouverte de dimension 3 et \bar{B}^3 est son adhérence).

2) Si $M_{-\infty}$ est **compact** et s'il se scinde localement, ce doit être un quotient de $S^2 \times S^1$ muni de son flot canonique (par l'argument précédent), or ce dernier n'est pas une κ -solution. Il ne reste donc que le cas où l'opérateur de courbure est strictement positif pour tous les temps. Le théorème 2.1 montre que le soliton doit tendre à être de courbure constante et, comme il est auto-similaire, il est de courbure constante. La variété $M_{-\infty}$ est un quotient de S^3 munie de sa métrique canonique.

En résumé, si $M_{-\infty}$ est compact c'est un quotient de la sphère canonique, et s'il est non compact c'est $S^2 \times_{\mathbf{Z}_2} \mathbf{R}$ ou $S^2 \times \mathbf{R}$. Pour classer les κ -solutions nous aurons recours au théorème de R. Hamilton, prouvé dans [25], qui montre qu'un flot de Ricci compact en dimension 3 et qui est de courbure sectionnelle positive ou nulle est de l'un des trois types suivants :

i) la métrique $g(t)$ est plate pour tout t ,

ii) le flot se scinde localement en un produit $N \times \mathbf{R}$, où N est une surface de courbure strictement positive,

iii) pour tout t , $g(t)$ est de courbure sectionnelle strictement positive.

1) Si $(M, g(t))$ est une **κ -solution compacte**, elle ne peut ni être plate (par hypothèse) ni être recouverte par $N \times S^1$ (qui n'est pas une κ -solution), elle est donc de courbure strictement positive et, par le théorème 2.1, est difféomorphe à un quotient fini de S^3 . On peut encore subdiviser ce cas en deux sous-cas :

i) le soliton asymptotique est compact. Celui-ci est alors, d'après la discussion précédente, le flot canonique sur un quotient de S^3 munie de sa métrique canonique. La κ -solution ressemble lorsque le temps se rapproche de $-\infty$ à ce soliton. Or, le théorème 2.1 montre que le flot tend à rendre la variété de courbure constante; si elle l'est en $-\infty$, elle le sera pour tout temps. Il est aisé de formaliser cette observation. Dans ce cas donc, la κ -solution est le flot canonique, de courbure constante pour chaque t , sur un quotient de S^3 . Notons que ce quotient peut être arbitrairement petit, au sens où il peut être un espace lenticulaire.

ii) Le soliton asymptotique est non-compact. Nous traitons ce cas ci-dessous car il réclame une autre construction.

2) Si la **κ -solution est non-compacte** elle peut être localement scindée ou de courbure sectionnelle strictement positive.

i) Si elle est localement scindée, comme précédemment, elle est isométrique au flot canonique sur $S^2 \times \mathbf{R}$ ou bien $S^2 \times_{\mathbf{Z}_2} \mathbf{R}$. Topologiquement c'est un cylindre ou bien $\mathbf{R}P^3 \setminus \bar{B}^3$.

ii) Si elle est de courbure sectionnelle strictement positive, elle est difféomorphe à une boule B^3 (ou, ce qui revient au même, à \mathbf{R}^3); c'est, en effet, une conséquence d'un théorème dû à D. Gromoll et W. Meyer (voir [5], chapitre 8 et [23], pour un survol). C'est un cas que l'on précise également ci-dessous. En résumé seuls restent à décrire un peu plus précisément les cas de courbure strictement positive, les autres tombant sous le coup du scindement. Si $(M, g(t))$ est une κ -solution et $(x_0, t_0) \in M \times (-\infty, 0]$, on

définit la κ -solution normalisée en (x_0, t_0) par :

$$g_0(t) = R(x_0, t_0)g\left(t_0 + \frac{t}{R(x_0, t_0)}\right).$$

C'est une dilatation parabolique de sorte que $R_{g_0}(x_0, 0) = 1$. On a alors,

THÉORÈME 5.8 ([44], 11.7). — *Pour tout $\kappa > 0$, l'ensemble des κ -solutions normalisées est compact pour la topologie de la convergence pointée.*

La topologie est celle suggérée dans l'énoncé du théorème de compacité 4.2.

DÉFINITION 5.9. — *Soient (M, g) une variété riemannienne de dimension 3 et $\epsilon > 0$. Une partie $T \subset M$ est appelée une ϵ -gorge (« neck » chez Hamilton et Perel'man) centrée en $x \in T$ s'il existe un difféomorphisme $\Psi : T \rightarrow S^2 \times (-1/\epsilon, 1/\epsilon)$, tel que $\Psi(x) \in S^2 \times \{0\}$ et tel que l'image par Ψ de la métrique normalisée $R(x)g$ (où R désigne la courbure scalaire de g) est à distance ϵ , dans la topologie C^N , du produit de la métrique de courbure scalaire constante égale à 1 sur S^2 par la métrique usuelle de l'intervalle.*

Nous dirons que $R(x)g$ est ϵ -proche du cylindre. L'expression « ϵ -gorge » est un raccourci pratique pour « gorge de taille $2/\epsilon$ ». Le nombre N est grand et de l'ordre de $[1/\epsilon]$.

Revenons à la κ -solution $(M, g(t))$. Pour $\epsilon > 0$ et pour un temps t , on note $M_\epsilon(t)$ l'ensemble des points qui ne sont pas centres d'une ϵ -gorge. L'étude des κ -solutions se fait en contrôlant la géométrie des ensembles $M_\epsilon(t)$. On procède *grosso modo* toujours de la même manière ; si ces parties ne vérifient pas les propriétés souhaitées, on trouve une suite de points et de temps tels que les métriques correspondantes convergent, par le théorème de compacité ci-dessus, vers une κ -solution qui possède une droite (*i.e.* une géodésique paramétrée par \mathbf{R} et minimisante sur toute sa longueur) ; on en déduit que la limite obtenue se scinde en un cylindre et, par conséquent, juste avant la limite, les points choisis pour marquer la suite sont dans des ϵ -gorges, contredisant ainsi le fait qu'ils ont été choisis dans $M_\epsilon(t)$. On différencie deux cas :

i) M_ϵ est contenu dans deux boules disjointes séparées par un tube ; c'est le cas lorsque la κ -solution est compacte. Elle est alors difféomorphe à S^3 ou $\mathbf{R}P^3$. On pourrait s'attendre à ce qu'apparaissent également $\mathbf{R}P^3 \# \mathbf{R}P^3$; ceci n'est pas possible car nous sommes dans une situation où la variété est difféomorphe à un quotient de S^3 et a donc un groupe fondamental fini.

ii) M_ϵ est compact et inclus dans une boule en dehors de laquelle la variété est un tube ; on sait déjà que la κ -solution est difféomorphe à une boule B^3 , elle est munie de métriques telles, qu'en dehors d'un compact, tout point est le centre d'une ϵ -gorge.

DÉFINITION 5.10. — *On appelle capuchon de taille $2/\epsilon$, ou, plus simplement, ϵ -capuchon (« cap » chez Perel'man), une métrique sur une boule euclidienne de dimension 3, B^3 , ou sur $\mathbf{R}P^3 \setminus \bar{B}^3$, telle que tout point en dehors d'un compact est le centre d'une ϵ -gorge et telle que la courbure scalaire soit bornée.*

Résumons ce que produit cette étude (voir [44], [36] et [51] pour les nombreux détails). Pour un flot de Ricci, un point x et un temps t , nous noterons $B(x, t, r)$ la boule de centre x et de rayon r pour la métrique $g(t)$.

THÉORÈME 5.11 ([46] section 1.5 et [36] section 53). — *Il existe $\epsilon_0 > 0$ tel que, pour tout $0 < \epsilon \leq \epsilon_0$, il existe une constante $C_1 = C_1(\epsilon) > 0$ telle que, pour toute κ -solution $(M, g(t))$ et tout (x, t) , il existe un rayon $r \in [\frac{1}{C_1 R(x, t)^{1/2}}, \frac{C_1}{R(x, t)^{1/2}}]$ et un voisinage U , $B(x, t, r) \subset U \subset B(x, t, 2r)$, qui vérifie une des assertions suivantes :*

- a) U est une ϵ -gorge,
- b) U est un ϵ -capuchon,
- c) U est une variété compacte difféomorphe à S^3 ou $\mathbf{R}P^3$ et de courbure strictement positive,
- d) U est une variété compacte de courbure sectionnelle constante strictement positive.

On peut même contrôler la courbure scalaire et minorer le volume dans les cas a), b) et c), par des quantités de l'ordre de $R(x, t)$, de même que la courbure sectionnelle dans le cas c). De plus, en utilisant le soliton asymptotique et le théorème de compacité on prouve aussi qu'il existe $\kappa_0 > 0$ et une constante universelle $\eta > 0$ tels que toute κ -solution soit ou bien un quotient de la sphère canonique, ou bien une κ_0 -solution, et qu'en chacun de ses points on ait

$$(5.3.1) \quad |\nabla R| \leq \eta R^{3/2}, \quad \left| \frac{\partial R}{\partial t} \right| \leq \eta R^2.$$

On a exclu les quotients de la sphère canonique de la possibilité d'être une κ_0 -solution, car ils peuvent être très petits, des espaces lenticulaires par exemple pour lesquels la constante κ n'est pas minorée. Ces dernières inégalités permettent de contrôler la courbure dans un voisinage en espace-temps d'un point donné.

5.4. Le théorème des voisinages canoniques

C'est le point culminant de [44], le résultat qui permet de pratiquer la chirurgie autorisant la poursuite du flot malgré les singularités. Utilisons, pour simplifier, une définition donnée dans [44]

DÉFINITION 5.12. — *On appelle voisinage parabolique, noté $P(x, t, r, \Delta t)$, l'ensemble des points (x', t') avec $x' \in B(x, t, r)$ et $t' \in [t, t + \Delta t]$ ou $t' \in [t + \Delta t, t]$ suivant le signe de Δt .*

On a montré dans la sous-section 5.2 que, si l'on regarde une solution du flot de Ricci au voisinage d'un maximum (en temps et en espace) de la courbure scalaire, alors la métrique ressemble à (converge vers) une κ -solution. On prouve alors un résultat plus précis qui permet de décrire les régions de grande courbure scalaire qui ne correspondent pas nécessairement à des maxima

THÉORÈME 5.13 ([44], 12.1). — *Soit $\epsilon > 0$, $\kappa > 0$; il existe $r_0 > 0$ tels que, si $(M, g(t))$ est un flot de Ricci défini pour $t \in [0, T]$, de courbure ϕ -presque positive, sur une variété compacte de dimension 3, κ -non-effondrée à l'échelle r_0 , alors, pour tout (x_0, t_0) , $t_0 \geq 1$ et $Q = R(x_0, t_0) \geq r_0^{-2}$, le voisinage parabolique $P(x_0, t_0, \frac{1}{\sqrt{\epsilon Q}}, -\frac{1}{\epsilon Q})$ est, après dilatation parabolique en t_0 de rapport Q , ϵ -proche du voisinage correspondant d'une κ -solution.*

La proximité de deux métriques est celle utilisée dans la définition d'une gorge. On dit qu'un voisinage parabolique est proche d'une solution si le flot sur l'intervalle considéré est proche du flot sur la solution pour un intervalle de temps identique. L'hypothèse $t_0 \geq 1$ est présente car il est nécessaire que le flot ait vécu assez longtemps (disons 1 seconde) afin de régulariser les voisinages des points de grande courbure. On peut toujours se ramener à cette hypothèse par dilatation parabolique qui rend la courbure petite au temps 0; les estimés de W.-X. Shi montrent alors que l'explosion ne survient pas avant 1 seconde.

Nous avons simplifié l'énoncé en omettant un paramètre qui est la fonction ϕ de l'inégalité de Hamilton-Ivey; le nombre r_0 dépend du choix de cette fonction (dont nous avons donné un exemple dans le théorème 3.4). Ceci, combiné avec le théorème de structure des κ -solutions, donne une description, comme dans 5.11, des voisinages des points de grande courbure.

Insistons sur le fait que ce théorème affirme que l'explosion a toujours lieu le long de sous-variétés difféomorphes à S^2 .

IDÉE DE LA PREUVE

La principale difficulté vient du fait que nous ne travaillons pas en un maximum de la courbure scalaire. La preuve est faite par contradiction. On suppose qu'il existe $\epsilon > 0$, $\kappa > 0$, une suite de flots de Ricci (de dimension 3), notés $(M_k, g_k(t))$, de courbure ϕ -presque positive, définis sur un intervalle $[0, T_k]$, $\hat{x}_k \in M_k$, $\hat{t}_k \geq 1$, $r_k \xrightarrow[k \rightarrow +\infty]{} 0$, tels que M_k est κ -non-effondrée aux échelles inférieures à r_k , $Q_k = R(\hat{x}_k, \hat{t}_k) \geq r_k^{-2}$ et le voisinage parabolique correspondant n'est pas ϵ -proche d'une κ -solution (après changement d'échelle de rapport Q_k). Si un tel point de l'espace temps (\hat{x}_k, \hat{t}_k) est appelé un mauvais point, on cherche parmi les mauvais points ceux qui ont (presque) la plus grande courbure scalaire. On trouve une suite (x_k, t_k) de mauvais points tels que tout point (y, t) , $t \leq t_k$, vérifiant $R(y, t) \geq 2R(x_k, t_k)$ est bon. Les estimés du gradient de la courbure scalaire en espace montrent que, dans un voisinage de ces points de taille contrôlée, la courbure scalaire est comparable à $R(x_k, t_k)$. On peut alors prendre une limite sur ces voisinages en renormalisant les métriques comme dans la sous-section 5.2. Il faut ensuite montrer que la limite de la suite renormalisée existe au sens de la convergence pointée, c'est-à-dire sur toute boule. Par ailleurs, l'estimé sur la dérivée de la courbure de la sous-section précédente montre que cette limite possède un « bout » de flot vers les temps négatifs et il faut là encore montrer qu'elle est munie d'un flot de Ricci qui est une solution antique. En conclusion on prouve que c'est une κ -solution,

ce qui contredit le choix des points (x_k, t_k) . Le théorème de compacité 4.2, dans des formes plus générales, est utilisé plusieurs fois à chaque étape. ■

6. LES TRAVAUX DE PEREL'MAN : LE FLOT AVEC CHIRURGIES

L'article [46] représente un travail techniquement difficile et qui utilise toutes les notions déjà introduites dans [44]. Dans ce rapport nous ne retiendrons que la construction du flot avec chirurgies, qui permet de franchir les singularités. C'est une idée inventée par R. Hamilton – pour la dimension 4 dans [30] (voir aussi [29]) – qui a fait une grande partie du travail dont une brève description suit. La seule faiblesse de l'argument développé dans [30] est que la chirurgie ne se pratique qu'aux points de courbure (presque) maximale. Quelques arguments sont incomplets, en particulier la preuve de la non-accumulation des chirurgies. Pour ce qui concerne la dimension 4, ceci est traité dans la récente prépublication [8], dans laquelle la méthode développée par G. Perelman, dans [46], est adaptée au contexte.

6.1. Premier temps singulier

Considérons un flot de Ricci $(M, g(t))$ défini sur un intervalle maximal $[0, T)$, où T est fini et M est une variété compacte, connexe, sans bord, orientée et de dimension 3. Nous supposons $g(0)$ normalisée pour satisfaire les hypothèses de 3.4. La courbure de $g(t)$ explose au voisinage de T (c'est-à-dire, son supremum tend vers $+\infty$). Par le théorème des voisinages canoniques, pour ϵ assez petit, il existe $r = r(\epsilon) > 0$ tel que tout point (x, t) vérifiant $R(x, t) \geq r^{-2}$ a un voisinage qui est soit une ϵ -gorge, soit un ϵ -capuchon, soit une variété compacte de courbure sectionnelle strictement positive. Dans ce dernier cas, par connexité, M est entièrement contenue dans ce voisinage et est une variété de courbure strictement positive; par le théorème 2.1, M peut être munie d'une métrique de courbure constante et est donc un quotient de la sphère S^3 par un groupe d'isométries de la métrique canonique.

Dans les autres cas, les voisinages canoniques sont des gorges ou bien des capuchons. On appelle Ω l'ensemble des points où la courbure de $g(t)$ reste bornée lorsque t tend vers T . Notons que Ω n'est pas nécessairement connexe mais, grâce à (5.3.1), on montre qu'il est ouvert dans M . Au premier temps singulier, l'ensemble Ω ne peut pas avoir de composante compacte; en effet, cette composante serait ouverte et fermée, et M étant connexe, elle serait égale à M en entier; mais alors la courbure sur M serait bornée, ce qui est incompatible avec la définition de T . Donc $M \setminus \Omega$ est l'ensemble des points où la courbure tend vers l'infini lorsque t tend vers T et est non vide. L'ensemble, $M \setminus \Omega$ peut être très compliqué; en effet, le théorème des voisinages canoniques affirme que l'explosion a lieu dans des gorges, le long de sphères et celles-ci pourraient très bien s'accumuler.

L'ensemble Ω peut être vide. Si c'est le cas, en prenant un temps t très proche de T , on constate que la variété M est entièrement recouverte par des ϵ -gorges et des ϵ -capuchons

de courbure bornée. Lorsque deux ϵ -gorges s'intersectent, leur réunion est difféomorphe à $S^2 \times \mathbf{R}$, si ϵ est choisi assez petit. De même, si une ϵ -gorge intersecte un ϵ -capuchon, leur réunion est difféomorphe à B^3 ou $\mathbf{R}P^3 \setminus \bar{B}^3$. Par ailleurs les gorges et les capuchons sont de diamètre minoré, donc, pour $t < T$, on peut choisir un recouvrement fini par de tels ensembles. Cela montre que, dans ce cas, M est difféomorphe à S^3 , $\mathbf{R}P^3$, $S^2 \times S^1$ ou bien $\mathbf{R}P^3 \# \mathbf{R}P^3$. C'est un cas où la variété disparaît en un temps fini.

Si Ω est non vide, la métrique $g(t)$ converge sur Ω vers une métrique de classe C^∞ notée $g(T)$ (pour prouver ce fait il faut utiliser les estimés de W.-X. Shi). Le théorème des voisinages canoniques s'applique, par continuité, à $(\Omega, g(T))$, en changeant éventuellement la constante ϵ . Pour un nombre $\rho < r$, considérons l'ensemble $\Omega(\rho) = \{(x, t); R(x, T) \leq \rho^{-2}\}$. Si $\Omega(\rho)$ est vide, on conclut comme dans le cas où Ω est vide; on suppose donc $\Omega(\rho)$ non vide. Le nombre ρ est de la forme δr , où $\delta > 0$ est un paramètre à choisir, qui permet d'avoir de la marge pour les opérations qui suivent. Les inégalités 5.3.1 montrent que Ω_ρ est compact. Nous allons décrire $\Omega \setminus \Omega_\rho$. La définition qui suit s'applique à des chirurgies ultérieures, c'est pourquoi ii) mentionne une composante compacte qui n'existe pas au premier temps singulier.

DÉFINITION 6.1. — *Dans l'ensemble Ω , on appelle*

i) ϵ -tube, une sous-variété difféomorphe à $S^2 \times I$ (I est un intervalle) dont chaque point est le centre d'une ϵ -gorge dans Ω ,

ii) ϵ -tore, une composante de Ω qui est une variété fermée dont chaque point est le centre d'une ϵ -gorge. Un ϵ -tore est difféomorphe à $S^2 \times S^1$,

iii) ϵ -pointe (« horn » chez G. Perelman), un sous-ensemble de Ω difféomorphe à $S^2 \times [0, 1)$ dont le bord est contenu dans $\Omega(\rho)$ et dont chaque point est le centre d'une ϵ -gorge. La courbure scalaire de $g(T)$ tend vers $+\infty$ à l'autre bout,

iv) double ϵ -pointe, une composante connexe de Ω difféomorphe à $S^2 \times (0, 1)$ dont chaque point est le centre d'une ϵ -gorge. La courbure scalaire de $g(T)$ tend vers l'infini aux deux bouts,

v) ϵ -pointe encapuchonnée, une composante de Ω difféomorphe à une boule ou à $\mathbf{R}P^3 \setminus \bar{B}^3$ dont chaque point est soit dans une ϵ -gorge, soit dans un ϵ -capuchon. La courbure scalaire tend vers l'infini au bout.

Dans la suite nous pourrions omettre la mention à ϵ pour alléger le discours. Nous travaillons au temps T . Partons d'une gorge ou d'un capuchon dans $\Omega \setminus \Omega(\rho)$, un point de son bord est contenu dans $\Omega(\rho)$, dans une gorge ou bien un capuchon adjacent à la première gorge. Si c'est une gorge, en poursuivant ce procédé on ne peut s'arrêter que lorsque l'on rencontre un capuchon ou bien $\Omega(\rho)$; s'il se poursuit indéfiniment on obtient une pointe; notons qu'on ne sait rien du diamètre des pointes qui peut être fini ou infini. On peut résumer ceci par l'assertion suivante :

FAIT 6.2. — *Tout ϵ -gorge ou ϵ -capuchon de $\Omega \setminus \Omega(\rho)$ est contenu dans l'un des ensembles suivants :*

a) un tube dont les bords sont dans $\Omega(\rho)$,

- b) *un capuchon dont le bord est dans $\Omega(\rho)$,*
- c) *une pointe dont le bord est dans $\Omega(\rho)$,*
- d) *une pointe encapuchonnée,*
- e) *une double pointe.*

Rappelons que nous avons exclu les composantes compactes de Ω pour ce premier temps singulier. On remarque qu'un point x du bord de $\Omega(\rho)$ est contenu dans une ϵ -gorge, c'est-à-dire un ensemble presque isométrique à $S^2 \times (-\frac{1}{\epsilon\sqrt{\rho}}, \frac{1}{\epsilon\sqrt{\rho}})$, dont les sphères sont de courbure presque égale à ρ . Dans ce cylindre la courbure peut osciller autour de ρ^{-2} , il se pourrait donc qu'il se décompose en une infinité de petits cylindres de courbure scalaire supérieure à ρ^{-2} séparés par d'autres de courbure scalaire inférieure à cette valeur. Ceci montre que l'affirmation, figurant dans [46] page 7, selon laquelle les ϵ -tubes dont les bords rencontrent $\Omega(\rho)$ ont un volume minoré n'est pas correcte, mais cela est sans importance pour la suite. Les ensembles de type b) ont un volume minoré en fonction de ρ d'après 5.11 et ceux de type c), dont la courbure tend vers l'infini au bout, contiennent une gorge de courbure de l'ordre de $2\rho^{-2}$; leur volume est donc également minoré par une fonction de ρ . Par ailleurs le volume de $(M, g(t))$ est majoré indépendamment de t sur l'intervalle $[0, T)$ car la courbure scalaire est minorée (le minimum est croissant) et,

$$\frac{d \operatorname{vol}(M, g(t))}{dt} = - \int_M R(x, t) dv_{g(t)} .$$

On en déduit que le volume de $(\Omega, g(T))$ est fini et, par conséquent, qu'il n'y a qu'un nombre fini de composantes de Ω contenant des points de $\Omega(\rho)$ et chacune d'elles a un nombre fini de bouts, tous de type c) (un capuchon n'est pas un bout et un tube relie deux parties de $\Omega(\rho)$). Les autres composantes de Ω sont de types d) et e).

Lorsque t se rapproche de T on peut voir apparaître un chapelet de doubles pointes avec d'un côté une pointe reliée à $\Omega(\rho)$ et de l'autre une pointe encapuchonnée ou bien se terminant des deux côtés par une pointe reliée à $\Omega(\rho)$. Le nombre de doubles pointes peut être infini. Pour un temps fixé juste avant l'explosion en T , la variété est de diamètre fini et les gorges de diamètre contrôlé par la courbure; on peut donc recouvrir $\Omega \setminus \Omega_\rho$ par un nombre fini de gorges et de capuchons. Les chapelets précédents proviennent donc de tubes bouchés soit par des boules ou des ensembles difféomorphes à $\mathbf{R}P^3 \setminus \bar{B}^3$, dans le premier cas, soit par des tubes reliant deux parties de l'ensemble qui converge vers $\Omega(\rho)$, dans le second cas (figures 1 et 2 ci-dessous).

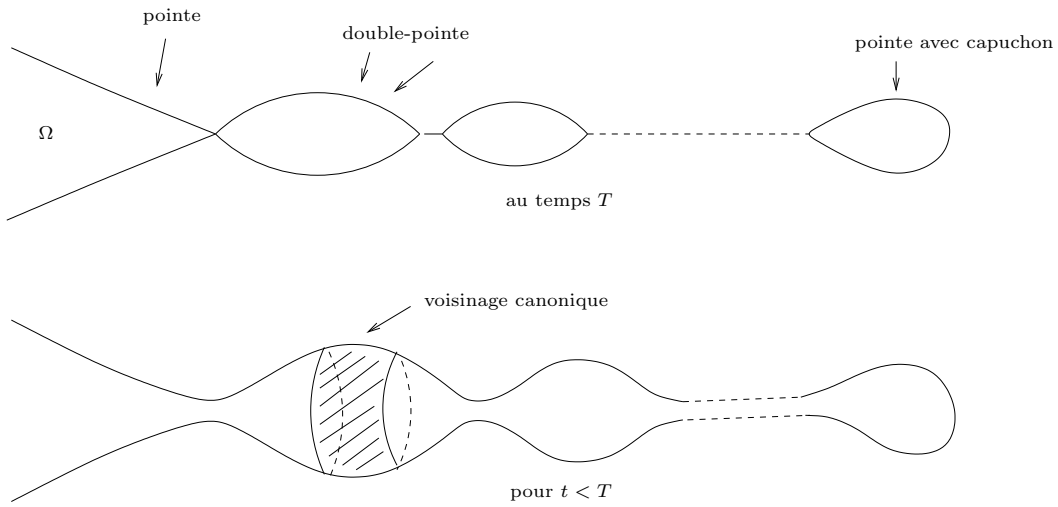


Fig. 1

La topologie de M peut être reconstituée comme suit : on considère la liste Ω_j , $1 \leq j \leq i$, des composantes connexes de Ω qui contiennent des points de $\Omega(\rho)$, on tronque les pointes et on colle sur les bords obtenus des tubes $S^2 \times I$ (I est un intervalle), ou des boules B^3 ou bien des ensembles difféomorphes à $\mathbf{R}P^3 \setminus \bar{B}^3$. En conséquence, si $\bar{\Omega}_j$ désigne la compactification de Ω_j qui consiste à « fermer » chaque pointe par un point, alors M est difféomorphe à la somme connexe des $\bar{\Omega}_j$ (ce qui correspond au recollement d'un tube entre deux Ω_j différents) avec un nombre fini de $S^2 \times S^1$ (qui correspondent au recollement de tubes entre deux bords différents du même Ω_j), un nombre fini de $\mathbf{R}P^3$ et un nombre fini de sphères (qui n'ont aucune influence sur la topologie). Au passage nous avons fait disparaître $M \setminus \Omega$; le théorème des voisinages canoniques montre que si la courbure scalaire tend vers l'infini en un point qui est dans une gorge, elle le fait sur toute la sphère de dimension deux correspondante; la structure des parties du complémentaire de Ω provenant des capuchons est moins claire.

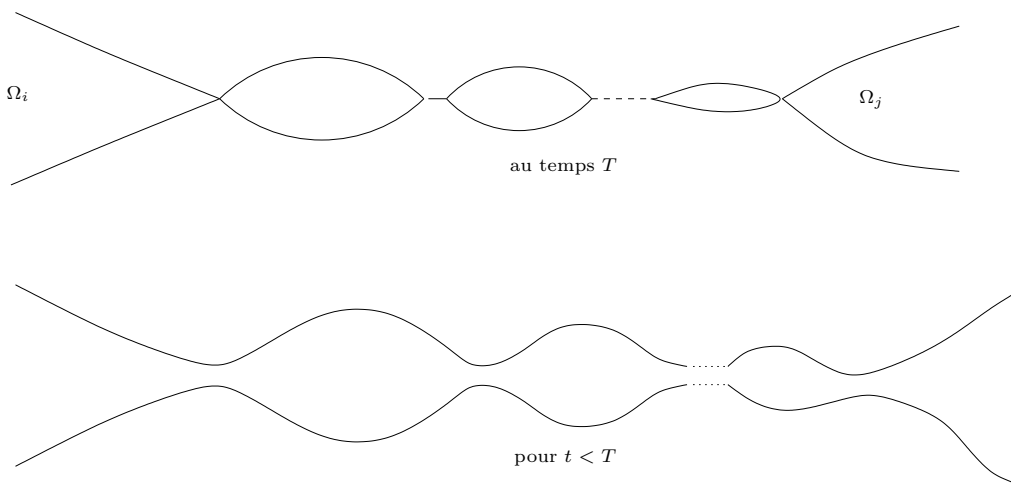
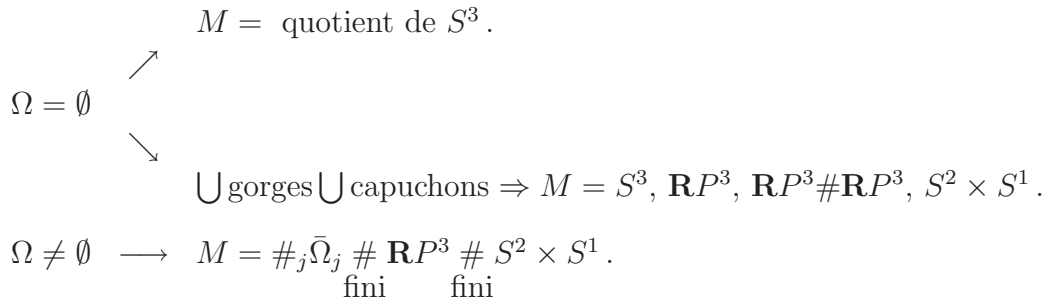


Fig. 2

Nous pouvons résumer cette section dans le diagramme suivant :



6.2. La solution standard

La condition initiale de la solution standard est une métrique complète sur \mathbf{R}^3 , d'opérateur de courbure positif ou nul et de courbure scalaire strictement positive qui est asymptote à un cylindre canonique et invariante par rotation. Il s'agit donc de recoller un hémisphère de courbure scalaire constante égale à 1 à un demi-cylindre $S^2 \times [0, +\infty)$ muni de la métrique produit (S^2 est de courbure scalaire constante égale à 1 également); le pôle de l'hémisphère est l'origine de \mathbf{R}^3 et est appelé le centre. Il faut utiliser une partition de l'unité afin de réaliser une métrique de classe C^∞ et la métrique de l'hémisphère est alors presque ronde. Il n'y a pas unicité et l'utilisation de l'article défini est abusive; toutefois, nous faisons un choix que nous appellerons « la » condition initiale de la solution standard. La variété n'étant pas compacte il n'est pas immédiat qu'il existe une solution ayant cette métrique pour donnée initiale en temps petit; la courbure étant bornée à l'infini ce résultat est dû à W.-X. Shi ([52]). De même, l'unicité est à prouver ([46], section 2 ou [8], appendice A). On démontre également que la solution est définie pour $t \in [0, 1)$, qu'elle est complète et invariante par rotation pour tout t (voir [46] et [8]). Enfin, la solution standard vérifie les conclusions du théorème 5.11 et on peut montrer que sa courbure scalaire vérifie, pour tout $x \in \mathbf{R}^3$ et $t \in [0, 1)$, $R(x, t) \geq \frac{\text{const.}}{1-t}$.

6.3. La première chirurgie

C'est l'idée inventée par R. Hamilton pour la dimension 4, dans [30], qui est reprise par G. Perelman et adaptée à la dimension 3. Elle consiste à opérer une chirurgie métrique pour éliminer les parties de la variété qui sont susceptibles de devenir des singularités. La chirurgie se pratique sur des composantes de Ω qui contiennent des points de $\Omega(\rho)$; ce faisant, nous sommes dans l'obligation d'éliminer des parties de Ω dont nous devons contrôler la topologie. La description donnée dans la section précédente joue alors un rôle essentiel. Insistons sur le fait que la chirurgie se pratique toujours le long de sphères et ce n'est donc pas elle qui produira la décomposition de Jaco-Shalen-Johannson mais plutôt celle de Kneser.

La chirurgie est faite à une échelle différente de ρ . En fait, il existe un nombre $0 < h < \delta\rho$ tel que, si (x, T) est un point qui est dans une pointe dont le bord rencontre $\Omega(\rho)$, vérifiant $R(x, T) \geq h^{-2}$, le voisinage parabolique $P(x, T, \delta^{-1}R(x, T)^{-\frac{1}{2}}, -R(x, T)^{-1})$

est contenu dans une δ -gorge (en un sens plus fort que nous omettons de préciser). Nous dirons que h est le paramètre de chirurgie; il dépend des données du problème que sont la fonction ϕ du théorème 3.4, celle du nombre r_0 apparaissant dans 5.13, de ϵ et δ . Tant que δ est minoré on peut choisir h minoré; le travail de la prochaine sous-section est justement l'existence d'un tel δ . Le nombre δ est choisi beaucoup plus petit que ϵ .

On procède alors de la manière suivante :

a) on élimine les composantes de Ω qui ne rencontrent pas $\Omega(\rho)$. On connaît leur impact sur la topologie de M ,

b) dans les pointes qui rencontrent $\Omega(\rho)$ on choisit un point x tel que $R(x, T) = h^{-2}$; il est le centre d'une δ -gorge (par définition de h). On coupe le long de la sphère de dimension 2 qui le contient, c'est-à-dire la sphère centrale,

c) sur le bord libre, difféomorphe à S^2 , on colle un homothétique de rapport h^2 d'un voisinage de taille fixe du centre de la condition initiale de la solution standard. Le recollement est décrit ci-dessous.

d) À partir de cette nouvelle variété on relance le flot en prenant T comme origine des temps.

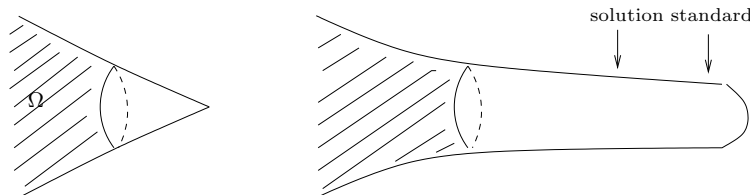


Fig. 3

Dans [30], R. Hamilton montre que la nouvelle variété vérifie le pincement de Hamilton-Ivey avec la même fonction ϕ , quitte à choisir correctement le recollement. Nous précisons ce dernier point.

La demi-gorge sur laquelle la chirurgie est opérée est identifiée à $S^2 \times [0, \frac{h}{\delta})$. Soit λ un nombre à choisir.

1) On munit $S^2 \times [0, \lambda]$ de la métrique g de la gorge,

2) On munit $S^2 \times [\lambda, 2\lambda]$ de la métrique $e^{-2f}g$, où f est une fonction de la coordonnée longitudinale seulement qui vaut 0 au voisinage de λ , qui est partout très proche de 0 et qui est choisie pour que la courbure sectionnelle soit positive à partir de 2λ . On peut également la choisir en sorte que les courbures sectionnelles de $e^{-2f}g$ soient supérieures à celles de g sur l'intervalle $[\lambda, 2\lambda]$ (voir [18] pour les détails). Par monotonie de la fonction ϕ , la condition de Hamilton-Ivey est alors vérifiée sur cet intervalle. Ensuite,

3) sur $S^2 \times [2\lambda, 3\lambda]$ la métrique est de la forme $e^{-2f}(\psi g + (1 - \psi)h^2\bar{g})$, où ψ est une fonction plateau de la coordonnée longitudinale qui vaut 1 au voisinage de 2λ et 0 au voisinage de 3λ , et \bar{g} est la métrique de la condition initiale de la solution standard. La fonction f étant fixée on peut choisir δ assez petit pour que la courbure sectionnelle de cette métrique soit positive. La condition de hamilton-Yvey est alors trivialement vérifiée. Enfin,

4) sur $S^2 \times [3\lambda, c]$, pour $3\lambda \leq c \leq 4\lambda$, la métrique est $e^{-2f} h^2 \bar{g}$ et le centre de la solution standard correspond à la valeur c du paramètre longitudinal. Sa courbure sectionnelle est strictement positive.

On peut décrire plus précisément \bar{g} comme un produit tordu sur le cylindre. Le nombre λ est de l'ordre de h/ϵ . On choisit donc les paramètres pour que la courbure soit supérieure à celle de g (c'est le cas pour $e^{-2f}g$) ou bien que la courbure sectionnelle soit positive; dans les deux cas le pincement de Hamilton-Ivey est vérifié. On note également que la distance du centre du capuchon ajouté à la sphère de chirurgie est de l'ordre de h/ϵ . En éliminant la demi-gorge on a perdu un volume de l'ordre de h^3/δ et, en collant le capuchon standard, on l'a augmenté de l'ordre de h^3/ϵ ; si δ est assez petit devant le paramètre ϵ fixe, alors la chirurgie fait perdre une quantité de volume de l'ordre de h^3 . Ce volume pourrait être récupéré très rapidement par l'évolution ultérieure du flot; il n'en est rien. En effet, cette chirurgie n'affecte pas le minimum de la courbure scalaire, puisqu'elle se pratique en des régions où celle-ci est grande, et ce minimum reste donc croissant. La formule déjà utilisée montre alors que la dérivée logarithmique du volume de la nouvelle variété est majorée; l'accroissement de volume dans un intervalle de temps fini donné est donc contrôlé. Cet argument montre que ce type de chirurgies ne peut intervenir qu'un nombre fini de fois dans un intervalle de temps fini, tant que δ est fixé; le nombre de chirurgies dépend de la condition initiale et des divers choix de paramètres.

Remarque 6.3. — On peut dire que r_0 est l'échelle à laquelle on contrôle la géométrie d'un voisinage (théorème des voisinages canoniques), h est l'échelle à laquelle on effectue la chirurgie. On pourrait effectuer la chirurgie à l'échelle ρ mais l'argument précédent montre qu'il faut se laisser un peu de marge. Le paramètre ϵ est un paramètre de contrôle qui décrit la proximité utilisée dans le théorème des voisinages canoniques. Il doit en particulier être assez petit pour que la réunion de deux gorges dont l'une est centrée sur le bord de l'autre soit difféomorphe à $S^2 \times I$; c'est une condition indépendante de la variété. Le paramètre δ est un paramètre de contrôle bien plus fin, nécessaire dans la chirurgie et, en particulier, dans l'argument précédent.

L'étape suivante consiste à montrer l'existence de ce flot modifié pour tout temps.

6.4. Existence pour tout temps du flot avec chirurgies

C'est la partie où l'on choisit les paramètres r_0 et δ afin de faire fonctionner le processus. Dans [46] il est appelé « Ricci flow with δ -cutoff »; nous garderons toutefois l'expression « flot avec chirurgies » sachant que cela implique l'utilisation du paramètre δ et du paramètre r_0 . Une variété riemannienne M , compacte orientable de dimension 3 est dite *normalisée* si $|\text{Rm}(x)| \leq 1$, pour tout $x \in M$, et le volume de toute boule de rayon 1 est au moins la moitié du volume de la boule unité de \mathbf{R}^3 . Dans [31], R. Hamilton montre que, pour des données initiales normalisées, la plus petite valeur

propre de l'opérateur de courbure vérifie

$$(1+t)R(x,t) \leq (1+t)(-\nu(x,t))(\ln((1+t)(-\nu(x,t)) - 3)),$$

dès que $-\nu > 0$. Ceci montre l'existence d'une fonction décroissante $\phi : (0, +\infty) \rightarrow \mathbf{R}$, telle que $\phi(x) \sim \frac{1}{\ln(x)}$, pour x grand, avec la propriété que

$$(1+t)\nu(\cdot, t) \geq -(1+t)R(\cdot, t)\phi((1+t)R(\cdot, t)).$$

Ce choix généralise celui fait pour le théorème 3.4. Il est surtout utile pour les grands temps (et donc, comme nous le verrons plus loin, pour la géométrisation); il produit les mêmes résultats que ceux prouvés dans le théorème 3.4. La chirurgie n'affecte pas cette propriété comme nous l'avons déjà signalé.

Après la première chirurgie la nouvelle donnée initiale n'est plus normalisée, la variété considérée n'est plus connexe et les paramètres utilisés pour la chirurgie suivante doivent être modifiés, par exemple en les faisant dépendre du temps. Ils pourraient se dégrader à tel point que le processus s'arrête. La condition de courbure ϕ -presque positive étant préservée, la seule raison qui empêche la poursuite du flot est que le théorème des voisinages canoniques ne soit plus vérifié pour la valeur de r_0 choisie, il faut donc la modifier (essentiellement diminuer r_0) mais alors le risque est que ce nombre tende vers 0 en un temps fini.

Nous appellerons flot de Ricci avec chirurgies sur un intervalle $[0, T)$ le résultat de l'itération de la procédure décrite ci-dessus. Après la première chirurgie, la variété a un nombre fini de composantes connexes car on ne garde de Ω que les composantes qui intersectent $\Omega(\rho)$; on étudie le flot de Ricci sur chacune d'elles simultanément. On atteint éventuellement un nouveau temps singulier et apparaît un nouvel ensemble Ω . On pratique une nouvelle chirurgie et on recommence, espérant poursuivre ce processus pour tout temps. Notons que maintenant Ω peut contenir des composantes compactes. On dira que la solution (ou une de ses composantes) est éteinte si $\Omega(\rho)$ (ou la composante correspondante) est vide. Par commodité on considérera, dans le cas d'une extinction totale, que le flot se poursuit pour tous les temps supérieurs en un flot sur une variété vide. La notion de flot avec chirurgies est formalisée de manière générale dans [36], 61.1, et peut donc être appliquée à d'autres types de chirurgies que celles décrites ci-dessus. Le résultat principal est la proposition suivante

PROPOSITION 6.4 ([46], 5.1). — *Il existe des suites décroissantes $0 < r_j < \epsilon^2$, $0 < \kappa_j$, et $\bar{\delta}_j < \epsilon^2$, pour $j = 1, 2, \dots$, telles que, pour toute condition initiale normalisée et toute fonction $\delta(t)$, satisfaisant $0 < \delta(t) < \bar{\delta}_j$ pour $t \in [2^{j-1}\epsilon, 2^j\epsilon]$, le flot de Ricci avec chirurgies correspondant est défini pour $t \in [0, +\infty)$. De plus, il est κ_j -non-effondré et vérifie les conclusions du théorème des voisinages canoniques à l'échelle r_j sur l'intervalle de temps $t \in [2^{j-1}\epsilon, 2^j\epsilon]$.*

REMARQUE SUR LA PREUVE

Comme nous l'avons indiqué plus haut, la seule chose à faire est de montrer que le théorème des voisinages canoniques est vrai avec un paramètre ne se dégradant pas

trop vite. En cela la preuve n'est qu'une nouvelle version du théorème 5.13 tenant compte des chirurgies. Elle est techniquement plus difficile, mais ne contient pas d'idée fondamentalement nouvelle. On constate au passage que la constante κ doit aussi être modifiée, ce qui est indispensable comme on peut s'en convaincre en étudiant l'exemple du produit d'une surface hyperbolique avec S^1 (exemple suggéré dans [39]). ■

En conclusion, si la variété obtenue après k chirurgies est appelée M_k , M s'obtient en faisant la somme connexe des composantes de M_k , d'un nombre fini de quotients de S^3 (par des sous-groupes finis du groupe d'isométries de la métrique ronde) et d'un nombre fini de copies de $S^2 \times S^1$.

7. TRAVAUX DE PEREL'MAN III

7.1. Extinction en temps fini

Il y a un cas où l'on peut connaître complètement la topologie de la variété M , c'est celui où le flot s'éteint totalement en un temps fini. En effet, sous cette hypothèse, il n'y a qu'un nombre fini de chirurgies et la discussion précédente montre que M est obtenue comme somme connexe d'un nombre fini de copies de quotients de S^3 (par des sous-groupes du groupe d'isométries canonique) et de $S^2 \times S^1$. Il est donc intéressant de chercher des hypothèses impliquant cette extension. C'est ce qui est fait dans [45]. Nous présentons ici la variante due à T. Colding et W. Minicozzi (voir [16]) qui nous semble plus simple.

Soit M une variété de dimension 3, compacte, connexe, orientée, de classe C^∞ et $g(t)$ un flot de Ricci **sans chirurgies** sur M . On suppose que M est première pour la décomposition de Kneser en somme connexe, c'est-à-dire que, si $M = P_1 \# P_2$, alors P_1 ou P_2 est homéomorphe à S^3 (voir [32], théorème 1.5); on suppose également que M a un groupe fondamental fini. On dit que M est irréductible si toute sphère $S^2 \subset M$ borde une boule. Si M est compacte, connexe, orientable, première et de groupe fondamental fini elle est irréductible ([32], proposition 1.4) et son revêtement universel est une sphère d'homotopie. En particulier, $\pi_3(M) \neq \{0\}$.

Considérons une situation générale où (N, g) est une variété riemannienne compacte, connexe et orientable; on peut définir l'espace $H = L_1^2(S^2, N)$ des applications de S^2 dans N dont la différentielle est de carré intégrable; il suffit pour cela de plonger N isométriquement dans \mathbf{R}^n (voir [41]). On peut également considérer des applications à différentielles Hölder afin d'avoir plus de régularité. Appelons i la fonction qui associe à un point $x \in N$ l'application $i(x) : S^2 \rightarrow N$ qui envoie tout S^2 sur x . Dans [41], les auteurs montrent que si $\pi_3(N) \neq \{0\}$, alors $\pi_1(H, i(N)) \neq \{0\}$, c'est-à-dire : l'espace des applications de S^2 dans N modulo les applications constantes n'est pas simplement connexe. Soit alors un chemin continu $\beta : [0, 1] \rightarrow C^0 \cap L_1^2(S^2, N)$, tel que $\beta(0)$ et $\beta(1)$ soient des applications constantes et que la classe d'homotopie $[\beta]$ de β soit non

triviale. On définit l'énergie

$$W(g) = \min_{\gamma \in [\beta]} \max_{s \in [0,1]} E(\gamma(s)),$$

où $E(\gamma(s))$ désigne l'énergie de l'application $\gamma(s)$ de S^2 dans N .

On montre que les propriétés topologiques de H impliquent que $W(g) > 0$, pour toute métrique g (voir [35]). T. Colding et W. Minicozzi prouvent le résultat suivant, pour un flot de Ricci $(M, g(t))$,

THÉORÈME 7.1 ([16]). — *Soit $(M, g(t))$ un flot de Ricci (sans chirurgies) où M est une variété de dimension 3, compacte, connexe, orientable, première et non asphérique (i.e. il existe $k > 1$ tel que $\pi_k(M) \neq \{0\}$). Alors,*

$$\frac{dW}{dt}(g(t)) \leq -4\pi + \frac{3}{4(t+C)}W(g(t)),$$

pour $C = -\frac{3}{2R_{\min}(0)}$ si $R_{\min}(0) < 0$, et $C = +\infty$ sinon (c'est-à-dire, le terme correspondant disparaît). En particulier, le flot s'éteint en temps fini.

Notons que, bien que la fonction W ne soit pas nécessairement dérivable, on peut donner un sens à l'expression ci-dessus. La méthode de preuve utilise des résultats standards de la théorie des applications harmoniques et des calculs d'aires assez faciles. G. Perelman considère, dans [45], des classes d'homotopie d'applications de S^1 dans M . Il remplit ces applications par des disques minimaux ; toutefois, comme le problème de Plateau est difficile à résoudre si la courbe bordante n'a pas une forme agréable, il la « régularise » en lui appliquant le flot de la courbure extrinsèque décrit par M. Grayson ([22]). On peut comparer cette idée à un argument similaire utilisé dans [31].

Le théorème 7.1 s'applique au cas où M est de groupe fondamental fini et irréductible.

Question : Peut-il s'étendre à un flot avec chirurgies ?

Admettons que ce soit le cas, cela montrerait qu'une variété irréductible de groupe fondamental fini est un quotient fini de S^3 par un groupe d'isométries canoniques. En effet, d'après la discussion précédente, la variété est le résultat de la somme connexe d'un nombre fini de quotients de S^3 (par des sous-groupes finis du groupe d'isométrie de la sphère canonique) et de $S^2 \times S^1$; l'indécomposabilité implique qu'il n'y a qu'un terme non égal à S^3 dans cette suite d'opérations et la finitude du groupe fondamental ne laisse que la possibilité d'un quotient de S^3 . Si la variété n'est pas première mais de groupe fondamental fini, ses composantes dans la décomposition de Kneser sont premières et de groupe fondamental fini, donc des quotients de S^3 ; une au maximum est non simplement connexe (sinon le groupe fondamental de M est infini). En appliquant la discussion précédente à chaque composante on montre que M est aussi un quotient de S^3 (voir aussi la discussion du cas non irréductible dans [45] et [16]). Ceci prouverait les conjectures 0.1 et 0.2.

Revenons à la variété compacte M irréductible et de groupe fondamental fini, munie d'un flot de Ricci avec chirurgies non trivial (il y a au moins une chirurgie). Après la première chirurgie, une, au plus, des composantes connexes n'est pas une sphère, nous

l'appellerons M_1 et la métrique riemannienne post-chirurgie sera désignée par $g_1(T)$; la composante connexe de Ω qui lui correspond est notée Ω_1 et T désigne le premier temps singulier. La proposition suivante permet d'adapter l'argument de T. Colding et W. Minicozzi,

PROPOSITION 7.2. — *Il existe une équivalence d'homotopie qui contracte les distances entre $(\bar{\Omega}_1, g(T))$ et $(M_1, g_1(T))$.*

ESQUISSE DE PREUVE

Pour simplifier, supposons qu'il n'y a qu'une pointe dans Ω_1 contenant des points de $\Omega(\rho)$. On tronque la pointe comme indiqué et on remplace le demi-cylindre noté \mathcal{H} par un voisinage du centre de la condition initiale de la solution standard pour obtenir M_1 . En utilisant les notations de la sous-section 6.3, l'application considérée est l'identité sur $(\Omega \setminus \mathcal{H}) \cup S^2 \times [0, c)$ et elle envoie $\mathcal{H} \setminus S^2 \times [0, c)$ sur le centre de la solution standard. Comme précédemment, la fonction f étant fixée, si δ est assez petit, l'application ci-dessus contracte les distances. On peut également déformer celle-ci en un homéomorphisme de $\bar{\Omega}_1$ sur M_1 .

Cet énoncé n'est pas parfaitement correct car la structure différentiable sur $\bar{\Omega}_1$ n'a pas été définie et donc $g(T)$ n'est pas définie au bout de la pointe. On peut contourner ce problème en considérant $(M, g(t))$ pour t proche de T . M est homéomorphe à $\bar{\Omega}_1$ et à M_1 et si on effectue l'opération ci-dessus avec $(M, g(t))$, on construit une application lipschitzienne de rapport $1 + \chi(t)$ avec $\chi(t) \rightarrow 0$ lorsque $t \rightarrow T$. Les arguments qui suivent s'adaptent sans difficultés. Pour simplifier l'expression nous conservons la formulation ci-dessus. ■

Maintenant, si $[\beta]$ est une classe non triviale de $\pi_1(H, i(M))$, elle persiste en une classe non triviale dans $\pi_1(H, i(M_1))$. De plus, grâce à la proposition ci-dessus, on voit que son énergie vérifie

$$W(M_1, g_1(T)) \leq W(\bar{\Omega}_1, g(T)) \quad \left(\text{ou bien } \leq \liminf_{t \rightarrow T} (1 + \chi(t))W(M, g(t)) \right).$$

Le théorème 7.1 s'applique alors pour le flot sur M_1 dont la donnée initiale est $g_1(T)$. La constante C peut éventuellement changer, mais le minimum de la courbure scalaire étant croissant le long d'un flot de Ricci sans chirurgies et non affecté par une chirurgie (à moins que toute la variété disparaisse), elle est croissante également le long d'un flot avec chirurgies; en conséquence, l'inégalité du théorème 7.1 est valable pour le flot avec chirurgies et montre l'extinction en temps fini (par itération de l'argument précédent).

Ceci termine la preuve des conjectures 0.1 et 0.2. dkdskdkdisd ; , ;wsdkldkljklmsdjf

Remarque 7.3. — Les arguments ci-dessus ont été vérifiés grâce à plusieurs échanges avec B. Kleiner et J. Lott. Je tiens à les remercier pour leur aide.

7.2. Conclusion : vers la géométrisation des variétés de dimension 3

S'il n'y a pas extinction en temps fini la situation est beaucoup plus complexe. Il peut, en effet, être nécessaire de pratiquer une infinité de chirurgies. Encore une fois le premier pas a été franchi par R. Hamilton, pour un flot de Ricci sans chirurgies, dans [31]. Nous reprenons l'énoncé tel qu'il est donné dans [43] ; le lecteur est renvoyé à [31] pour les détails et des énoncés précis.

THÉORÈME 7.4. — *Soit $(M, g(t))$ un flot de Ricci sur une variété M de dimension 3, compacte, connexe et orientée. On suppose qu'il existe pour tout $t \in [0, +\infty)$ et que la courbure normalisée $t \operatorname{Rm}(x, t)$ est bornée lorsque t tend vers l'infini. Alors, il existe un nombre fini de variétés hyperboliques complètes H_i de volume fini, et, pour tout t assez grand, un plongement $\phi_t : \bigsqcup H_i \longrightarrow M$ vérifiant les propriétés suivantes. L'image réciproque de la métrique (normalisée) $t^{-1}g(t)$ par ϕ_t converge, uniformément sur tout compact de $\bigsqcup H_i$, vers une métrique de courbure constante négative. Les tores des H_i , sections des cusps, sont envoyés, par ϕ_t , sur des tores incompressibles dans M . La métrique $t^{-1}g(t)$ sur le complémentaire de l'image de ϕ_t s'effondre à courbure sectionnelle bornée en valeur absolue.*

On rappelle qu'un tore est dit incompressible si son groupe fondamental s'injecte dans celui de M . Les parties de M qui s'effondrent sont classifiées par les résultats de J. Cheeger et M. Gromov ([6] et [7]). On ne sait pas montrer que le flot de Ricci uniformise les variétés qui s'effondrent et la conclusion doit venir d'un ingrédient extérieur à la théorie.

G. Perel'man annonce essentiellement le même résultat dans la situation plus compliquée du flot avec chirurgies et où la courbure normalisée n'est plus nécessairement bornée à l'infini. Une des difficultés supplémentaires est que le complémentaire des variétés hyperboliques s'effondre à courbure sectionnelle minorée et leur classification ne relève plus des travaux de J. Cheeger et M. Gromov. G. Perel'man affirme qu'en étudiant les espaces limites de ces effondrements, qui sont des espaces d'Alexandrov de courbure minorée (voir [4] pour une définition), il arrive à conclure. Une autre étude des effondrements à courbure sectionnelle minorée est faite dans [55] et [54]. Insistons sur le fait que les chirurgies ne réalisent pas la décomposition le long des tores incompressibles ; de plus, s'il y en a un nombre infini, elles ont lieu dans les régions qui s'effondrent (car leur complémentaire converge). Il manque à cette théorie, outre la vérification des énoncés, une meilleure compréhension de l'évolution des effondrements.

L'état actuel de l'expertise ne permet pas de se prononcer sur ces développements. Toutefois, la démarche est cohérente et compatible avec les résultats espérés. Pour terminer, il faut insister sur l'apport de G. Perel'man dans le premier article ([44]) : le non-effondrement local, l'étude des solitons contractants et des κ -solutions et la description des régions où la courbure est grande sans être maximale. Bon nombre de ces résultats sont valables en dimension quelconque et donc utilisables dans d'autres

contextes (voir [8], par exemple). Il est aussi important de saluer la contribution de R. Hamilton à cet édifice dont il a posé les fondations.

RÉFÉRENCES

- [1] M. ANDERSON. Geometrization of three-manifolds via the Ricci flow. *Notices A.M.S.*, 51(2) :184–193, 2004.
- [2] L. BESSIÈRES. Conjecture de Poincaré : la preuve de R. Hamilton et G. Perelman. *La gazette des mathématiciens*, 106, 2005.
- [3] J.-P. BOURGUIGNON. L'équation de la chaleur associée à la courbure de Ricci. In *Séminaire Bourbaki 1985-86*, Exp. n° 653, Astérisque 145-146, pages 45–61. Société Mathématique de France, 1987.
- [4] D. BURAGO, Y. BURAGO, and S. IVANOV. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. Amer. Math. Soc. Providence R.I., 2001.
- [5] J. CHEEGER and D. EBIN. *Comparison theorems in Riemannian geometry*. North-Holland Publishing Co., Amsterdam-Oxford, 1975.
- [6] J. CHEEGER and M. GROMOV. Collapsing Riemannian manifolds while keeping their curvature bounded I. *J. Differential Geometry*, 23 :309–346, 1986.
- [7] J. CHEEGER and M. GROMOV. Collapsing Riemannian manifolds while keeping their curvature bounded II. *J. Differential Geometry*, 32 :269–298, 1990.
- [8] B.-L. CHEN and X.-P. ZHU. Ricci flow with surgery on four-manifolds with positive isotropic curvature. ArXiv : math.DG/0504478v1, april 2005.
- [9] X. CHEN, P. LU, and G. TIAN. A note on uniformization of Riemann surface by Ricci flow. ArXiv : math.DG/0505163, may, 10 2005.
- [10] B. CHOW. On the entropy estimate for the Ricci flow on compact 2-orbifolds. *J. Differential Geometry*, 33 :597–600, 1991.
- [11] B. CHOW. The Ricci flow on the 2-sphere. *J. Differential Geometry*, 33 :325–334, 1991.
- [12] B. CHOW and S.-C. CHU. A geometric interpretation of Hamilton's Harnack inequality for the Ricci flow. ArXiv : math.DG/0211349, November, 22 2002.
- [13] B. CHOW and D. KNOPF. *The Ricci flow : an introduction*, volume 110 of *Mathematical surveys and monographs*. A.M.S., 2004.
- [14] B. CHOW and P. LU. The time-dependent maximum principle for systems of parabolic equations subject to an avoidance set. ArXiv : math.DG/0211209, november 2002. To appear in *Pacific J. Math*.
- [15] B. CHOW and L.-F. WU. The Ricci flow on compact 2-orbifolds with curvature negative somewhere. *Comm. on Pure and Appl. Math.*, 44 :275–286, 1991.

- [16] T. COLDING and W. MINICOZZI. Estimates for the extinction time for the Ricci flow on certain three-manifolds and a question of Perelman. *Journal of the A.M.S.*, 18(3) :561–569, 2005.
- [17] D. DETURCK. Deforming metrics in the direction of their Ricci tensors. *J. Differential Geometry*, 18 :157–162, 1983.
- [18] Y. DING. Notes on Perelman’s second paper.
<http://www.math.uci.edu/~yding/perelman.pdf>.
- [19] M. GAGE and R. HAMILTON. The heat equation shrinking convex plane curves. *J. Differential Geometry*, 23 :69–96, 1986.
- [20] S. GALLOT, D. HULIN, and J. LAFONTAINE. *Riemannian geometry*. Universitext. Springer-Verlag, 2004.
- [21] M. GRAYSON. The heat equation shrinks embedded plane curves to round points. *J. Differential Geometry*, 26 :285–314, 1987.
- [22] M. GRAYSON. Shortening embedded curves. *Annals of Mathematics*, 129 :71–111, 1989.
- [23] R. GREENE. A genealogy of noncompact manifolds of nonnegative curvature : history and logic. In *Comparison geometry*, volume 30 of *M.S.R.I. publications*, pages 99–134. M.S.R.I., 1997.
- [24] R. HAMILTON. Three-manifolds with positive Ricci curvature. *J. Differential Geometry*, 17 :255–306, 1982.
- [25] R. HAMILTON. Four-manifolds with positive curvature operator. *J. Differential Geometry*, 24 :153–179, 1986.
- [26] R. HAMILTON. The Ricci flow on surfaces. In *Mathematics and general relativity (Santa Cruz 1986)*, volume 71 of *Contemp. Math.*, pages 237–262. Amer. Math. Soc., Providence, RI, 1988.
- [27] R. HAMILTON. The Harnack estimate for the Ricci flow. *J. Differential Geometry*, 37(1) :225–243, 1993.
- [28] R. HAMILTON. A compactness property for solutions of the Ricci flow. *Amer. J. Math.*, 117(3) :545–572, 1995.
- [29] R. HAMILTON. The formation of singularities in the Ricci flow. In *Surveys in differential geometry*, volume II, pages 7–136. International Press, Cambridge MA, 1995.
- [30] R. HAMILTON. Four-manifolds with positive isotropic curvature. *Comm. Anal. Geom.*, 1 :1–92, 1997.
- [31] R. HAMILTON. Non-singular solutions of the Ricci flow on three-manifolds. *Comm. Anal. Geom.*, 7(4) :625–729, 1999.
- [32] A. HATCHER. Notes on basic 3-manifold topology.
<http://www.math.cornell.edu/~hatcher>, 1997.

- [33] G. HUISKEN. Ricci deformation of the metric on a Riemannian manifold. *J. Differential Geometry*, 21(1) :47–62, 1984.
- [34] T. IVEY. Ricci solitons on compact three-manifolds. *Diff. Geom. Appl.*, pages 301–307, 1993.
- [35] J. JOST. *Two-dimensional geometric variational problems*. J. Wiley-Intersciences, Chichester N.Y., 1991.
- [36] B. KLEINER and J. LOTT. Notes on Perelman’s papers.
http://www.math.lsa.umich.edu/research/Ricci_flow/perelman.html, december 30 2004.
- [37] O.A. LADYSENSKAJA, V.A. SOLONNIKOV, and N.N. URALCEVA. *Linear and quasilinear equations of parabolic type*, volume 23 of *Transl. Amer. Math. Soc.* Amer. Math. Soc., 1968.
- [38] P. LI and S.T. YAU. On the parabolic kernel of the Schrödinger operator. *Acta Math.*, 156(3-4) :153–201, 1986.
- [39] S. MAILLOT. Flot de Ricci et géométrisation des variétés de dimension 3. Notes informelles, juin 2004.
- [40] Ch. MARGERIN. Pointwise pinched manifolds are space forms. In *Geometric measure theory and the calculus of variations (Arcata, Calif., 1984)*, volume 44 of *Proc. Sympos. Pure Math.*, pages 307–328. Amer. Math. Soc., Providence, RI, 1986.
- [41] M. MICALLEF and J. MOORE. Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes. *Annals of Mathematics*, 127 :199–227, 1988.
- [42] J. MILNOR. Towards the Poincaré conjecture and the classification of 3-manifolds. *Notices A.M.S.*, 50 :1226–1233, 2003.
- [43] J. MORGAN. Recent progress on the Poincaré conjecture. *Bulletin of the Amer. Math. Soc.*, 42(1) :57–78, 2005.
- [44] G. PERELMAN. The entropy formula for the Ricci flow and its geometric applications. ArXiv : math.DG/0211159, november 2002.
- [45] G. PERELMAN. Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. ArXiv : math.DG/0307245, july 2003.
- [46] G. PERELMAN. Ricci flow with surgery on three-manifolds. ArXiv : math.DG/0303109, march 2003.
- [47] V. POENARU. Poincaré et l’hypersphère. *Pour la Science*, Dossier hors-série n° 41 :52–57, 2003.
- [48] H. POINCARÉ. Cinquième complément à l’analysis situs. *Rend. Circ. Mat. Palermo*, 18 :45–110, 1904.
- [49] M.H. PROTTER and H.F. WEINBERGER. *Maximum Principles in differential equations*, volume 110 of *Mathematical surveys and monographs*. A.M.S., 2004.

- [50] P. SCOTT. The geometries of 3-manifolds. *Bull. London Math. Soc.*, 15 :401–487, 1983.
- [51] N. SESUM, G. TIAN, and X. WANG. Notes on Perelman’s paper on the entropy formula for the Ricci flow and its applications. Notes, september 24 2003.
- [52] W.-X. SHI. Deforming the metric on complete Riemannian manifolds. *J. Differential Geometry*, 30(1) :223–301, 1989.
- [53] W.-X. SHI. Ricci deformation of the metric on complete noncompact Riemannian manifolds. *J. Differential Geometry*, 30(2) :303–394, 1989.
- [54] T. SHIOYA and T. YAMAGUCHI. Collapsing three-manifolds under a lower curvature bound. *J. Differential Geometry*, 56 :1–66, 2000.
- [55] T. SHIOYA and T. YAMAGUCHI. Volume collapsed three-manifolds with a lower curvature bound. ArXiv : math.DG/0304472, April, 15 2003.
- [56] W. P. THURSTON. *The geometry and topology of 3-manifolds*. Lecture Notes. Princeton University, 1979.
- [57] W. P. THURSTON. Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bulletin of the Amer. Math. Soc.*, 6(3) :357–381, 1982.
- [58] P. TOPPING. Lectures on the Ricci flow. Communication privée, 9 mai 2005.
- [59] L.-F. WU. The Ricci flow on 2-orbifolds with curvature. *J. Differential Geometry*, 33 :575–596, 1991.
- [60] L.-F. WU. The Ricci flow on 2-orbifolds with positive curvature. *J. Differential Geometry*, 33 :575–596, 1991.
- [61] R. YE. Notes on the reduced volume and asymptotic Ricci solitons of κ -solutions. [http ://www.math.ucsb.edu/~yer/ricciflow.html](http://www.math.ucsb.edu/~yer/ricciflow.html), december 20.

G erard BESSON

Institut Fourier de Math ematiques

UMR 5582 du CNRS

B.P. 74

F-38402 SAINT-MARTIN-D’H ERES C EDEX

E-mail : G.Besson@ujf-grenoble.fr

ON THE PROOF OF THE PARISI FORMULA
BY GUERRA AND TALAGRAND

by Erwin BOLTHAUSEN

1. THE SHERRINGTON-KIRKPATRICK MODEL

We consider “Ising spins” $\sigma_i \in \{-1, 1\}$, $i = 1, \dots, N$. Spin configurations will be denoted by $\sigma = (\sigma_i)_{i=1, \dots, N} \in \Sigma_N \stackrel{\text{def}}{=} \{-1, 1\}^N$. As the Sherrington-Kirkpatrick model (SK-model for short) is a mean-field model, there is no geometric structure of $\{1, \dots, N\}$ assumed.

Let further J_{ij} , $1 \leq i < j \leq N$, be i.i.d. standard Gaussian random variables, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. These random variables form the “random environment”. The **Hamiltonian** is the following random function $\Sigma_N \rightarrow \mathbb{R}$:

$$(1) \quad H_{N,\omega}(\sigma) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} J_{ij}(\omega) \sigma_i \sigma_j, \quad \omega \in \Omega,$$

We will often drop ω and N in such expressions. Remark that for any σ , this is a random variable, and indeed a centered Gaussian one. The covariances are given by:

$$(2) \quad \begin{aligned} \mathbb{E}(H_N(\sigma) H_N(\sigma')) &= \frac{1}{N} \sum_{1 \leq i < j \leq N} \sigma_i \sigma_j \sigma'_i \sigma'_j = \frac{1}{2N} \sum_{i,j=1}^N \sigma_i \sigma_j \sigma'_i \sigma'_j - \frac{1}{2} \\ &= \frac{N}{2} \left(\frac{1}{N} \sum_{i=1}^N \sigma_i \sigma'_i \right)^2 - \frac{1}{2}. \end{aligned}$$

The quantity in brackets is the so-called **overlap** of the two spin configurations

$$R_N(\sigma, \sigma') \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \sigma_i \sigma'_i.$$

The (random) Gibbs distribution $\mathcal{G}_{N,\beta,h,\omega}$ with inverse temperature $\beta > 0$, and external field with strength $h \in \mathbb{R}$ on Σ_N is defined by

$$\mathcal{G}_{N,\beta,h,\omega}(\sigma) \stackrel{\text{def}}{=} \frac{1}{Z_{N,\beta,h,\omega}} \exp \left[\beta H_{N,\omega}(\sigma) + h \sum_{i=1}^N \sigma_i \right],$$

where

$$Z_{N,\beta,h,\omega} \stackrel{\text{def}}{=} \sum_{\sigma} \exp \left[\beta H_{N,\omega}(\sigma) + h \sum_{i=1}^N \sigma_i \right]$$

is the appropriate norming constant⁽¹⁾. h here is non-random. There exist also variants where h is a random variable or where $h \sum_{i=1}^N \sigma_i$ is replaced by $\sum_{i=1}^N h_i \sigma_i$, where the h_i are random variables, e.g. $h_i = \gamma g_i + h$, $\gamma > 0$, $h \in \mathbb{R}$, and the g_i again being independent standard Gaussian random variables. It has some advantages to include such a Gaussian external field, as we will see later, but for the moment, we do not consider this possibility.

We write F_N for the finite volume free energy

$$F_N(\beta, h) \stackrel{\text{def}}{=} \frac{1}{N} \log Z_{N,\beta,h},$$

which is a random variable, defined on Ω , and

$$f_N(\beta, h) \stackrel{\text{def}}{=} \mathbb{E} F_N(\beta, h)$$

its expectation, the so-called “quenched” free energy. Sometimes, “quenched” refers to the random quantity only, but there is not much difference, as we will explain. In contrast, the so-called “annealed” free energy is obtained by taking the expectation inside the logarithm. By Jensen’s inequality, f_N is dominated by the annealed free energy.

Before proceeding with the discussion of the model, we try to explain why it is interesting.

The usual models of (non-random) Ising type are defined as follows. Consider a finite set Λ , and let $\Sigma_\Lambda \stackrel{\text{def}}{=} \{-1, 1\}^\Lambda$. Let further $A = (a_{ij})_{i,j \in \Lambda}$ be a real symmetric matrix, and $\mathbf{h} = (h_i)_{i \in \Lambda}$ be a real vector. Then the Gibbs measure $\mathcal{G}_{A,\mathbf{h}}$ on Σ_Λ is defined by

$$\mathcal{G}_{\Lambda,A,\mathbf{h}}(\sigma) \stackrel{\text{def}}{=} \frac{1}{Z_{\Lambda,A,\mathbf{h}}} \exp \left[\frac{1}{2} \sum_{i,j \in \Lambda} a_{ij} \sigma_i \sigma_j + \sum_{i \in \Lambda} h_i \sigma_i \right],$$

where of course

$$Z_{\Lambda,A,\mathbf{h}} \stackrel{\text{def}}{=} \sum_{\sigma} \exp \left[\frac{1}{2} \sum_{i,j \in \Lambda} a_{ij} \sigma_i \sigma_j + \sum_{i \in \Lambda} h_i \sigma_i \right].$$

Of great importance is the (finite volume) free energy, defined by

$$F_\Lambda(A, \mathbf{h}) = \frac{1}{|\Lambda|} \log Z_{A,\mathbf{h}}.$$

The importance of this quantity is coming from the fact that most of the physical interesting quantities can be expressed through it, like mean magnetization, entropy, etc.

The best known example is the Ising model where Λ is a finite (large) box in \mathbb{Z}^d , and

$$a_{ij} \stackrel{\text{def}}{=} \begin{cases} \beta & |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

⁽¹⁾In contrast to the habit in physics, we do not take a minus in front of β , and we also do not apply β to h .

Short range models are usually rather difficult to analyze, and often a qualitatively good approximation is obtained from *mean field models* where every spin interacts with any other one on equal footing. The simplest mean-field model is the *Curie-Weiss-model*. Here

$$a_{ij} \stackrel{\text{def}}{=} \beta / |\Lambda|, \quad \forall i, j \in \Lambda.$$

In that case one has with $N \stackrel{\text{def}}{=} |\Lambda|$

$$\frac{1}{2} \sum_{i,j \in \Lambda} a_{ij} \sigma_i \sigma_j = \frac{\beta}{2N} \left\{ \sum_{i \in \Lambda} \sigma_i \right\}^2,$$

and anything one wants to know can be derived from the Stirling approximation, and it becomes an easy exercise in elementary probability.

Spin glasses are models where the interactions are “disordered”, which typically means that they are obtained as a random object. A topic which is still very poorly understood is the case of *short range* random interactions, for instance when $\Lambda = \{-n, \dots, n\}^d$, and the a_{ij} are independent Gaussians for $|i - j| = 1$, and 0 otherwise. This is the *Edwards-Anderson model* on which there are ongoing controversial discussions in the physics community, the more so as it is very difficult to simulate on computers with a reasonably large box and in interesting dimensions. The SK-model is a mean-field model of this random interaction type, and it was invented in [18] certainly with the aim to have a simple model with disordered interaction. The $1/\sqrt{N}$ factor is easy to understand. In the Curie-Weiss model, each spin variable interacts with the other ones with a total interaction strength of order 1. Due to the cancellations between positive and negative J 's, the situation is essentially the same for the SK-model.

The model is evidently closely connected with questions probabilists have been interested in for a long time, namely maxima (or minima) of (Gaussian) random vectors. For instance, $\lim_{\beta \rightarrow \infty} (1/\beta) \log Z_{N,\beta,0}$ is simply $\max_{\sigma} H_N(\sigma)$, which is just the maximum of a family of correlated Gaussians with a simple covariance structure. Probabilists have developed methods to investigate such questions for a long time, e.g. Dudley, Fernique, Talagrand, and many others. It is not difficult to see that $\max_{\sigma} H_N(\sigma)$ is of order N and to prove that there are constants $0 < C_1 < C_2$ satisfying

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(C_1 N \leq \max_{\sigma} H_N(\sigma) \leq C_2 N \right) = 1.$$

However, the standard probabilistic techniques cannot derive the exact constant, which the Parisi-theory does, revealing a marvelous mathematical structure behind the problem, which is still *very* poorly understood, to this day.

The Parisi-theory applies to many other problems besides to the SK-model, e.g. to the assignment problem from combinatorial optimization, to the perceptron and the Hopfield net from neural networks, to coding theory, and to others. For some of these applications, see Nishimori [14].

Back to the SK-model, the first question one typically answers is the existence of the free energy in the thermodynamical limit (here just $N \rightarrow \infty$). It is however not at all

clear that the free energy

$$\lim_{N \rightarrow \infty} F_N(\beta, h)$$

exists. In principle, even if the limit exists, it could be a random variable. This possibility is however ruled out by Gaussian concentration inequalities. One says that the free energy is “self-averaging”, meaning that no randomness remains in the $N \rightarrow \infty$ limit. For a proof of the following inequality, see for instance [12].

PROPOSITION 1.1. — *Let γ_n be the standard Gaussian distribution on \mathbb{R}^n . Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant L . Then for any $u > 0$*

$$\gamma_n \left(f > \int f d\gamma_n + u \right) \leq \exp [-u^2/2L^2].$$

If we apply this inequality to $F_N(\beta, h)$, regarded as a function of the standard Gaussian vector $(J_{ij})_{1 \leq i < j \leq N}$, then one gets

$$\mathbb{P} \left(\left| \frac{1}{N} \log Z_{N,\beta,h} - \frac{1}{N} \mathbb{E} \log Z_{N,\beta,h} \right| \geq N^{-1/4} \right) \leq 2 \exp \left[-\frac{N^{1/2}}{\beta^2} \right].$$

It is therefore clear that instead of investigating $\lim_{N \rightarrow \infty} F_N(\beta, h)$, one can as well investigate the non-random object $\lim_{N \rightarrow \infty} f_N(\beta, h)$. The existence of this limit had been open for a long time, until Guerra and Toninelli [10] found a very nice, and not so obvious superadditivity property:

$$(3) \quad \mathbb{E} \log Z_{N_1+N_2} \geq \mathbb{E} \log Z_{N_1} + \mathbb{E} \log Z_{N_2},$$

from which one easily derives that

$$f(\beta, h) = \lim_{N \rightarrow \infty} f_N(\beta, h)$$

exists.

For the SK-model, the inequality came somewhat as a surprise. The proof is by a simple but very clever interpolation scheme which interpolates between the $(N_1 + N_2)$ -system, and the two independent smaller systems. Such interpolation schemes are at the very base of the recent progress in the understanding of the SK model, as we will see later. I am not going to prove the inequality here, but I will explain the interpolation method in another case below (Section 3).

There are many quantities in the SK-model which are *not* self-averaging in the $N \rightarrow \infty$ limit, i.e. which stay random (or at least are believed to be so). An example is the overlap of two independent “replicas”: Take σ, σ' to be two independent realizations under $\mathcal{G}_{N,\beta,h,\omega}$ for a fixed ω , and calculate $R_N(\sigma, \sigma')$, and then take the Gibbs-expectation. This is still a random variable (being a function of the interaction strengths). For small β , these random variables have a non-random limit for $N \rightarrow \infty$, but the limit stays random for large β .

The case $h = 0$ has some evident symmetry properties which make life easier, particularly in the high-temperature region. This case is however somewhat misleading. In

particular, the high temperature behavior for $h \neq 0$ is far from trivial, and it is actually of crucial importance for the understanding of the low temperature region.

For $h = 0$ and small enough β , the (“quenched”) free energy equals the “annealed” free energy.

THEOREM 1.2 (Aizenman-Lebowitz-Ruelle). — *For $h = 0$, and $\beta \leq 1$, one has*

$$(4) \quad f(\beta) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E} Z_{N,\beta} = \frac{\beta^2}{4} + \log 2.$$

The second equation is evident:

$$\begin{aligned} \mathbb{E} Z_{N,\beta} &= \sum_{\sigma} \mathbb{E} \exp [\beta H_N(\sigma)] = \sum_{\sigma} \exp \left[\frac{\beta^2}{2} \text{var} (H_N(\sigma)) \right] \\ &= 2^N \exp \left[\frac{\beta^2}{2} \text{var} (H_N(\sigma)) \right] = 2^N \exp \left[\frac{\beta^2}{2} \left(\frac{N}{2} - \frac{1}{2} \right) \right] \end{aligned}$$

from which the claim follows. The somewhat astonishing fact is that one can interchange the expectation with the logarithm. Of course, by Jensen, one always has

$$(5) \quad \mathbb{E} \log Z_{N,\beta} \leq \log \mathbb{E} Z_{N,\beta},$$

and therefore $f(\beta) \leq \beta^2/4 + \log 2$. We will indeed show later that $f(\beta) < \beta^2/4 + \log 2$ for $\beta > 1$. The proof of the above result is surprisingly simple and can be done by a second moment computation, proving that $\mathbb{E} Z^2 \leq \text{const} \times (\mathbb{E} Z)^2$ for $\beta < 1$, which is easy. Together with Gaussian isoperimetry (Proposition 1.1), this proves (4). The original proof in [1] was more complicated, but it derived also a much more detailed picture of the remaining fluctuations of $\log Z_N$.

There are other models like directed polymers for which one can prove that the quenched free energy equals the annealed one in certain regions, but typically, this is not possible by a simple second moment method in the full region where it is true. The fact that a second moment computation gives the result in the SK-model up to the correct critical value (for $h = 0$) is rather surprising.

For $h \neq 0$, “quenched=annealed” is never true, which reveals that this is a much more interesting situation, even where β is small.

2. THE REPLICA COMPUTATION AND THE PARISI FORMULA

The first evaluation of the free energy $f(\beta, h)$ was by Sherrington and Kirkpatrick [18], who applied the so-called “replica trick”. This is based on the observation that for a positive number x , one has $\log x = \lim_{n \downarrow 0} (x^n - 1)/n$. If X is positive random variable, one therefore has, provided the interchange of limits with the expectation is justified,

$$\mathbb{E} \log X = \lim_{n \downarrow 0} \frac{\mathbb{E} X^n - 1}{n}.$$

As integer moments are often easier to evaluate than non-integer ones, the “trick” is to evaluate $\mathbb{E}X^n$ for integer n , then somehow extend things analytically, and perform the above limit. This is the folk tale how the replica trick works, but for the SK-model, this is not quite the way it is done. In fact, one just *starts* the computation of $\mathbb{E}Z_N^n$ assuming that n is an integer, but as soon as convenient, one gives up this illusion and lets $n \rightarrow 0$, *before* really finishing the computation. From computations of the integer moments (in the $N \rightarrow \infty$ limit), one cannot derive the value of $f(\beta, h)$. I do not repeat the computation here, as it is done in many textbooks (see e.g. [14]), and the most interesting issue starts *after* the (non-rigorous) replica computation. The variational formula one obtains from the replica trick is

$$(6) \quad f(\beta, h) = \inf_{\mathbf{q}} \lim_{n \rightarrow 0} \left\{ -\frac{\beta^2}{2n} \sum_{\alpha < \beta \leq n} q_{\alpha\beta}^2 + \frac{1}{n} \log \text{tr}_\sigma e^{L(\mathbf{q}, \sigma)} + \frac{\beta^2}{4} \right\} + \log 2,$$

where

$$L(\mathbf{q}, \sigma) \stackrel{\text{def}}{=} \beta^2 \sum_{\alpha < \beta} q_{\alpha\beta} \sigma^\alpha \sigma^\beta + h \sum_{\alpha} \sigma^\alpha,$$

and where tr_σ means taking the average over an n -fold “replicated” spin $(\sigma^1, \dots, \sigma^n)$, and the infimum taken over $n \times n$ -matrices. The above expression (6) lacks a proper mathematical meaning as it is not clear what it should mean to take the $n \rightarrow 0$ limit, and what an $n \times n$ -matrix is for $n \sim 0$: What one does is to make an appropriate ansatz for the minimizing q -matrix, and then take a (formal) $n \rightarrow 0$ limit. Sherrington-Kirkpatrick made short work of the problem and assumed that there is no sufficient reason why the n replicas should behave “asymmetric”, and put $q_{\alpha\beta} = q$, and then take the infimum only over $q \geq 0$. This leads to the so-called “replica symmetric solution”. The computation is easy: We have that $\sum_{\alpha < \beta} q_{\alpha\beta}^2 = n(n-1)q^2/2$, and an n cancels out. Furthermore, the $n \rightarrow 0$ limit in this part is no longer particularly demanding: $\lim_{n \rightarrow 0} (n-1) = -1$. Taking the $n \rightarrow 0$ limit in the other part is only a bit more tricky. A simple computation yields

$$\frac{1}{n} \log \text{tr}_\sigma e^{L(\mathbf{q}, \sigma)} = -q\beta^2/2 + \frac{1}{n} \log E(1 + n \log \cosh(g\beta\sqrt{q} + h)) + o(1),$$

where g is a standard Gaussian variable, and E the expectation for this Gaussian. Expanding in n and letting $n \rightarrow 0$, one obtains for this $-q\beta^2/2 + E \log \cosh(g\beta\sqrt{q} + h)$. Summing things together, we get for this “replica symmetric solution”

$$(7) \quad RS(\beta, h) \stackrel{\text{def}}{=} \inf_q \left\{ \frac{\beta^2}{4} (1-q)^2 + \int \log \cosh(x\beta\sqrt{q} + h) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right\} + \log 2.$$

For later use, we find the minimizing q . A simple computation leads to the following fixed point equation for the optimal q :

$$(8) \quad q = \int \tanh^2(h + \beta\sqrt{q}x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

For $h = 0$, $q = 0$ is always a solution, and for $\beta \leq 1$, it is the only one, as one can easily check. Therefore $RS(\beta, 0) = \beta^2/4$ for $\beta \leq 1$. For $\beta > 1$, there is however at least one other solution of this fixed point equation, which follows easily by calculating the

derivative of the expression on the right hand side at $q = 0$. In fact, there is just *one* other solution $q(\beta) > 0$ which gives the minimum, and therefore $RS(\beta, 0) < \beta^2/4 + \log 2$ for $\beta > 1$. As we will prove $f(\beta, h) \leq RS(\beta, h)$ for all β, h , this proves that value in Theorem 1.2 is never correct for $\beta > 1$.

For $h > 0$, the equation (8) does have a unique positive solution:

LEMMA 2.1. — *Let $\beta, h > 0$ be arbitrary. Then (8) has a unique solution $q(\beta, h)$.*

The proof is due to Guerra and is short but a bit tricky. Talagrand has it in his book ([19]).

The main question is whether $f(\beta, h) = RS(\beta, h)$. It is certainly correct for $h = 0$ and $\beta \leq 1$, as we have seen before. However, for $\beta > 1$, it is not correct. This is far from trivial to see. It will however turn out that for $h \neq 0$, the formula is correct again for small β , but not for large ones. Even the small β case is highly non-trivial. That the solution cannot be correct for large β was already realized by Sherrington and Kirkpatrick by calculating the entropy, which has to be positive, but it can also be computed from the free energy, and if one uses RS , it becomes negative for large β . So already Sherrington and Kirkpatrick concluded that their own solution is not correct for large β .

The RS -solution is supposed to be correct for β up to the celebrated AT-line (de Almayda-Thouless line [3]), i.e. for β satisfying

$$(9) \quad \beta^2 \int \frac{1}{\cosh^4(h + \beta\sqrt{q(\beta, h)}x)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx < 1,$$

but this is not yet proved; it is now simply a nasty analytical problem, as the Parisi-formula for $f(\beta, h)$ is proved for the whole temperature region. (The above condition comes up through a local stability computation.)

In order to overcome the problem with the replica symmetric solution for large β , there had been various proposals for a different ansatz for the minimizing problem in (6), no longer assuming that all the $q_{\alpha\beta}$ are equal. This is the famous “replica symmetry breaking”. A particular ansatz for this is due to Parisi. The ansatz makes a *very special* assumption on the matrix $Q = (q_{\alpha\beta})$, namely that it has a kind of hierarchical organization. The question then remained if there could not be a better choice not satisfying the Parisi-ansatz. A justification of the Parisi-ansatz before Talagrand’s proof was the proof that it is in a sense locally stable, by computing Hessians, and that it was the only one found having this property, but the really convincing argument was that the outcome had interesting consequences also outside the “replica formulation”. Very nice explanations of these issues can be found in [14]. Here just a cursory explanation of what is going on.

The replica symmetric ansatz fixes the matrix Q to be of the following form

$$Q = \begin{pmatrix} 0 & q & q & \cdots & \cdots & q \\ & 0 & q & \cdots & \cdots & q \\ & & 0 & q & \cdots & q \\ & & & \ddots & & \vdots \\ & & & & 0 & q \\ & & & & & 0 \end{pmatrix}.$$

In the Parisi ansatz, one uses more complicated matrices. There are a number of levels. In the end, this number has to go to infinity, but let us first look at the simplest case, the case with one level of replica symmetry breaking. Here one takes a matrix of the form:

$$\begin{pmatrix} \begin{pmatrix} 0 & q_2 & q_2 \\ & 0 & q_2 \\ & & 0 \end{pmatrix} & \begin{pmatrix} q_1 & q_1 & q_1 \\ q_1 & q_1 & q_1 \\ q_1 & q_1 & q_1 \end{pmatrix} \\ \begin{pmatrix} 0 & q_2 & q_2 \\ & 0 & q_2 \\ & & 0 \end{pmatrix} & \begin{pmatrix} q_1 & q_1 & q_1 \\ q_1 & q_1 & q_1 \\ q_1 & q_1 & q_1 \end{pmatrix} \end{pmatrix}$$

The rule is that one divides the $n \times n$ -matrix by choosing $n_1 \leq n$ such that n/n_1 is an integer, and then one divides the matrix into $(n/n_1)^2$ submatrices of the form $n_1 \times n_1$. The diagonal blocks get q_2 above the diagonal, and the off-diagonal blocks all get q_1 . In the above example, one has $n = 6$ and $n_1 = 3$. Then one does the computation analogously as above, keeps $m_1 \stackrel{\text{def}}{=} n_1/n$ fixed, and lets formally $n \rightarrow 0$. This leads to a variational problem. One can check that one can always assume that $0 \leq q_1 \leq q_2 \leq 1$. For β small it turns out, that nothing new is achieved: The optimal choice for the q 's is $q_1 = q_2$, but for large β , some $q_1 < q_2$ give a lower value. This is the “one level symmetry breaking”, but one can proceed by dividing the q_2 blocks in a similar fashion, which leads to a “two level symmetry breaking”, and one can go on in this way with an arbitrary number of symmetry breakings. The calculations are somewhat lengthy but not difficult. Here is the outcome:

Let $K \in \mathbb{N}$ (the number of symmetry breakings), and then we choose parameters

$$(10) \quad 0 = m_0 < m_1 < \dots < m_{K-1} < m_K = 1,$$

$$(11) \quad 0 = q_0 \leq q_1 < \dots < q_K < q_{K+1} = 1.$$

For $i = 0, \dots, K$ let g_i be Gaussian with variance $\beta^2 (q_{i+1} - q_i)$, and set $Y_{K+1} \stackrel{\text{def}}{=} \cosh \left(h + \sum_{i=0}^K g_i \right)$. Then one defines

$$(12) \quad Y_K \stackrel{\text{def}}{=} [E_K (Y_{K+1}^{m_K})]^{1/m_K} = E_K (Y_{K+1}),$$

where E_K means that one integrates out g_K , so that Y_K still depends on g_0, \dots, g_{K-1} . Then one defines

$$Y_{K-1} \stackrel{\text{def}}{=} [E_{K-1} (Y_K^{m_{K-1}})]^{1/m_{K-1}}$$

and so on, until one gets Y_1 . Y_1 is still a random variable as it depends on g_0 . Remark however, that in case $q_1 = 0$ which we do not exclude, there is no randomness left. In any case, we set

$$(13) \quad \mathcal{P}_K(m, q; \beta, h) \stackrel{\text{def}}{=} E \log Y_1 - \frac{\beta^2}{4} \sum_{i=1}^K m_i (q_{i+1}^2 - q_i^2) + \log 2.$$

Then $\inf_{m,q} \mathcal{P}_K(m, q)$ is the value one obtains by optimizing (6) with the Parisi-ansatz at K levels of replica symmetry breaking, and therefore, believing that first of all the replica trick works, and secondly that the ansatz of Parisi finds the minimum, we get

$$(14) \quad f(\beta, h) = \inf_{K,m,q} \mathcal{P}_K(m, q) = \lim_{K \rightarrow \infty} \inf_{m,q} \mathcal{P}_K(m, q).$$

THEOREM 2.2 (Parisi Formula). — *The Parisi-formula (14) is correct for all β, h .*

The proof is due to Guerra [9] who proved the upper bound, and Talagrand [20] who then finished the proof.

In the case of the SK-model, either one has $K = 1$, which gives the true value in the region where the replica-symmetric solution is correct, or one has to take $K \rightarrow \infty$, and therefore one has “replica symmetry breaking” at infinitely many levels. There are other models, with the minimum assumed at one level of symmetry breaking, i.e. $K = 2$. One can artificially cook up cases with arbitrary K , but $K = 1, 2, \infty$ seem to be the only ones coming up “naturally”. In the case $K = \infty$, one can phrase the limit $K \rightarrow \infty$ directly as a variational problem involving continuous functions $q \rightarrow x(q)$. The finite K case then corresponds to taking step functions $x(q) = m_i$ for $q \in [q_i, q_{i+1})$.

Here is an outline of what the physicists believe to be the picture behind the RS-solution ($K = 1$), and the replica symmetry breaking ($K > 1$). This picture emerged partly from another non-rigorous approach, the so-called “cavity method” which led to the same formula for the free energy, and gave a clearer picture about the Gibbs distribution (see [13]).

The region where the RS-solution (7) is valid is characterized by the property that the σ_i under the Gibbs measure are still “fairly independent”. The $h = 0$ case is simple because, due to symmetry, the expectation under the Gibbs measure is 0. For $h \neq 0$, the expectation of σ_i under the Gibbs measure $\mathcal{G}_{N,\beta,h,\omega}$ is $m_i \stackrel{\text{def}}{=} \mathcal{G}(\sigma_i)$ which satisfies $\mathbb{E}m_i^2 = q(\beta, h)$, q being the solution of (8), equality in the $N \rightarrow \infty$ limit. The m_i are themselves approximately independent under the measure \mathbb{P} . One therefore has the following picture (for large N): The randomness of the disorder (i.e. the J_{ij}) produces the nearly i.i.d. random variables m_i , and given the disorder, the Gibbs measure has approximately independent spin variables σ_i with mean m_i . The property that the σ_i are approximately independent is reflected in the physics community saying that there is just “one pure state”.

Given this picture, $q(\beta, h)$ has a precise mathematical interpretation in terms of the Gibbs measure. It is the almost sure limit (as $N \rightarrow \infty$) of the overlaps of two

independent realizations of the spin variables:

$$(15) \quad R_N(\sigma, \sigma') = \frac{1}{N} \sum_{i=1}^N \sigma_i \sigma'_i \simeq \frac{1}{N} \sum_{i=1}^N m_i^2 \simeq q(\beta, h)$$

by the law of large numbers. The precise statement is as follows: Let $\nu_N^{(2)}$ be the measure on $\Sigma_N \times \Sigma_N$ defined by

$$(16) \quad \nu_N^{(2)}(\sigma, \sigma') \stackrel{\text{def}}{=} \int \mathbb{P}(d\omega) \mathcal{G}_{N,\omega}^{\otimes 2}(\sigma, \sigma'),$$

where $\mathcal{G}^{\otimes 2}$ denotes the twofold product Gibbs measure. Then for small enough β

$$\lim_{N \rightarrow \infty} \nu_N^{(2)}(|R_N(\sigma, \sigma') - q(\beta, h)| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

This means that the overlap of independent replicas is self-averaging. The high-temperature regime is now mathematically very well understood, mainly through the work of Michel Talagrand (see [19], Chapter 2).

In the low temperature regime things become much more complicated. First of all, the RS-solution is no longer correct, but this is only one aspect. The overlaps are no longer self-averaging but stay random. The Gibbs distribution splits into a “countable number of pure states”, a statement made in the physics literature which is difficult to make mathematically precise. Essentially the “pure states” under the Gibbs-distribution should be organized in a hierarchical way. This hierarchy somehow reflects the hierarchical ansatz in the Parisi-matrices above. Nothing of this has been proved mathematically, and probably not all statements made in the physics literature should be taken (mathematically) too literally. One important aspect is “ultrametricity”, which has the following precise mathematical meaning: Take three independent realizations $\sigma, \sigma', \sigma''$ under the Gibbs measure. Define $\nu_N^{(3)}$ for the three replicas, similarly as $\nu_N^{(2)}$ defined above. The claim is that for any $\varepsilon > 0$

$$(17) \quad \lim_{N \rightarrow \infty} \nu_N^{(3)}[R_N(\sigma, \sigma'') \geq \min(R_N(\sigma, \sigma'), R_N(\sigma', \sigma'')) - \varepsilon] = 0.$$

There is no proof of this. Together with the so-called Ghirlanda-Guerra identities, which I am not discussing here, this would essentially characterize the model completely.

Despite the recent progress on the SK-model in the full temperature regime, which is explained below, much of the above picture is mathematically not understood. In a case with one level of symmetry breaking only, the so-called p -spin SK model, there are results by Michel Talagrand, which confirm the physicists predictions, if properly formulated (see [19] Chapter 6).

3. GUERRA'S INTERPOLATION SCHEME: THE REPLICA SYMMETRIC BOUND

Much of the recent progress on the SK-model is based on a very clever argument invented by Guerra, which leads to bounds on the free energy. These bounds are obtained by interpolating continuously between the system one is interested in, and a much easier one.

I will explain this in the simplest case, where one proves that the replica symmetric solution is a strict bound for the free energy, for all N and in the full region of parameters.

THEOREM 3.1. — *For all $\beta > 0$, $h \in \mathbb{R}$, and $N \in \mathbb{N}$ one has*

$$f_N(\beta, h) \leq RS(\beta, h),$$

where $RS(\beta, h)$ is defined by (7).

Proof. — The proof is by interpolation. Let for an arbitrary number $q \geq 0$, and $t \in [0, 1]$

$$(18) \quad H(t, \sigma) \stackrel{\text{def}}{=} \sqrt{\frac{t}{N}} \sum_{1 \leq i < j \leq N} J_{ij} \sigma_i \sigma_j + \sqrt{1-t} \sum_{i=1}^N \sqrt{q} g_i \sigma_i,$$

$$\Phi(t, \sigma; \beta, h) \stackrel{\text{def}}{=} \beta H(t, \sigma) + h \sum_{i=1}^N \sigma_i,$$

where g_i is a set of standard Gaussian variables, independent of the J 's. The basic idea of this interpolation is to relate the Hamiltonian we are interested in with a much simpler one with independent σ_i , which however have the right overlap structure: For $t = 0$, we have, conditioned on the g_i , independent spins with mean $\tanh(h + \beta\sqrt{q}g_i)$. Therefore, if we take two independent realizations σ, σ' (still conditioned on the g_i), we get

$$\frac{1}{N} \sum_{i=1}^N \sigma_i \sigma'_i \simeq \frac{1}{N} \sum_{i=1}^N \tanh^2(h + \beta\sqrt{q}g_i)$$

$$\simeq \int \tanh^2(h + \beta\sqrt{q}x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

which equals q , if we take for q the solution of (8). The clever idea by Guerra is that one can control what happens along the path from $t = 0$ to $t = 1$. For the moment, we have not even to assume that q is the right one, and we can just take it arbitrary ≥ 0 . We again define the partition function

$$\xi(t) \stackrel{\text{def}}{=} \sum_{\sigma} e^{\Phi(t, \sigma; \beta, h)},$$

and we write $\mathcal{G}_t(\sigma)$ for the corresponding Gibbs measure. Let

$$(19) \quad \phi(t) \stackrel{\text{def}}{=} \frac{1}{N} \mathbb{E} \log \xi(t).$$

Remark that

$$(20) \quad \begin{aligned} \phi(0) &= \int \log \cosh(\beta\sqrt{q}x + h) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + \log 2 \\ \phi(1) &= f_N(\beta, h). \end{aligned}$$

We compute the derivative of $\phi(t)$ with respect to t .

$$\begin{aligned} \phi'(t) &= \frac{\beta}{2\sqrt{t}N^{3/2}} \sum_{\sigma} \sigma_i \sigma_j \sum_{i < j} \mathbb{E} J_{ij} \frac{\exp[\Phi(t, \sigma)]}{\xi(t)} - \frac{\sqrt{q}}{2\sqrt{1-t}} \sum_{\sigma} \sigma_i \sum_i \mathbb{E} g_i \frac{\exp[\Phi(t, \sigma)]}{\xi(t)} \\ &= A(t) - B(t), \text{ say.} \end{aligned}$$

We use Gaussian partial integration to remove the J_{ij}, g_i : If Z is a standard Gaussian, and $F: \mathbb{R} \rightarrow \mathbb{R}$ any decent function, then $\mathbb{E}(ZF(Z)) = \mathbb{E}F'(Z)$. Applying this first to $Z = J_{ij}$, we get

$$\mathbb{E} J_{ij} \frac{\exp[\Phi(t, \sigma)]}{\xi(t)} = \beta \sqrt{\frac{t}{N}} \left\{ \mathbb{E} \frac{\sigma_i \sigma_j \exp[\Phi(t, \sigma)]}{\xi(t)} - \sum_{\tau} \mathbb{E} \frac{\exp[\Phi(t, \sigma) + \Phi(t, \tau)]}{\xi(t)^2} \tau_i \tau_j \right\}.$$

After some elementary manipulations, we obtain

$$A(t) = \frac{\beta^2}{4} \mathbb{E} \sum_{\sigma} \frac{\exp[\Phi(t, \sigma)]}{\xi(t)} - \frac{\beta^2}{4N^2} \mathbb{E} \sum_{\sigma, \tau} \sum_{i, j} \sigma_i \sigma_j \tau_i \tau_j \frac{\exp[\Phi(t, \sigma) + \Phi(t, \tau)]}{\xi(t)^2}.$$

The first summand is simply $\beta^2/4$. In the second, we remark that $N^{-2} \sum_{i, j} \sigma_i \sigma_j \tau_i \tau_j$ is $R_N(\sigma, \tau)^2$. The summation over σ, τ is therefore just the expectation of $R_N(\sigma, \tau)^2$ under the product Gibbs measure at parameter t . We therefore have

$$A(t) = \frac{\beta^2}{4} \left(1 - \nu_{N,t}^{(2)}(R_N^2(\sigma, \tau)) \right),$$

where $\nu_{N,t}^{(2)}$ is the measure (16), but with the interpolated Hamiltonian at t . (We sometimes do not clearly distinguish between a measure and taking the expectation with respect to it.) Similarly, we get

$$B(t) = \frac{\beta^2 q}{2} \left(1 - \nu_{N,t}^{(2)}(R_N(\sigma, \tau)) \right),$$

and therefore

$$(21) \quad \begin{aligned} \frac{d\phi}{dt} &= \frac{\beta^2}{4} \nu_{N,t}^{(2)}(1 - R_N^2(\sigma, \tau) - 2q(1 - R_N(\sigma, \tau))) \\ &= \frac{\beta^2}{4} \left\{ (1 - q)^2 - \nu_{N,t}^{(2)}((R_N(\sigma, \tau) - q)^2) \right\} \\ &\leq \frac{\beta^2}{4} (1 - q)^2. \end{aligned}$$

$$\phi(1) - \phi(0) \leq \frac{\beta^2}{4} (1 - q)^2.$$

We therefore get from (20) for *any* N and *any* $q \geq 0$

$$f_N(\beta, h) \leq \frac{\beta^2}{4} (1 - q)^2 + \int \log \cosh(h + \beta\sqrt{q}x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + \log 2.$$

Taking the infimum over q of the right hand side implies the theorem. \square

The proof does not only give the desired result, but gives also an expression of the difference, namely

$$(22) \quad RS(\beta, h) - f_N(\beta, h) = \frac{\beta^2}{4} \int_0^1 \nu_{N,t}^{(2)} ((R_N(\sigma, \tau) - q)^2) dt.$$

In order to prove that $f(\beta, h) = RS(\beta, h)$, one therefore “only” has to show that for the optimal q (i.e. the one given by (8)), one has $R_N(\sigma, \tau) \simeq q$ with large $\nu_{N,t}^{(2)}$ -probability, at least in the t -average. This is not true for large β , but it is true for small β , as we will discuss later.

It should also be emphasized that the Gibbs measure \mathcal{G}_t structurally is not much different from the original measure. In fact it is of the form

$$(23) \quad \frac{1}{Z} \exp \left[\beta' H_N(\sigma) + \gamma \sum_i g_i \sigma_i + h \sum_i \sigma_i \right],$$

where the g_i are new independent Gaussians, and γ is an additional parameter.

4. DERRIDA’S RANDOM ENERGY MODEL AND RUELLE’S CASCADES

This section is a deviation from the SK model and introduces a class of simple models invented by Derrida which in a certain vague sense are supposed to be “universal attractors” of much more complicated models like SK. On a mathematical level this is very far from being understood. Nonetheless, computations on Derrida’s model (in Ruelle’s asymptotic version) can be used to give a transparent proof of Guerra’s bound of the free energy by the Parisi expression. We will explain this in the next section. This application is not the only reason that we spend some space explaining Derrida’s models. At the core lies one of the main open problems in spin glass theory, namely the claim of universal ultrametricity. Derrida’s models are ultrametric by construction, and as Parisi’s formula is closely connected with them, as we will see, it appears natural to conclude, that ultrametricity should hold. However, there seems to be no promising idea around how to prove that.

The basic difficulty of the SK-model is coming from the fact that the “energies” $H(\sigma)$ are correlated random variables. Derrida [7] realized that already something interesting is happening assuming that they are just independent random variables having (about) the correct variances. Therefore, we consider independent Gaussian random variables $(H_N(\alpha))_{\alpha \in \Sigma_N}$. Σ_N does not need to have any structure here, so we just let $1 \leq \alpha \leq 2^N$. In order to match the variance of the Hamiltonian in the SK-case, we should take $N/2$,

but for convenience, we take variance N , and define the partition function, the free energy, and the Gibbs measure in the usual way

$$(24) \quad Z_N(\beta) \stackrel{\text{def}}{=} \sum_{\alpha} e^{\beta H_N(\alpha)}$$

$$(25) \quad f(\beta) \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta), \quad \mathcal{G}_{N,\beta}(\alpha) \stackrel{\text{def}}{=} Z_N(\beta)^{-1} e^{\beta H_N(\alpha)}$$

It is easy to see that the free energy is self-averaging, so that $f(\beta)$ is also the limit of the expectations, and therefore non-random. The Gibbs measure is again a *random* probability distribution on Σ_N , as the $H(\sigma)$ are random variables. The limiting free energy is not difficult to determine and is given by

$$(26) \quad f(\beta) = \begin{cases} \beta^2/2 + \log 2 & \text{for } \beta \leq \beta_{\text{cr}} = \sqrt{2 \log 2} \\ \sqrt{2 \log 2} \beta & \text{for } \beta \geq \beta_{\text{cr}} = \sqrt{2 \log 2}. \end{cases}$$

Much more interesting is the Gibbs distribution in the $N \rightarrow \infty$ limit. This can be derived from a well known probabilistic result on extreme values of i.i.d. Gaussian random variables. There exists a sequence $a_N \uparrow \infty$ (the exact value is of no importance, they are of order $\sqrt{2 \log 2N}$) such that the random measure

$$\sum_{\alpha} \delta_{H_N(\alpha) - a_N}$$

converges weakly to a Poisson point process on \mathbb{R} with intensity measure

$$\sqrt{2 \log 2} e^{-\sqrt{2 \log 2} t} dt.$$

We write PPP ($t \rightarrow ae^{-at}$) for a Poisson point process with such a density. Remark that there is a largest point, simply because ae^{-at} is integrable at $+\infty$. In contrast, there is no smallest point, and the points are lying dense and denser the further one goes down the negative real axis. We can represent such a point process as $\sum_{i=0}^{\infty} \delta_{\xi_i}$, where $\xi_1 > \xi_2 > \dots$ are real-valued random variables. We also just talk of the “point process (ξ_i)”, meaning $\sum_{i=0}^{\infty} \delta_{\xi_i}$, but we tacitly always assume that the points are ordered downwards. (The point processes we consider will always have a largest point.) We are not really interested in the energy levels, but rather in the Gibbs weights, which are given as $\exp[\beta H_N(\alpha)]$. As we are interested only in the relative weights, we can as well consider $\exp[\beta(H_N(\alpha) - a_N)]$. Of course, we could normalize the weights to a (random) probability distribution, but it turns out to better be not too hasty with that, and to consider first the limiting point process of these points which evidently converges in distribution to the transformation of the point process PPP ($t \rightarrow \sqrt{2 \log 2} e^{-\sqrt{2 \log 2} t}$) obtained by applying the mapping $\xi \rightarrow \eta \stackrel{\text{def}}{=} e^{\beta \xi}$ to the points. This is a PPP ($t \rightarrow x t^{-x-1}$), with the parameter $x = x(\beta) = \sqrt{2 \log 2} / \beta$, i.e. we have

$$(27) \quad \sum_{\alpha} \delta_{\exp[\beta(H_N(\alpha) - a_N)]} \rightarrow \text{PPP} (t \rightarrow x(\beta) t^{-x(\beta)-1})$$

in distribution. The PPP ($t \rightarrow xt^{-x-1}$) (which of course are point processes on the positive real line) have a number of remarkable properties which are absolutely crucial for their appearance in the Parisi picture.

PROPOSITION 4.1. — *Assume (η_i) are the points of a PPP ($t \rightarrow xt^{-x-1}$), and let Y_1, Y_2, \dots be i.i.d. positive real random variables satisfying $EY^x < \infty$, being also independent of the point process. Set $\psi(x) \stackrel{\text{def}}{=} (EY^x)^{1/x}$. Then $\sum_i \delta_{\psi(x)^{-1}Y_i\eta_i}$ is also a PPP ($t \rightarrow xt^{-x-1}$). In plain words, multiplying the points η_i by Y_i amounts to the same (when regarded as a point process) than multiplying the points with the constant $\psi(x)$. (We will see that this property is at the core of the Parisi formula).*

The proof is an easy exercise and I do not give it here. Note that the properties crucially depend on the special form of the intensity measure of the Poisson process. The property actually characterizes PPP ($t \rightarrow xt^{-x-1}$) as has recently been shown by Ruzmaikina and Aizenman [17].

In order to describe the limiting Gibbs distribution, one still has to apply a normalization, and it is plausible that we can interchange the normalizing operation with taking the limit in (27), i.e. we would like to conclude that the point process $\sum_\alpha \delta_{\mathcal{G}_{N,\beta}(\alpha)}$ converges weakly to the proper normalization of PPP ($t \rightarrow xt^{-x-1}$). There is however a difficulty. Let $\eta_1 > \eta_2 > \dots > 0$ be the ordered (random) points of a PPP ($t \rightarrow xt^{-x-1}$). We would like to apply a normalization procedure by normalizing the weights η_i , setting

$$\bar{\eta}_i = \eta_i / \sum_j \eta_j .$$

This we can only do if the sum converges. One easily proves the following statement for the points of a PPP ($t \rightarrow xt^{-x-1}$)

$$\sum_j \eta_j < \infty \text{ a.s. } \iff x < 1.$$

If $x < 1$, we can therefore define the normalization procedure, obtaining the point process $\sum_i \delta_{\bar{\eta}_i}$ which we denote by $\mathcal{N}(\text{PPP}(t \rightarrow xt^{-x-1}))$. This is no longer a Poisson point process as is evident from the fact that the points sum up to 1. The following result is plausible, but its proof still requires some work as the above normalization is not a continuous operation.

PROPOSITION 4.2. — *Assume $\beta > \sqrt{2 \log 2}$. Then $\sum_\alpha \delta_{\mathcal{G}_{N,\beta}(\alpha)}$ converges weakly to $\mathcal{N}(\text{PPP}(t \rightarrow xt^{-x-1}))$, where $x(\beta) = \sqrt{2 \log 2} / \beta$.*

For a proof (in a more general setting), see [6] or [19], Chap 1. The result states that for low temperature, there are configurations α which have Gibbs weight of order 1 in the $N \rightarrow \infty$ limit, but these Gibbs weights stay random. So the limiting Gibbs distribution is not “self-averaging”. Furthermore, there is a “countable” number of such configurations in the limit. More precisely: For any $\varepsilon > 0$ there exists a number $K(\varepsilon)$ such that the total Gibbs weight of the $K(\varepsilon)$ configurations with the largest weight is $\geq 1 - \varepsilon$, with \mathbb{P} -probability larger than $1 - \varepsilon$, and that uniformly in N . Furthermore $K(\varepsilon)$ has to go to ∞ for $\varepsilon \rightarrow 0$. The situation is easy to understand: For $\beta > \sqrt{2 \log 2}$,

the Gibbs weights concentrate on the configurations α for which the energies $H_N(\alpha)$ are maximal or close to the maximum. These energies (near the maximum) are spaced at a distance of order 1: The second largest is below the largest by a random distance which stays stochastically of order 1 in the $N \rightarrow \infty$ limit. The maximum energy is approximately at $\sqrt{2 \log 2} N$, with some correction of order $\log N$.

If $\beta < \sqrt{2 \log 2}$, the situation is completely different. The main contribution comes from energies approximately at a level aN , where $a < \sqrt{2 \log 2}$ (actually $a = \beta$, by accident). At this level, the energies are lying tightly, with exponentially small typical spacings. Therefore, the maximum Gibbs weight for $\beta < \sqrt{2 \log 2}$ is exponentially small in N , and in order to catch a macroscopic weight one has to sum over exponentially many individual configurations. Therefore, in the limit, “uncountable” many configurations contribute to the Gibbs measure.

A prediction of the Parisi theory is that the point process described above is a universal object in spin glass theory and appears as the distribution of the “pure states” in essentially all systems exhibiting “spin glass behavior”, in particular in the SK model. It is difficult to give the notion of a “pure state”, which is often appearing in the physics literature, a precise mathematical sense. This has been achieved only for the p -spin SK model which has a simpler structure than the regular SK model, by Talagrand (see [19], Chap 6).

The above model is called the “random energy model”, REM for short. It is certainly an oversimplification, and Derrida [8] a bit later introduced a model which has hierarchical organized correlations. Shortly afterwards, Ruelle [16] in an attempt to get a clearer mathematical picture of the physicists predictions in spin glass theory introduced a point process version, which is the limiting object of Derrida’s model. This model was then further investigated in [4] and elsewhere. These models are now called “generalized random energy models”, or GREM for short. In contrast to the random energy model, they have a non-trivial notion of “overlaps”.

Here is Derrida’s version. We consider a tree with 2^N leaves and K branching levels, where K stays fixed (for the moment), and we let then $N \rightarrow \infty$. We write the elements of the tree as $\alpha = (\alpha_1, \dots, \alpha_K)$ where $\alpha_i \in \{1, \dots, 2^{N/K}\}$. For convenience, we always assume that N/K is an integer. We again write Σ_N for the collection of such α ’s. Evidently, we have 2^N elements in Σ_N . For $i \leq K$, we identify $(\alpha_1, \dots, \alpha_i)$ with the “bond” from node $(\alpha_1, \dots, \alpha_{i-1})$ to $(\alpha_1, \dots, \alpha_i)$. To the bonds of the tree, we attach Gaussian random variables with variances proportional to N , but depending on the level inside the tree. We choose parameters $\sigma_1^2, \dots, \sigma_K^2 > 0$ with $\sum_i \sigma_i^2 = 1$, and for $i \leq K$, $(\alpha_1, \dots, \alpha_i)$ as above, we choose Gaussian random variables $X_{\alpha_1, \dots, \alpha_i}^{(i)}$ which have variance $\sigma_i^2 N$. All these variables are independent. Then we define the random Hamiltonian

$$(28) \quad H_N(\alpha) \stackrel{\text{def}}{=} \sum_{i=1}^K X_{\alpha_1, \dots, \alpha_i}^{(i)},$$

i.e. for any leaf of the tree we sum the independent Gaussian variables attached to the bonds along the path from the root to this leaf.

The $H_N(\alpha)$ are evidently Gaussians with variance N , like in the REM case, but there are now correlations. Defining for $\alpha, \alpha' \in \Sigma_N$ the overlap

$$R(\alpha, \alpha') \stackrel{\text{def}}{=} \max \{i : (\alpha_1, \dots, \alpha_i) = (\alpha'_1, \dots, \alpha'_i)\}$$

one has

$$\mathbb{E}H_N(\alpha)H_N(\alpha') = N \sum_{i=1}^{R(\alpha, \alpha')} \sigma_i^2.$$

We impose the following condition

$$(29) \quad \sigma_1^2 > \sigma_2^2 > \dots > \sigma_K^2 > 0.$$

If this is not satisfied, just some levels disappear in the $N \rightarrow \infty$ limit, so we can as well make this assumption⁽²⁾. The partition function and the Gibbs weights are defined as before in (24), (25). Despite its simplicity the model has a number of surprising properties which can be summarized as follows:

- The limiting Gibbs measure is always that of random energy model if “properly interpreted”. This is true despite the fact that the limiting point process of the energy levels is *not* that of a random energy model, but by normalizing, the “not-REM” part cancels out.
- The model keeps a non-trivial overlap structure of the configurations, even in the $N \rightarrow \infty$ limit, which also stays random (i.e. non-selfaveraging). Surprisingly however, the overlap structure becomes stochastically independent of the Gibbs-weights in the limit (which is not true for finite N). The overlap structure has a simple Markovian structure as a coalescent with explicitly defined transition probabilities, a fact worked out in [4].
- The GREM overlap structure is of direct relevance for the Sherrington-Kirkpatrick model. It turns out that the Parisi-formula for the free energy of the SK shows up through a “one spin perturbation” of GREMs. This aspect can be used to prove that the Parisi-expression is an upper bound for $f_{SK}(\beta, h)$. This will be explained in the next section.

The free energy of the GREM can be computed (see [8]), but it is not of great relevance for the aspects discussed here.

We now describe Ruelle’s limiting point process versions. One can perform the $N \rightarrow \infty$ separately on each level. On the first level, one simply has $2^{N/K}$ independent Gaussians with variance $\sigma_1^2 N$. After subtracting a suitable constant from these “energies”, one arrives in the limit $N \rightarrow \infty$, using the same argument as for the REM, at a PPP ($t \rightarrow ae^{-at}$), with $a = \sqrt{2 \log 2} / \sqrt{K \sigma_1^2}$. As we are interested in the Gibbs distribution we can as well consider the point process where one maps the points ξ of

⁽²⁾There is a delicate issue in case of equalities in (29) which we do not address here.

the above point process to $\eta = \exp[\beta\xi] \in \mathbb{R}^+$. This leads to a PPP ($t \rightarrow x_1 t^{-x_1-1}$) on the positive real line where $x_1 = \sqrt{2\log 2} / \sqrt{K}\beta\sigma_1$.

In this way, one proceeds along the tree and arrives at the following object. Set

$$x_i(\beta) \stackrel{\text{def}}{=} \frac{\sqrt{2\log 2}}{\sqrt{K}\beta\sigma_i}.$$

The point process PPP ($t \rightarrow x_1 t^{-x_1-1}$) for the first level consists of countably many random points which we can assume to be ordered downwards. Call them $\eta_1^1 > \eta_2^1 > \dots > 0$. For any $i \in \mathbb{N}$, i.e. for any point from the first level, we choose independent point processes PPP ($t \rightarrow x_2 t^{-x_2-1}$), whose points we again order downwards: $\eta_{i,1}^2 > \eta_{i,2}^2 > \dots > 0$, and in this way we proceed: For $j \leq K$ and i_1, \dots, i_{j-1} fixed, $(\eta_{i_1 i_2 \dots i_j}^j)_{i_j \in \mathbb{N}}$ are the points of a PPP ($t \rightarrow x_j t^{-x_j-1}$). They are independent for different i_1, \dots, i_{j-1} and also for different levels j , and we again assume that for each of these point processes, the points are ordered downwards. This is essentially Ruelle’s cascade construction.

We can compose these point processes of the individual levels by just multiplying the “abstract Gibbs weights” along the tree (which corresponds to summing the energy levels of Derrida’s GREM along the tree). We therefore arrive at a point process with random points indexed by $\mathbf{i} = (i_1, \dots, i_K)$, $i_j \in \mathbb{N}$,

$$(30) \quad \eta_{\mathbf{i}} = \eta_{i_1}^1 \eta_{i_1, i_2}^2 \cdot \dots \cdot \eta_{i_1, i_2, \dots, i_K}^K.$$

This is not a Poisson point process, but after normalization, surprisingly, it is simply a normalized REM. For convenience assume $\beta > \beta_K^{\text{cr}} \stackrel{\text{def}}{=} \sqrt{2\log 2} / \sqrt{K}\sigma_K$, so that $x_K(\beta) < 1$. (If this is not satisfied, one has to collapse some of the latter levels and one arrives essentially to the same conclusion for the remaining ones.) In that case

$$\sum_{\mathbf{i}} \eta_{\mathbf{i}} < \infty$$

with probability one, and so one can normalize the point process, defining

$$\bar{\eta}_{\mathbf{i}} \stackrel{\text{def}}{=} \frac{\eta_{\mathbf{i}}}{\sum_{\mathbf{j}} \eta_{\mathbf{j}}}.$$

Then one has the following properties:

- The point process

$$\Xi = \sum_{\mathbf{i}} \delta_{\bar{\eta}_{\mathbf{i}}}$$

is the normalization of a PPP ($t \rightarrow x_K t^{-x_K-1}$).

- The point process of the Gibbs-distributions of Derrida’s GREM converges weakly to the point process of Ruelle’s GREM:

$$\sum_{\alpha} \delta_{\mathcal{G}_{\beta}(\alpha)} \rightarrow \Xi,$$

in distribution, as $N \rightarrow \infty$. For a proof, see [6].

The point process Ξ does not keep track of the way the points were produced through the tree, so it “forgets” the tree structure. This structure is however important for the Parisi picture. The tree structure can be retained in the following way. As usual we order the energy levels η_i downwards, i.e. we define a (random) bijection $\pi : \mathbb{N} \rightarrow \mathbb{N}^K$ such that $\eta_{\pi(k)}$ is the k -th largest element in the set $\{\eta_i\}$. This leads to an overlap structure on \mathbb{N} , by measuring the hierarchical distance between $\pi(i)$ and $\pi(i')$, i.e. we set for $i, i' \in \mathbb{N}$

$$q(i, i') \stackrel{\text{def}}{=} \max \{r : \pi(i)_1 = \pi(i')_1, \dots, \pi(i)_r = \pi(i')_r\}.$$

This leads to a sequence of (random) partitions of \mathbb{N} , which for $k \leq K - 1$ clumps together points in \mathbb{N} whose π -value agrees on level k , i.e. we introduce the equivalence relation

$$i \sim_k i' \iff q(i, i') \geq k,$$

which leads to a partition of \mathbb{N} in the equivalence classes of \sim_k . If k decreases, the partitions become coarser. For $k = 0$, evidently all of \mathbb{N} is clumped into one set. Remarkably, this sequence of random partitions is stochastically independent of Ξ itself. Furthermore the sequence of clustering has a very simple Markovian structure, of viewed backwards in k (see [4]).

5. GUERRA’S REPLICA SYMMETRY BREAKING BOUND: THE AIZENMAN-SIMS-STARR PROOF

In a remarkable paper [9], Guerra extended the bound derived in Section 3 to a bound of $f(\beta, h)$ by the Parisi solution. The proof is not very complicated, but hard to understand without knowledge of the cascade picture introduced in the last section. A bit later, Aizenman, Sims, and Starr [2] reproved the bound, and generalized it by introducing what they call “random overlap structures”, which serve as an abstract model for measures on a countable set which have a notion of “overlaps”.

DEFINITION 5.1. — *A **random overlap structure** \mathcal{R} (ROSt for short) consists of a finite or countable set A , a probability space $(\Gamma, \mathcal{G}, \mathbb{P})$, and random variables $\eta_\alpha \geq 0$, $q_{\alpha, \alpha'}$, $\alpha, \alpha' \in A$, satisfying the following properties*

1. $\sum_\alpha \eta_\alpha < \infty$
2. $(q_{\alpha, \alpha'})$ is positive definite and satisfies $q_{\alpha, \alpha} = 1$.

The η_α play the rôle of (unnormalized) Gibbs weights, and the q ’s are the abstract overlaps.

Example 5.2. — As an example take $A = \Sigma_N \stackrel{\text{def}}{=} \{-1, 1\}^N$. The η_σ , $\sigma \in \Sigma_N$, can be arbitrary. For $q_{\sigma, \sigma'}$ we take the standard overlap $R_N(\sigma, \sigma')$, as introduced before. We write $\mathcal{R}_N^{\text{SK}}$ for this overlap structure. The q here are nonrandom. On the other hand, we can use a (random) reordering of the set A by ordering the η_σ downwards:

$\eta_1 > \eta_2 > \dots > \eta_{2^N}$. After this random reordering, the q become random: $q_{1,2}$ for instance is the overlap of the two indices with the largest η -weight.

Example 5.3. — Another overlap structure is defined by Ruelle’s probability cascades (30) introduced in the last section. Fix $0 = m_0 < m_1 < \dots < m_K = 1$. We take $A = \mathbb{N}^K$, and the η are the (unnormalized) weights $\eta_{\mathbf{i}}$ as in the last section with $x_i \stackrel{\text{def}}{=} m_i$, $1 \leq i \leq K$ (see (30)). There is a slight problem because we have to take the last parameter $x_K = 1$, which implies that $\sum_{\mathbf{i}} \eta_{\mathbf{i}} = \infty$. This will not cause any difficulties for what we do below. The overlaps are defined in the following way. Fix a sequence $0 \leq q(1) < q(2) < \dots < q(K) < q(K+1) = 1$, and we set

$$q_{\mathbf{i},\mathbf{i}'} = q(\max\{k : (i_1, \dots, i_k) = (i'_1, \dots, i'_k)\} + 1),$$

i.e. we measure the hierarchical distance on the tree, and weight it with the function q . For this random overlap structure, we write $\mathcal{R}_K^{\text{Ruelle}}$.

Given any ROSt, we attach to it families of Gaussian random variables $(y_{\alpha,i})_{\alpha \in A, i \in \mathbb{N}}$, $(\kappa_{\alpha})_{\alpha \in A}$ by requiring

$$(31) \quad \mathbb{E}(\kappa_{\alpha}\kappa_{\alpha'}) = q_{\alpha,\alpha'}^2/2,$$

and the “cavity field” by

$$(32) \quad \mathbb{E}(y_{\alpha,j}y_{\alpha',j'}) = q_{\alpha,\alpha'}\delta_{j,j'}.$$

The κ and the y are independent. In case, the q ’s themselves are random variables, these are just the conditional distributions, given (ξ, q) . It is not difficult to see that such random variables exist. By an extension of the probability space, we can assume that all the random variables are defined on a single probability space.

For later use, we give the construction of the cavity variables for $\mathcal{R}_K^{\text{Ruelle}}$. We simply write

$$(33) \quad y_{\mathbf{i}} = g^{(0)} + \sum_{k=1}^K g_{i_1, \dots, i_k}^{(k)},$$

where the g ’s are independent centered Gaussians, with $\text{var}(g^{(0)}) = q(1)$, $\text{var}(g^{(k)}) = q(k+1) - q(k)$. Furthermore, the $y_{\mathbf{i},j}$, $j \in \mathbb{N}$, are independent copies of $y_{\mathbf{i}}$. The $\kappa_{\mathbf{i}}$ are constructed in a similar way.

The above notion of a ROSt needs some explanation. The basic idea comes from what in the physics literature is called the “cavity method”. We consider the standard SK-Hamiltonian, but now with $N + M$ spins, where one should think of N being much larger than M . We then try to write the Hamiltonian in terms of the Hamiltonian on N

spin variables acting on the M “newcomers”. We write $\tau_i = \sigma_{N+i}$ for the newcomers.

$$\begin{aligned} & \frac{\beta}{\sqrt{N+M}} \sum_{i < j \leq N+M} J_{ij} \sigma_i \sigma_j + h \sum_{i=1}^{N+M} \sigma_i \\ &= \frac{\sqrt{N}}{\sqrt{N+M}} \frac{\beta}{\sqrt{N}} \sum_{i < j \leq N} J_{ij} \sigma_i \sigma_j + h \sum_{i=1}^N \sigma_i + \frac{\beta}{\sqrt{N+M}} \sum_{j=1}^M \left(\sum_{i=1}^N J_{i,N+j} \sigma_i \right) \tau_j \\ &+ \frac{\beta}{\sqrt{N+M}} \sum_{i < j \leq M} J_{N+i,N+j} \tau_i \tau_j + h \sum_{j=1}^M \tau_j. \end{aligned}$$

If $M \ll N$, then we can neglect the interaction among the newcomers, i.e. we can drop the fourth summand on the right hand side above. Furthermore, we may as well replace $\sqrt{N+M}$ by \sqrt{N} in the third summand. Defining the cavity variables

$$y_{\sigma,j} \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{i=1}^N J_{i,N+j} \sigma_i,$$

we see that they have exactly the right distribution as required in (32), with respect to the random overlap structure $\mathcal{R}_N^{\text{SK}}$ coming from the N system. In the first summand, we have to be more careful: $N^{-1/2} \sum_{i < j \leq N} J_{ij} \sigma_i \sigma_j$ has variance $(N-1)/2$, and $(N+M)^{-1/2} \sum_{i < j \leq N} J_{ij} \sigma_i \sigma_j$ has variance $(N+M)^{-1} N(N-1)/2 \simeq (N-1)/2 - M/2$. We can therefore (approximately) represent the former by the latter plus an independent Gaussian $\sqrt{M/2} \kappa_\sigma$, where $(\kappa_\sigma)_{\sigma \in \Sigma_N}$ is a field with the covariances given by (31). If we set

$$\eta_\sigma \stackrel{\text{def}}{=} \exp \left[\frac{\beta}{\sqrt{N+M}} \sum_{i < j \leq N} J_{ij} \sigma_i \sigma_j + h \sum_{i=1}^N \sigma_i \right],$$

we see that

$$(34) \quad \frac{Z_{N+M}}{Z_N} \simeq \frac{\sum_{\sigma \in \Sigma_N, \tau \in \Sigma_M} \eta_\sigma \exp \left[\sum_{i=1}^M (\beta y_{\sigma,i} + h) \tau_i \right]}{\sum_{\sigma \in \Sigma_N} \eta_\sigma \exp \left[\beta \sqrt{M/2} \kappa_\sigma \right]}.$$

Here we have used the ROSt from the N -spin SK model (with Gibbs weights coming from a slightly changed temperature parameter). Aizenman, Sims and Starr had the idea to investigate the above object when the N system is replaced by an *arbitrary* ROSt \mathcal{R} , and they considered the “relative finite M free energy” in the following way

$$(35) \quad G_M(\beta, h, \mathcal{R}) \stackrel{\text{def}}{=} \frac{1}{M} \mathbb{E} \left(\log \frac{\sum_{\alpha, \tau \in \Sigma_M} \eta_\alpha \exp \left[\sum_{j=1}^M (\beta y_{\alpha,j} + h) \tau_j \right]}{\sum_{\alpha} \eta_\alpha \exp \left[\beta \sqrt{M/2} \kappa_\alpha \right]} \right),$$

where the \mathbb{E} expectation is taken with respect both to the law of the random overlap structure and the cavity variables $y_{\alpha,i}$ and κ_α . The variant of Guerra’s theorem in the generalized version of Aizenman, Sims and Starr is the following remarkable inequality, which holds for *any* M , and *any* random overlap structure.

THEOREM 5.4. — For any M , and any random overlap structure \mathcal{R} one has

$$(36) \quad f_M(\beta, h) \leq G_M(\beta, h, \mathcal{R}).$$

Remark 5.5. — Aizenman, Sims and Starr actually show that

$$f(\beta, h) = \inf_{\mathcal{R}} \lim_{M \rightarrow \infty} G_M(\beta, h, \mathcal{R}),$$

which follows by plugging in the SK-random overlap structure itself, using (34), and applying the subadditivity result of Guerra-Toninelli (3).

I am not going to prove the theorem. To a large extent it is a rerun of the computation done above in Section 3. Here is an outline. One uses the following interpolation:

$$H_M(\tau, \alpha, t) \stackrel{\text{def}}{=} \frac{\sqrt{1-t}}{\sqrt{M}} \sum_{1 \leq i < j \leq M} J_{ij} \tau_i \tau_j + \sqrt{\frac{M(1-t)}{2}} \kappa_\alpha + \sqrt{t} \sum_{i=1}^M y_{\alpha, i} \tau_i$$

and defines

$$\hat{G}_M(\beta, h, t, \mathcal{R}) \stackrel{\text{def}}{=} \frac{1}{M} \mathbb{E} \left(\log \frac{\sum_{\alpha \in A, \tau \in \Sigma_M} \eta_\alpha \exp \left[\beta H_M(\tau, \alpha, t) + h \sum_{j=1}^M \tau_j \right]}{\sum_{\alpha \in A} \eta_\alpha \exp \left[\beta \sqrt{M/2} \kappa_\alpha \right]} \right),$$

where \mathbb{E} is taken with respect to the overlap structure *and* the J 's (which are supposed to be independent). For $t = 0$, the κ -part cancels, and one just gets $f_M(\beta, h)$. For $t = 1$, one gets $G_M(\beta, h, \mathcal{R})$. By a computation similar to the one in Section 3, one gets

$$(37) \quad \frac{d\hat{G}_M(\beta, h, \cdot, t, \mathcal{R})}{dt} = \frac{1}{2} \nu_t^{(2)} \left((R_M(\tau, \tau') - q_{\alpha, \alpha'})^2 \right) \geq 0,$$

which immediately implies the theorem. $\nu_t^{(2)}$ is to be understood in the following way: For fixed environment (from J and the ROST), one takes two independent copies of the (τ, α) distributed according to

$$p_t(\tau, \alpha) \stackrel{\text{def}}{=} \frac{\eta_\alpha \exp \left[\beta H_M(\tau, \alpha, t) + h \sum_i \tau_i \right]}{\text{Normalization}},$$

and afterwards, one takes the environment expectation.

In principle, one should of course take $\mathcal{R}_N^{\text{SK}}$ as the random overlap structure, in which case one gets $f(\beta, h)$ by Guerra-Toninelli in the $M \rightarrow \infty$ limit, as remarked above, but one has obviously not gained much. The striking fact however is that the inequality is true for *any* ROST, and one can try to obtain good upper bounds by choosing random overlap structures for which one can compute G_M . It turns out that the good choice is Ruelle's $\mathcal{R}_K^{\text{Ruelle}}$ from Example 5.3:

LEMMA 5.6. —

$$(38) \quad G_M(\beta, h, \mathcal{R}_K^{\text{Ruelle}}) = G_1(\beta, h, \mathcal{R}_K^{\text{Ruelle}}) = \mathcal{P}_K(m, q; \beta, h).$$

Proof. — We give a sketch of the computation. We can handle numerator and denominator in (35) separately. The denominator is simpler, so I will only discuss the numerator. We take $M = 1$. It will be clear from the computation that for general M the outcome is the same. With the representation of the $y_{\mathbf{i}}$ by (33), we get

$$(39) \quad \begin{aligned} \frac{1}{2} \sum_{\mathbf{i}, \tau \in \Sigma_1} \eta_{\mathbf{i}} \exp [(\beta y_{\mathbf{i}} + h) \tau] &= \sum_{\mathbf{i}} \eta_{\mathbf{i}} \cosh (\beta y_{\mathbf{i}} + h) \\ &= \sum_{(i_1, \dots, i_K)} \eta_{i_1}^1 \eta_{i_1 i_2}^2 \cdots \eta_{i_1 i_2 \dots i_K}^K \cosh \left(\beta \sum_{n=0}^K g_{i_1, \dots, i_n}^{(n)} + h \right). \end{aligned}$$

We condition on $\eta_{i_1}^1, \eta_{i_1 i_2}^2, \dots, \eta_{i_1 i_2 \dots i_{K-1}}^{K-1}$ and $g_{i_1}^{(1)}, g_{i_1 i_2}^{(2)}, \dots, g_{i_1 i_2 \dots i_{K-1}}^{(K-1)}$. Then $(\eta_{i_1 i_2 \dots i_K}^K)_{i_K \in \mathbb{N}}$ is a PPP ($t \rightarrow m_K t^{-m_K - 1}$) whose points are multiplied by the independent random variables $\left(\cosh \left(\beta \sum_{n=0}^K g_{i_1, \dots, i_n}^{(n)} + h \right) \right)_{i_K}$. From Proposition 4.1, we know that the conditional law (conditioned on anything up to level $K - 1$) of the point process

$$\left(\eta_{i_1 i_2 \dots i_K}^K \cosh \left(\beta \sum_{n=0}^K g_{i_1, \dots, i_n}^{(n)} + h \right) \right)_{i_K}$$

is the same as that of

$$\left(C_K \left(\beta \sum_{n=0}^{K-1} g_{i_1, \dots, i_n}^{(n)} \right) \eta_{i_1 i_2 \dots i_K}^K \right)_{i_K},$$

where

$$C_K(\xi) \stackrel{\text{def}}{=} [E_Z \cosh^{m_K} (\xi + h + \beta \sqrt{q_{K+1} - q_K} Z)]^{1/m_K}, \quad \xi \in \mathbb{R},$$

Z being a standard Gaussian random variable, and E_Z the expectation with respect to Z . C_K is a random variable which still depends on the $g^{(n)}$, $n \leq K - 1$. We proceed in the same way, replacing $\left(C_K \left(\beta \sum_{n=0}^{K-1} g_{i_1, \dots, i_n}^{(n)} \right) \eta_{i_1 i_2 \dots i_{K-1}}^{K-1} \right)_{i_{K-1} \in \mathbb{N}}$ by $\left(C_{K-1} \left(\beta \sum_{n=0}^{K-2} g_{i_1, \dots, i_n}^{(n)} \right) \eta_{i_1 i_2 \dots i_{K-1}}^{K-1} \right)_{i_{K-1} \in \mathbb{N}}$, where

$$C_{K-1}(\xi) \stackrel{\text{def}}{=} [E_Z C_K^{m_{K-1}} (\xi + h + \beta \sqrt{q_{K-1} - q_{K-2}} Z)]^{1/m_{K-1}},$$

and so on. We finally see that multiplying the points $\eta_{\mathbf{i}}$ by $\cosh(\beta y_{\mathbf{i}} + h)$ amounts (for the corresponding point process) in multiplying the points by the constant $E \log Y_1$ from (13). The denominator in (35) is simpler because there one has in every step just an integration of a Gaussian in the exponent. We therefore see that multiplying the points $\eta_{\mathbf{i}}$ by $\exp [(\beta/\sqrt{2}) \kappa_{\mathbf{i}}]$ simply leads to a multiplication of the point process by $\exp [(\beta^2/4) \sum_{i=0}^K m_i (q_{i+1}^2 - q_i^2)]$. In the definition of G_1 (35), we would now like to argue that $\sum_{\mathbf{i}} \eta_{\mathbf{i}}$ cancels out. There is the slight difficulty that this sum diverges almost surely, because of $m_K = 1$, but we can choose m_K slightly less than 1, in which case

the sum is finite, and so cancels, and then we can let $m_K \rightarrow 1$ in the end. The upshot of this computation is that

$$\begin{aligned} G_1(\beta, h, \mathcal{R}) &= E \log Y_1 - \frac{\beta^2}{4} \sum_{i=0}^K m_i (q_{i+1}^2 - q_i^2) + \log 2 \\ &= \mathcal{P}_K(m, q; \beta, h), \end{aligned}$$

the $\log 2$ is coming from dividing by 2 in (39). It is evident from this computation that we get the same for arbitrary M . (One is just having M factors of $\cosh(\cdot)$ with independent contents, so in every step of the above argument, the factoring remains.) \square

Combining this result with Theorem 5.4, one gets

$$f_M(\beta, h) \leq \mathcal{P}_K(m, q; \beta, h)$$

for any K , and any sequence m and q . Therefore

$$f_M(\beta, h) \leq \inf_{K, m, q} \mathcal{P}_K(m, q; \beta, h).$$

This is Guerra’s upper bound.

6. ON TALAGRAND’S PROOF OF THE PARISI FORMULA

From (37) it is evident that one could prove the Parisi formula, provided one gets the right hand side of this equality under control and shows that it approaches 0 by correctly choosing a sequence of random overlap structures. A difficulty arises from the fact that one does not know from the outset the correct random overlap structures. In his remarkable paper [20], Talagrand had been able to finish this approach and to prove the Parisi formula. This is too complicated to be explained in any details, but I will sketch the strategy in the simplest situation where there is no replica symmetry breaking, i.e. in the region where the replica symmetric formula is correct. In order to prove that, Talagrand considers a replicated system with fixed overlap of the replicas. As one evidently wishes to apply this to the replicated interpolated system of Section 3, we work with measures of the type (23) with the additional parameter γ . Let therefore

$$Z_N^{(2)}(\beta, \gamma, h, u) \stackrel{\text{def}}{=} \sum_{\sigma, \tau \in \Sigma_N: R_n(\sigma, \tau) = u} \exp \left[\beta (H_N(\sigma) + H_N(\tau)) - \sum_i (h + \gamma g_i) (\sigma_i + \tau_i) \right],$$

and

$$(40) \quad r_N(\beta, \gamma, h, u) \stackrel{\text{def}}{=} \frac{1}{N} \mathbb{E} \log Z_N^{(2)}(\beta, \gamma, h, u).$$

(One restricts u to the possible overlaps only, but this is not causing any difficulties).

The strategy is to prove first an upper bound for the free energy r_N of the replicated system with restricted overlaps again through an interpolation procedure similar to the

proof of Guerra’s bound, obtaining a Parisi formula for the replicated system with overlap restrictions. The expression is more complicated than (13). To give the reader an impression of the complexity the procedure gains quickly, I describe the modifications which are needed. We choose K and sequences m, q as in (10), (11), but we need another sequence $(c_j)_{0 \leq j \leq K}$, $c_j \in [-1, 1]$, describing correlations. Instead of the sequence (g_i) used in (13), we now take a sequence of independent two-dimensional Gaussian vectors (g_i^1, g_i^2) , which have the same variances as before, $\text{var}(g_i^j) = \beta^2 (q_{i+1} - q_i)$, but with correlations between g_i^1 and g_i^2 given by the c_i . Then one defines the sequence $(Y_j)_{0 \leq j \leq K+1}$ by downwards recursion similar to that in (12), with additional parameter $x, x_1, x_2 \in \mathbb{R}$, starting with

$$Y_{K+1}(x, x_1, x_2) = \cosh(x) \cosh\left(x_1 + \sum_{i=0}^K g_i^1\right) \cosh\left(x_2 + \sum_{i=0}^K g_i^2\right) + \sinh(x) \sinh\left(x_1 + \sum_{i=0}^K g_i^1\right) \sinh\left(x_2 + \sum_{i=0}^K g_i^2\right),$$

and finally one sets

$$\mathcal{P}_K(m, q, c, x; \beta, h, \gamma, u) = 2 \log 2 + EY_0(x, h + \gamma g, h + \gamma g) - ux - \frac{\beta^2}{2} \sum_{i=1}^K m_i (q_{i+1}^2 - q_i^2) - \frac{\beta^2}{2} \sum_{i=1}^K m_i (d_{i+1}^2 - d_i^2) + \frac{\beta^2}{2} (u - d_{k+1})^2,$$

where g is an additional standard Gaussian, and $d_k \stackrel{\text{def}}{=} \sum_{i=0}^{k-1} c_i (q_{i+1} - q_i)$. Then with

$$(41) \quad U(\beta, \gamma, h, u) \stackrel{\text{def}}{=} \inf_{K, q, m, c, x} \mathcal{P}_K(m, q, c, x; \beta, h, \gamma, u)$$

one proves

$$(42) \quad r_N(\beta, \gamma, h, u) \leq U(\beta, \gamma, h, u)$$

in a way parallel to the one for Guerra’s replica symmetry breaking bound.

Fix now some parameter (β, h) . Talagrand’s strategy to show $f(\beta, h) = RS(\beta, h)$ (which is true only for small enough β) is to prove that for the parameter from the interpolation in Section 3, i.e. $\beta_t = \sqrt{t}\beta$, $\gamma_t = \sqrt{1-t}\sqrt{q}$, q being the “correct one”, i.e. the solution of equation (8), one has

$$\limsup_{N \rightarrow \infty} r_N(\beta_t, \gamma_t, h, u) < 2f(\beta_t, \gamma_t, h),$$

for $u \neq q$, where $f(\beta, \gamma, h)$ is the free energy of the slightly extended SK-model defined in (23). By (42), this follows from

$$(43) \quad U(\beta_t, \gamma_t, h, u) < 2f(\beta_t, \gamma_t, h).$$

This would lead to the conclusion that the probability for the replicated system (at interpolation parameter t) not to nearly match the correct q with the overlap, is negligible. By (21), this then implies that $f(\beta, h) = RS(\beta, h)$ (essentially, of course, as there are many details to fit in). At first sight, this seems to be a hopeless enterprise, as on the right hand side of the desired inequality, one has the free energy one is shooting for. Talagrand however remarked that it essentially suffices to prove an inequality where the right hand side in (43) is replaced by $2RS(\beta_t, \gamma_t, h)$.

The main line of argument is running as follows: Let $\psi(\beta, h)$ be the Parisi solution,

$$\psi(\beta, h) = \inf_{K, q, m} \mathcal{P}_K(m, q; \beta, h),$$

\mathcal{P}_K from (13), and let $\psi(\beta, \gamma, h)$ be the Parisi solution for the Hamiltonian with the additional parameter γ . Let also $RS(\beta, \gamma, h)$ be the replica symmetric solution for the Hamiltonian with the additional parameter γ . Let β, h be such that $\psi(\beta, h) = RS(\beta, h)$, and furthermore, that this relation is correct on our path joining the coupled system with the uncoupled one, i.e. $\psi(\beta_t, \gamma_t, h) = RS(\beta_t, \gamma_t, h)$, with $\beta_t = \sqrt{t}\beta$, $\gamma_t = \sqrt{(1-t)q}$.

Assume we can prove that

$$(44) \quad \sup_{u: |u-q| \geq y} U(\beta_t, \gamma_t, h, u) \leq 2RS(\beta_t, \gamma_t, h) - \varepsilon y^2$$

for some constant ε (possibly depending on β, h , but not on t), then we get

$$\nu_{N,t}^{(2)} \{ (R_N(\sigma, \tau) - q)^2 \geq y \} \leq e^{o(N)} \exp [2N (RS(\beta_t, \gamma_t, h) - \varepsilon y - f_N(\beta_t, \gamma_t, h))],$$

and plugging in $y = C (RS(\beta_t, \gamma_t, h) - f_N(\beta_t, \gamma_t, h)) + \varepsilon_1$, with $C = 1/\varepsilon$, $\varepsilon_1 > 0$ arbitrary, we get that

$$\nu_{N,t}^{(2)} \{ (R_N(\sigma, \tau) - q)^2 \geq C (RS(\beta_t, \gamma_t, h) - f_N(\beta_t, \gamma_t, h)) + \varepsilon_1 \}$$

is exponentially decaying in N , uniformly in t , with a rate depending on ε_1 . Therefore, we can choose N large enough (depending on β, h, ε_1) such that the above probability is small. Using that, and (22), it is then possible to prove

$$\lim_{N \rightarrow \infty} f_N(\beta, h) = RS(\beta, h).$$

(44) needs considerable work, but given Parisi-like formula (41), and the Guerra type bound (42), it suffices to (very) cleverly plug in parameters K, m, q, c, x on the right hand side to get the desired inequality.

One would hope that the regions where this is correct coincide with the one bounded by the AT-line, i.e. when (9) is satisfied. This however has not yet been proved. Anyway, Talagrand gives a not totally explicit characterization of the region in the (β, h) -plane where this line of arguments works (see [19], p. 153), and where the replica symmetric solution is correct.

In a remarkable tour de force, Talagrand has been able to extend this type of arguments to cover the whole temperature regime, and to verify the Parisi formula in the whole temperature region. An apparent difficulty is coming from the fact that in the low temperature region, one needs to let $K \rightarrow \infty$ in the infimum (14). A very pertinent observation of Talagrand is that along the interpolation, for any fixed $t_0 < 1$, one can work with finitely many symmetry breakings on $[0, t_0]$, where the number of levels of symmetry breaking has to increase to ∞ only when $t_0 \uparrow 1$. Therefore, along the interpolation path, one can work with a finite K , and instead of minimizing over one parameter q as in the high temperature region, one minimizes over $2K - 1$ parameters $q_1, \dots, q_K, m_1, \dots, m_{K-1}$, and tries to use estimates like (44). It goes without saying that the whole argument becomes more complicated than the replica symmetric case by an order of magnitude.

7. SOME OPEN PROBLEMS

Despite the success of the approach by Guerra and Talagrand, the whole field is still very much in its beginning, and there remain many open problems. Much of the recent progress depends on the possibility to derive inequalities of the type explained in Section 3 and Section 5. It is completely open if there exist similar inequalities, perhaps only in an asymptotic sense, for other models like the perceptron, or the Hopfield net. Talagrand had used a different method to obtain results on these models in the high temperature region which is based on the “cavity method”. There is not space enough here to give any explanations of this procedure (see [19], Chap. 3-5). The low temperature region for these models seems to be completely out of reach, presently.

Another open question concerns ultrametricity, even for the SK-model, and for instance to prove (17). A natural approach would be to try to somehow extend the approach used by Talagrand for a tripled system with three replicas $\sigma^1, \sigma^2, \sigma^3$, define a free energy like (40) with the overlaps $R_N(\sigma^1, \sigma^2)$, $R_N(\sigma^1, \sigma^3)$, $R_N(\sigma^2, \sigma^3)$ fixed, and derive a “Guerra type” bound, like (42) with a Parisi-formula like (41), and then to exclude overlaps which do not satisfy the ultrametricity condition. It seems that this natural approach leads to tremendous difficulties, and it has not been possible to make progress along this line, at least to my knowledge.

Another approach would be more in the line of Theorem 5.4 and Remark 5.5, and try to prove that the only random overlap structures which minimize the right hand side of (36) are ultrametric ones. This could follow from a characterization of the Ruelle point process through some invariance properties, like the one given in [17], but including correlations. This had been proposed by Michael Aizenman, but first of all, the necessary extension of [17] seems to be rather delicate, and secondly, it is not clear to me if this really would lead to a proof of ultrametricity in the form (17).

On a much more modest level, there has recently been some progress, namely a proof that some natural “non-hierarchical” generalizations of the GREM are in the limit hierarchical organized (see [5]), but it is very questionable if this sheds much light on similar problems for instance for the SK model.

REFERENCES

- [1] Aizenman, M., Lebowitz, J. and Ruelle, D.: *Some rigorous results on the Sherrington-Kirkpatrick model*. Comm. Math. Phys. **112** (1987), 3-20.
- [2] Aizenman, M., Sims, R. and Starr, S.L.: *An extended variational principle for the SK spin-glass model*. arXiv:cond-mat/0306386.
- [3] de Almeida, J.R.L. and Thouless, D.J.: *Stability of the Sherrington-Kirkpatrick solution of spin glasses*. J. Phys. A **11** (1978), 983.
- [4] Bolthausen, E. and Sznitman, A.-S.: *On Ruelle’s probability cascades and an abstract cavity method*. Comm. Math. Phys. **197** (1998), 247-276.
- [5] Bolthausen, E. and Kistler, N.: *On a non-hierarchical version of the Generalized Random Energy Model*. To appear in Ann. Appl. Prob.
- [6] Bovier, A. and Kurkova, I.: *Derrida’s Generalized Random Energy Models I & II*. Annales de l’Institut Henri Poincaré **40** (2004), 439-495.
- [7] Derrida, B.: *Random energy model: An exactly solvable model of disordered systems*. Phys. Rev. B **24** (1981), 2613-2626.
- [8] Derrida, B.: *A generalization of the random energy model that includes correlations between the energies*. J. Physique. Lett **46** (1986), 401-407.
- [9] Guerra, F.: *Replica broken bounds in the mean field spin glass model*. Comm. Math. Phys. **233** (2003), 1-12.
- [10] Guerra, F. and Toninelli, F. L.: *The thermodynamic limit in mean field spin glass models*. Comm. Math. Phys. **230** (2002), 71-79.
- [11] Guerra, F. and Toninelli, F. L.: *Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model*. J. Math. Phys. **43** (2002), 3704-3716.
- [12] Ledoux, Michel: *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs **89**, AMS.
- [13] Mézard, M., Parisi, G. and Virasoro, M.A.: *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.
- [14] Nishimori, H.: *Statistical Physics of Spin Glasses and Information Processing*. Oxford Science Publications, 1999.
- [15] Parisi, G.: *A sequence of approximate solutions to the S-K model for spin glasses*. J. Phys. A **13**, L-115 (1980).
- [16] Ruelle, D.: *A mathematical reformulation of Derrida’s REM and GREM*. Comm. Math. Phys. **108** (1987), 225-239.

- [17] Ruzmaikina, A. and Aizenman, M.: *Characterization of invariant measures at the leading edge for competing particle systems*. Preprint (2005).
- [18] Sherrington, D., and Kirkpatrick, S.: *Solvable model of a spin glass*. Phys. Rev. Lett. **35** (1972), 1792-1796 .
- [19] Talagrand, M.: *Spin Glasses: A Challenge for Mathematicians*. Springer, Heidelberg, 2003.
- [20] Talagrand, M.: *The Parisi formula*. To appear in Ann. of Math.

Erwin BOLTHAUSEN
Institute of Mathematics
University of Zurich
Winterthurerstrasse 190
CH-8057 ZURICH
(Suisse)
E-mail : `eb@math.unizh.ch`

**ALGÈBRES SIMPLES CENTRALES
SUR LES CORPS DE FONCTIONS DE DEUX VARIABLES**
[d'après A. J. de Jong]

par **Jean-Louis COLLIOT-THÉLÈNE**

1. INTRODUCTION

À toute algèbre simple centrale A de dimension finie sur un corps F sont associés deux entiers, l'indice et l'exposant. L'indice de A est le minimum des degrés $[E : F]$, où E parcourt les corps, extensions finies de F , pour lesquels la E -algèbre $A \otimes_F E$ est isomorphe à une algèbre de matrices. L'exposant de A est l'ordre de la classe de A dans le groupe de Brauer du corps F . L'exposant divise l'indice, mais ne lui est pas nécessairement égal. Lorsque F est un corps de nombres, c'est un théorème des années 1930 qu'exposant et indice coïncident ([BHN]). A. J. de Jong [dJ] montre qu'ils coïncident aussi lorsque F est un corps de fonctions de deux variables sur le corps des complexes.

Au §1 de cet exposé, on trouvera des rappels sur les algèbres simples centrales sur un corps, le groupe de Brauer, les extensions de ces notions au-dessus d'un schéma. On passe en revue les propriétés connues des corps de déploiement des algèbres simples centrales et on cite plusieurs problèmes ouverts dans cette direction, particulièrement en ce qui concerne la relation entre exposant et indice.

Le §2 décrit l'essentiel de la démonstration de de Jong, qui passe par des déformations d'algèbres d'Azumaya sur une surface, rendues possibles par des transformations élémentaires. On décrit la démonstration « simplifiée » évoquée à la fin de l'article de de Jong. Cette démonstration requiert un bon théorème de type Bertini pour démarrer, mais elle évite les études fines de variétés singulières de [dJ].

Le manque de temps, de place, et de compétence nous ont fait renoncer à décrire la nouvelle et très intéressante preuve, par Lieblich [Lie], du théorème de de Jong dans le cas non ramifié. Cette preuve, qui se place dans le cadre des champs algébriques, s'appuie sur le théorème de Graber, Harris, Starr et de Jong ([GHS], [dJS1]) selon lequel une variété (séparablement) rationnellement connexe sur un corps de fonctions d'une variable sur un corps algébriquement clos a un point rationnel. Il est par ailleurs trop tôt pour parler

d'une troisième démonstration, annoncée par de Jong et Starr, et qui s'inscrit dans l'étude toute nouvelle des variétés 1-rationnellement connexes.

Au §3, on donne une liste de propriétés des groupes algébriques linéaires sur un corps de fonctions de deux variables sur le corps des complexes que la coïncidence entre exposant et indice permet d'obtenir.

Les corps de fonctions de deux variables sur le corps des complexes sont de dimension cohomologique 2. Au §4, on discute divers résultats récents sur des corps qui sont de dimension cohomologique 3, à savoir les corps de fonctions d'une variable sur un corps p -adique.

Pour de nombreuses discussions sur le travail de de Jong, je remercie M. Ojanguren et R. Parimala. Je me suis en particulier fortement inspiré d'un cours donné par cette dernière à Lens en juin 2004. Je remercie O. Gabber, P. Gille, B. Kahn, T. Szamuely et tout particulièrement L. Moret-Bailly pour leurs critiques d'une première version du présent texte.

2. LA THÉORIE CLASSIQUE

2.1. Algèbres simples centrales sur un corps ([A], [D], [Bki], [Bld], [GSz])

Soient k un corps et V un k -espace vectoriel de dimension finie $n \geq 1$. La k -algèbre $\text{End}(V)$ satisfait les propriétés suivantes :

- (a) Elle est de dimension finie, n^2 , sur k .
- (b) Son centre est k .
- (c) Elle ne possède pas d'idéal bilatère non trivial.

Par le choix d'une base de V , $\text{End}_k(V)$ s'identifie à l'algèbre $M_n(k)$ des matrices carrées de dimension n sur le corps k .

Une k -algèbre simple centrale est une forme tordue d'une telle algèbre. Plus précisément :

DÉFINITION 2.1. — *Une k -algèbre simple centrale est une k -algèbre de dimension finie, de centre réduit à k , sans idéal bilatère non trivial.*

Pour une k -algèbre A , les propriétés suivantes sont équivalentes :

- (i) A est une k -algèbre simple centrale.
- (ii) Il existe un corps K contenant k tel que la K -algèbre $A \otimes_k K$ est simple centrale.
- (iii) Pour tout corps K contenant k , la K -algèbre $A \otimes_k K$ est simple centrale.
- (iv) Il existe un corps K contenant k et un entier $n \geq 1$ tels que la K -algèbre $A \otimes_k K$ est K -isomorphe à une K -algèbre $M_n(K)$.
- (v) Il existe un corps K extension finie séparable de k et un entier $n \geq 1$ tels que la K -algèbre $A \otimes_k K$ est K -isomorphe à une K -algèbre $M_n(K)$.

Ceci implique en particulier que la dimension de A sur k est un carré, soit n^2 . L'entier n est appelé le *degré* de A (sur k).

Un corps K comme en (iv) est dit *corps de déploiement* de l'algèbre A . D'après (v), une clôture séparable de k est un corps de déploiement de A .

D'après Wedderburn, toute k -algèbre simple centrale A est isomorphe à une k -algèbre de matrices $M_r(D)$, où D est simple centrale à division, c'est-à-dire que c'est un corps gauche de centre k (on emploiera indifféremment l'une ou l'autre terminologie). Un tel corps gauche D est déterminé, comme k -algèbre, à isomorphisme non unique près.

Le degré de D sur k est appelé l'*indice* de A (sur k). Il est noté $\text{ind}_k(A) = \text{ind}_k(D)$.

Soit A une k -algèbre simple centrale de degré n . Pour tout entier m avec $1 \leq m \leq n$ on peut considérer la k -variété algébrique dont les points sur une clôture séparable k_s de k sont les idéaux à droite de $A \otimes_k k_s$ de k_s -dimension mn . On note cette variété $SB(A, m)$. La k -variété $SB(A, 1)$ est la variété de Severi-Brauer associée par F. Châtelet à la k -algèbre simple centrale A . On a les faits suivants ([Bl]). La k -variété $SB(A, m)$ est une forme de la grassmannienne $\text{Grass}(m, n)$. La k -variété $SB(A, m)$ possède un k -point si et seulement $\text{ind}_k(A)$ divise m . Ainsi l'indice $\text{ind}_k(A)$ est le plus petit entier m tel que la k -variété $SB(A, m)$ possède un k -point. C'est aussi le p.g.c.d. des entiers m satisfaisant cette propriété.

Pour $A = M_r(D)$ comme ci-dessus, et K un corps contenant k , de degré fini sur k , les propriétés suivantes sont équivalentes :

- (i) Être k -isomorphe à un sous-corps commutatif maximal de D .
- (ii) Être de degré minimal sur k parmi toutes les extensions finies de k déployant A .

Les sous-corps commutatifs maximaux de D sont tous de degré $\text{ind}_k(D)$. Il existe de tels sous-corps qui sont séparables sur k .

L'indice $\text{ind}_k(A)$ d'une k -algèbre simple centrale A peut donc aussi se définir comme le degré commun des corps satisfaisant l'une des propriétés ci-dessus. C'est aussi le p.g.c.d. des degrés $[K : k]$ des extensions finies de corps K/k déployant D .

Une k -algèbre simple centrale à division D/k de degré $n = \prod_i p_i^{n_i}$ avec les p_i premiers distincts s'écrit comme un produit tensoriel $D = D_1 \otimes_k \cdots \otimes_k D_n$, chaque D_i étant un corps gauche de degré $p_i^{n_i}$.

Rappelons la notion d'algèbre cyclique. Soit K/k une extension finie cyclique du corps k , de degré n , soit σ un générateur de $\text{Gal}(K/k)$ et soit $b \in k^*$. On munit le k -vecteuriel $\bigoplus_{i=0}^{n-1} K.Y^i$ d'une structure de k -algèbre simple centrale par les relations $Y^n = b$ et $Yx = \sigma(x)Y$ pour $x \in K$. On note $A = (K/k, \sigma, b)$ cette k -algèbre. Lorsque k contient une racine n -ième de 1, soit ζ_n , on peut écrire $K = k(\alpha)$ avec $\alpha^n = a \in k^*$. Soit σ tel que $\sigma(\alpha) = \zeta_n \alpha$. On note alors $A = (a, b)_{\zeta_n}$. Cette k -algèbre, considérée par Dickson, est engendrée par deux éléments X et Y soumis aux relations $X^n = a$, $Y^n = b$, $YX = \zeta_n XY$. Pour $n = 2$, on retrouve la définition des algèbres de quaternions (a, b) .

2.2. Groupe de Brauer d'un corps ([A], [Bki], [GSz], [S2])

Soient A et B deux k -algèbres simples centrales. Le produit tensoriel $A \otimes_k B$ est une k -algèbre simple centrale. Etant donnée une k -algèbre simple centrale A , on dispose de l'algèbre opposée A^{op} , et l'homomorphisme $A \otimes_k A^{op} \rightarrow \text{End}_{k\text{-vect}}(A)$ qui à $a \otimes b$ associe $x \mapsto axb$ est un isomorphisme de k -algèbres. Considérons l'ensemble des classes d'isomorphie de k -algèbres simples centrales. Si l'on introduit la relation d'équivalence « L'algèbre A est équivalente à l'algèbre B s'il existe des entiers r, s avec $M_r(A) \simeq M_s(B)$ », on voit que le produit tensoriel induit sur les classes d'équivalence une structure de groupe abélien, dont l'élément neutre est la classe des algèbres $M_n(k)$ (n arbitraire). C'est le groupe de Brauer $\text{Br}(k)$ du corps k .

Notons $\text{Az}_n(k)$ l'ensemble des classes d'isomorphie de k -algèbres simples centrales de degré n . On dispose d'une application naturelle $\text{Az}_n(k) \rightarrow \text{Br}(k)$. Cette application est une injection : si deux k -algèbres simples centrales A et B de même degré ont même classe dans le groupe de Brauer, elles sont isomorphes. On a donc :

$$\text{Br}(k) = \cup_{n=1}^{\infty} \text{Az}_n(k),$$

la loi de groupe étant induite par les produits tensoriels

$$\text{Az}_n(k) \times \text{Az}_m(k) \rightarrow \text{Az}_{nm}(k).$$

On a une seconde définition du groupe de Brauer du corps k ([S1]). On note k_s une clôture séparable de k , g le groupe de Galois de k_s sur k . Alors

$$\text{Br}(k) = H^2(g, k_s^*).$$

On passe de l'une à l'autre définition en utilisant la cohomologie galoisienne de la suite exacte de k -groupes lisses :

$$1 \rightarrow \mathbb{G}_{m,k} \rightarrow GL_{n,k} \rightarrow PGL_{n,k} \rightarrow 1.$$

L'ensemble pointé de cohomologie galoisienne $H^1(k, PGL_n) = H^1(g, PGL_n(k_s))$ est en bijection avec $\text{Az}_n(k)$ ([S1], Chap. X, §4 et §5).

Une k -algèbre simple centrale A est déployée, c'est-à-dire k -isomorphe à une algèbre de matrices sur k , si et seulement si sa classe $\alpha = [A] \in \text{Br}(k)$ est nulle. Ainsi la K -algèbre $A \otimes_k K$ est déployée si et seulement si $\alpha_K = 0 \in \text{Br}(K)$.

L'indice $\text{ind}_k(A)$ d'une k -algèbre simple centrale A ne dépend que de la classe $\alpha = [A]$ de A dans $\text{Br}(k)$. Cet entier qu'on peut donc noter $\text{ind}_k(\alpha)$ est donc

(i) le plus petit degré d'une extension finie (séparable) de corps K/k telle que l'on ait $\alpha_K = 0 \in \text{Br}(K)$;

(ii) le p.g.c.d. des degrés des extensions finies de corps K/k (séparables) telles que l'on ait $\alpha_K = 0 \in \text{Br}(K)$.

On définit par ailleurs l'*exposant* d'une k -algèbre simple centrale A comme l'exposant de $\alpha = [A]$ dans le groupe de Brauer de k .

On utilise traditionnellement le mot « exposant » (ou parfois « période ») plutôt que le mot « ordre » pour éviter la confusion avec les ordres (maximaux ou autres) lorsque le corps k est le corps des fractions d'un anneau.

PROPOSITION 2.2 (R. Brauer). — (i) *L'exposant divise l'indice.*

(ii) *Les nombres premiers qui divisent l'indice divisent l'exposant.*

PREUVE — Le premier énoncé résulte de l'existence, pour une extension finie de corps K/k , d'un homomorphisme de corestriction $\text{Br}(K) \rightarrow \text{Br}(k)$ pour lequel la composition avec la restriction $\text{Br}(k) \rightarrow \text{Br}(K)$ est la multiplication par le degré $[K : k]$.

Montrons le second énoncé. Soit l premier ne divisant pas l'exposant de A . Soit K/k une extension finie galoisienne déployant A . Soit $F \subset K$ le corps fixe d'un l -sous-groupe de Sylow de $\text{Gal}(K/k)$. L'exposant de $[A \otimes_k F] \in \text{Br}(F)$ divise celui de $[A] \in \text{Br}(k)$, il est donc premier à l . Par ailleurs la restriction de $A \otimes_k F$ à K est triviale, l'argument de corestriction montre que la classe $[A \otimes_k F] \in \text{Br}(F)$ est annulée par $[K : F]$ qui est une puissance de l . Ainsi $[A \otimes_k F] = 0 \in \text{Br}(F)$, la k -algèbre A est déployée par l'extension F/k qui est de degré premier à l . Ainsi l ne divise pas l'indice. \square

2.3. Algèbres d'Azumaya, cohomologie étale, ramification ([Gr], [Mi])

Nous nous contenterons ici de rappeler quelques résultats. Pour les démonstrations, on renvoie aux trois exposés de Grothendieck [Gr] et au chapitre IV du livre de Milne [Mi].

Soit X un schéma. Une algèbre d'Azumaya sur X de degré n est un faisceau de \mathcal{O}_X -algèbres localement libres de rang fini qui localement pour la topologie étale sur X est isomorphe à $M_n(\mathcal{O}_X)$.

Soit A une telle algèbre. À tout entier m avec $1 \leq m \leq n$ on associe un X -schéma $SB(A, m)$. C'est le schéma des \mathcal{O}_X -idéaux à droite de A qui sont localement libres de rang mn sur X et qui sont localement facteurs directs dans A . C'est un X -schéma projectif, lisse, à fibres connexes. Pour $m = 1$, c'est le schéma de Severi-Brauer ([Gr]) associé à A .

On notera $Az_n(X)$ l'ensemble des classes d'isomorphie d'algèbres d'Azumaya sur X de degré n . Le produit tensoriel de telles \mathcal{O}_X -algèbres induit un produit

$$Az_n(X) \times Az_m(X) \rightarrow Az_{n+m}(X).$$

Si l'on considère l'application induite sur les classes d'isomorphie, et que de plus on considère comme triviales les algèbres de la forme $\mathcal{E}nd(V)$ pour V un fibré vectoriel sur X , on obtient un groupe abélien $\text{Br}_{Az}(X)$ (ce groupe est noté $\text{Br}(X)$ dans [Gr]). Pour X quasi-compact, ce groupe est de torsion.

On dispose par ailleurs du groupe de Brauer défini par Grothendieck. C'est le deuxième groupe de cohomologie étale $\text{Br}(X) = H^2(X, \mathbb{G}_m)$ (ce groupe est noté $\text{Br}'(X)$ dans [Gr]).

Si le schéma X est régulier, ou plus généralement si tout ouvert étale de X a ses anneaux locaux factoriels, le groupe $\text{Br}(X)$ est de torsion : ceci résulte du théorème 2.4 ci-dessous.

Pour tout schéma X , il y a un plongement naturel

$$\text{Br}_{\text{Az}}(X) \hookrightarrow \text{Br}(X).$$

Un théorème de Gabber, dont de Jong [dJ2] a donné une démonstration, assure que si X est quasi-compact et quasi-séparé, et possède un fibré inversible ample, alors le plongement ci-dessus induit un isomorphisme entre $\text{Br}_{\text{Az}}(X)$ et le sous-groupe de torsion de $\text{Br}(X)$.

Soit R un anneau de valuation discrète de corps des fractions K , de corps résiduel κ . Soit n un entier inversible sur R . On dispose alors d'une suite exacte naturelle

$$0 \rightarrow {}_n\text{Br}(R) \rightarrow {}_n\text{Br}(K) \rightarrow H^1(\kappa, \mathbf{Z}/n) \rightarrow 0.$$

(Pour un groupe abélien M et un entier $n > 0$, on note ${}_nM = \{x \in M, nx = 0\}$.)

On note ∂_R l'homomorphisme ${}_n\text{Br}(K) \rightarrow H^1(\kappa, \mathbf{Z}/n)$, qu'on appelle l'application résidu. Un élément de $\alpha \in {}_n\text{Br}(K)$ est dit *ramifié* (par rapport à l'anneau de valuation discrète R) si $\partial_R(\alpha) \neq 0$.

Soit $R \subset S$ une inclusion locale d'anneaux de valuation discrète induisant une extension finie $K \subset L$ des corps des fractions et une extension finie $\kappa_R \subset \kappa_S$ des corps résiduels. Soit $e = e_{S/R}$ l'indice de ramification de S sur R . Soit $n > 0$ premier aux caractéristiques résiduelles. On a alors le diagramme commutatif :

$$\begin{array}{ccc} {}_n\text{Br}(K) & \xrightarrow{\partial_R} & H^1(\kappa_R, \mathbf{Z}/n) \\ \text{Res}_{K,L} \downarrow & & \downarrow \times e_{S/R} \cdot \text{Res}_{\kappa_R, \kappa_S} \\ {}_n\text{Br}(L) & \xrightarrow{\partial_S} & H^1(\kappa_S, \mathbf{Z}/n). \end{array}$$

Soit X un schéma noethérien intègre. Soit K le corps des fonctions rationnelles de X . On note $X^{(i)}$ l'ensemble des points de codimension i de X . On note κ_x le corps résiduel d'un point $x \in X$. On note μ_l le faisceau étale des racines l -ièmes de 1, et $\mu_l^{\otimes 2} = \mu_l \otimes \mu_l$.

PROPOSITION 2.3. — *Soit X un schéma noethérien intègre, excellent, de dimension 2. Soit l un nombre premier inversible sur X .*

(i) *Il existe un complexe naturel de groupes de cohomologie étale*

$$0 \rightarrow H^2(X, \mu_l^{\otimes 2}) \rightarrow H^2(K, \mu_l^{\otimes 2}) \rightarrow \bigoplus_{x \in X^{(1)}} H^1(\kappa_x, \mu_l) \rightarrow \bigoplus_{x \in X^{(2)}} H^0(\kappa_x, \mathbf{Z}/l) \rightarrow 0.$$

(ii) *Si X est régulier, ce complexe est exact en le terme $H^2(K, \mu_l^{\otimes 2})$.*

(iii) *Si X est un schéma local régulier, le complexe est aussi exact en $H^2(X, \mu_l^{\otimes 2})$.*

QUELQUES RÉFÉRENCES — Pour une description du complexe, et en particulier des flèches dans le complexe, je renvoie à l'article de Kato [Kt] (qui décrit de tels complexes dans un cadre bien plus large que celui utile pour le présent exposé). Pour établir les autres énoncés on peut adjoindre les racines l -ièmes de 1, on est alors ramené à deux théorèmes sur le groupe de Brauer, dus à Auslander-Goldman et Grothendieck. Pour X local, soit $X = \text{Spec}(R)$, on a $H^2(R, \mu_l) = {}_l H^2(R, \mathbb{G}_m)$. L'énoncé (ii) est alors une conséquence du théorème 2.5 ci-après, et l'énoncé (iii) une conséquence du théorème 2.4 ci-après. (Dans le cas local régulier, on conjecture que tout le complexe est exact, ce n'est pas utile pour notre propos.)

THÉORÈME 2.4 ([AG], [Gr]). — *Soit X un schéma noethérien intègre, et soit K son corps des fonctions. Si X est géométriquement localement factoriel, par exemple si X est régulier, l'homomorphisme $\text{Br}(X) \rightarrow \text{Br}(K)$ est injectif.* \square

THÉORÈME 2.5. — *Soit X un schéma régulier intègre de dimension au plus 2, et soit K son corps des fonctions. Tout élément de $\text{Az}_n(K)$ dont l'image dans $\text{Br}(K)$ est dans $\text{Br}(X) \subset \text{Br}(K)$ est dans l'image de l'application de restriction $\text{Az}_n(X) \rightarrow \text{Az}_n(K)$.*

Soit $n > 0$ un entier inversible sur X . Si un élément de $\text{Az}_n(K)$ a une image dans ${}_n \text{Br}(K)$ dont les résidus en tous les points de codimension 1 de X sont nuls, alors il provient d'un élément de $\text{Az}_n(X)$.

Ce théorème combine un énoncé sur les anneaux de valuation discrète ([R]) et le théorème que tout module réflexif sur un anneau local régulier de dimension 2 est libre. On renvoie le lecteur à [AG], [Gr] (II, Thm. 2.1), [CTS] Cor. 6.14), [OP] (§1). \square

Soient X et K comme dans la proposition 2.3. À tout élément $\alpha \in H^2(K, \mu_l^{\otimes 2})$ on associe son lieu de ramification : c'est l'adhérence dans X de l'ensemble des points x de codimension 1 de X en lesquels α a un résidu non nul dans $H^1(\kappa_x, \mu_l)$.

À tout couple d'éléments $u, v \in K^*$ on associe le symbole $(u, v) \in H^2(K, \mu_l^{\otimes 2})$: c'est le cup-produit des classes de u et v dans $K^*/K^{*l} = H^1(K, \mu_l)$. Lorsque les racines l -ièmes de 1 sont dans K , après avoir fait un choix d'une racine primitive, on trouve ainsi les classes des algèbres cycliques dans $H^2(K, \mu_l) = {}_l \text{Br}(K)$.

PROPOSITION 2.6 (Saltman [Sa1], [CTOP]). — *Soit R un anneau local régulier excellent de dimension 2, d'idéal maximal m , de corps des fractions K . Soit $\alpha \in H^2(K, \mu_l^{\otimes 2})$ un élément dont le lieu de ramification sur $X = \text{Spec}(R)$ est un diviseur strictement à croisements normaux.*

(i) *Si ce lieu est vide, alors α est dans l'image de $H^2(R, \mu_l^{\otimes 2})$.*

(ii) *Si ce lieu est de la forme $s = 0$, où $s \in m$ est un paramètre régulier de R (i.e. R/s régulier), alors il existe $u \in R^*$ tel que $\alpha - (u, s)$ soit dans l'image de $H^2(R, \mu_l^{\otimes 2})$.*

(iii) *Si ce lieu est de la forme $st = 0$, où s et t engendrent l'idéal m , alors il existe $u, v \in R^*$ et $r \in \mathbf{Z}/l$ tels que $\alpha - (u, s) - (v, t) - r(s, t) \in H^2(K, \mu_l^{\otimes 2})$ soit dans l'image de $H^2(R, \mu_l^{\otimes 2})$.*

PREUVE — Le (i) résulte de la proposition 2.3 (ii).

Notons κ_s le corps des fractions de R/s . Le seul résidu non trivial de α est en $s = 0$, c'est une classe $\partial_s(\alpha) = \xi_s \in \kappa_s^*/\kappa_s^{*l}$ dont l'image dans \mathbf{Z}/l par l'application valuation est zéro (2.3 (i)). On a la suite exacte évidente

$$1 \rightarrow (R/s)^*/(R/s)^{*l} \rightarrow \kappa_s^*/\kappa_s^{*l} \rightarrow \mathbf{Z}/l \rightarrow 0.$$

Il existe donc un élément $u_s \in (R/s)^*$ qui a pour image ξ_s via l'application de réduction $(R/s)^*/(R/s)^{*l} \rightarrow \kappa_s^*/\kappa_s^{*l}$. L'application de réduction $R^* \rightarrow (R/s)^*$ est surjective (R est un anneau local), on peut donc relever u_s en $u \in R^*$. Les formules usuelles pour le résidu d'un cup-produit montrent que le résidu de (u, s) en $s = 0$ est ξ_s , et il est zéro partout ailleurs, comme celui de α . La proposition 2.3 (ii) donne alors l'énoncé (ii).

Montrons (iii). Notons κ_s , resp. κ_t , le corps des fractions de R/s , resp. R/t , et κ le corps résiduel en l'idéal maximal m de R . Notons $\partial_s : H^1(\kappa_s, \mu_l) \rightarrow H^0(\kappa, \mathbf{Z}/l)$ et de même $\partial_t : H^1(\kappa_t, \mu_l) \rightarrow H^0(\kappa, \mathbf{Z}/l)$ les homomorphismes apparaissant dans la proposition 2.3. Soient ξ_s et ξ_t les résidus de α en $s = 0$ et $t = 0$. Comme la proposition 2.3 donne un complexe de groupes abéliens, on a $\partial_s(\xi_s) + \partial_t(\xi_t) = 0 \in \mathbf{Z}/l$. Soit $r \in \mathbf{Z}$ avec $\partial_s(\xi_s) = r \in \mathbf{Z}/l$ et $\partial_t(\xi_t) = -r \in \mathbf{Z}/l$. Procédant comme ci-dessus, on trouve $u, v \in R^*$ tels que l'image de $ut^r \in R[1/t]^*$ dans κ_s^*/κ_s^{*l} soit ξ_s et que l'image de $vs^{-r} \in R[1/s]^*$ dans κ_t^*/κ_t^{*l} soit ξ_t . On vérifie que tous les résidus de

$$\alpha + (s, u) + (t, v) + r(s, t) \in H^2(K, \mu_l^{\otimes 2})$$

aux points de codimension 1 de X sont nuls, la proposition 2.3 (ii) donne alors que cet élément de $H^2(K, \mu_l^{\otimes 2})$ est l'image d'un (unique) élément de $H^2(R, \mu_l^{\otimes 2})$. \square

2.4. Questions sur les corps de déploiement

Sauf mention explicite du contraire, on suppose ici que les corps considérés sont de caractéristique zéro, ou du moins que l'indice des algèbres considérées est premier à la caractéristique du corps de base. Etant donné un corps k et une k -algèbre simple centrale A , que peut-on dire sur les extensions finies de corps K/k qui déploient A ?

Il existe des k -algèbres simples centrales qui ne sont pas déployées par une extension cyclique du corps de base (Tignol-Amitsur, Tignol-Wadsworth [TW] Exemples 3.6 et Thm. 4.7 (v)). Un exemple simple est le produit tensoriel $(x, y) \otimes_k (z, t)$ des algèbres de quaternions (x, y) et (z, t) sur le corps $k = \mathbf{C}(x, y, z, t)$ et même déjà sur le corps de séries formelles itérées $k = \mathbf{C}((x))((y))((z))((t))$.

Toute k -algèbre simple centrale A d'exposant n sur un corps k contenant une racine primitive n -ième de 1, soit ζ_n , est semblable à un produit d'algèbres cycliques $(a, b)_{\zeta_n}$, elle est en particulier déployée par une extension multicyclique de k (conséquence immédiate du théorème de Merkur'ev et Suslin [MS]). Il en résulte que toute k -algèbre simple centrale A est déployée par une extension résoluble de k .

Étant donné un corps K et une k -algèbre simple centrale à division D , on dit qu'elle est un produit croisé si elle admet un sous-corps commutatif maximal $K \subset D$ galoisien sur k . On dit qu'elle est cyclique si elle admet un sous-corps commutatif maximal $K \subset D$ cyclique sur k .

Amitsur (1972) montra qu'il existe des algèbres à division qui ne sont pas des produits croisés : elles ne possèdent pas de sous-corps commutatif maximal galoisien sur le corps de base. Les indices de ses exemples ainsi que de ceux qui ont été construits par la suite sont de la forme $2^r \prod_i p_i^{n_i}$ avec les $p_i \geq 3$ premiers, $r \geq 3$, et $n_i \geq 2$ pour tout i . Lorsque l'indice est premier, la question suivante est ouverte :

Problème *Sur tout corps k , toute algèbre à division d'indice premier l est-elle cyclique ?*

C'est le cas de façon triviale pour $l = 2$, c'est un résultat de Wedderburn pour $l = 3$, la question est ouverte déjà pour $l = 5$.

Dès les années 1930, des exemples furent donnés (par Albert, Brauer, Köthe, Nakayama, voir les références dans [CTG]) pour montrer qu'indice et exposant d'une algèbre ne coïncident pas nécessairement.

Un énoncé général permettant de fabriquer de tels exemples est le suivant :

PROPOSITION 2.7 (Tignol [T]). — *Soient K/k une extension cyclique de corps, σ un générateur du groupe de Galois de K sur k et t une variable. Soit A une k -algèbre simple centrale. On a les formules*

$$\text{ind}_{k(t)}(A_{k(t)} \otimes_{k(t)} (K(t)/k(t), \sigma, t)) = \text{ind}(A_K) \cdot [K : k]$$

et

$$\text{ind}_{k((t))}(A_{k((t))} \otimes_{k((t))} (K((t))/k((t)), \sigma, t)) = \text{ind}(A_K) \cdot [K : k].$$

En particulier, si A est un corps gauche, les conditions suivantes sont équivalentes :

- (i) *La K -algèbre A_K est un corps gauche.*
- (ii) *La $k(t)$ -algèbre $A_{k(t)} \otimes_{k(t)} (K(t)/k(t), \sigma, t)$ est un corps gauche.*
- (iii) *La $k((t))$ -algèbre $A_{k((t))} \otimes_{k((t))} (K((t))/k((t)), \sigma, t)$ est un corps gauche. □*

Donnons deux exemples. Sur le corps $K = \mathbf{C}(x_1, \dots, x_d)$ des fonctions rationnelles en d variables sur le corps des complexes, le produit tensoriel d'algèbres de quaternions $A = (x_1 + 2, x_2) \otimes_K \cdots \otimes_K (x_1 + d, x_d)$ est un corps gauche, d'exposant 2 et d'indice 2^{d-1} . Il en est ainsi déjà sur le corps $K = \mathbf{C}(x_1)((x_2)) \dots ((x_d))$. Par ailleurs soit \mathbf{F} un corps fini de caractéristique différente de 2, et soit $a \in \mathbf{F}$ non carré. Sur le corps de fonctions rationnelles $K = \mathbf{F}(x, y)$, l'algèbre $(x, a) \otimes_K (x + 1, y)$ est un corps gauche, d'exposant 2 et d'indice 4. Ces exemples font intervenir la ramification des algèbres considérées en des valuations convenables du corps de base.

Dans [Kr] et [CTG] on trouvera diverses constructions d'algèbres d'Azumaya A (donc sans ramification) sur des variétés projectives, lisses, connexes X sur les complexes (de dimension au moins 3) telles que l'exposant divise proprement l'indice de la $\mathbf{C}(X)$ -algèbre obtenue par évaluation de A au point générique de X . Pour tout $d \geq 2$, et tout nombre premier l , O. Gabber donne des exemples de variétés projectives, lisses, connexes X de dimension d , en fait des produits de courbes, qui possèdent une algèbre d'Azumaya d'exposant l et d'indice (générique) l^{d-1} .

Les exemples obtenus de cette façon sont définis sur des corps dont la dimension cohomologique [S2] est au moins 3. Par d'autres méthodes, pour tout entier n , Merkur'ev [M2] construit un (très gros) corps K_n de dimension cohomologique 2 sur lequel il existe un produit tensoriel de n algèbres de quaternions qui est un corps gauche, et donc définit un élément de $\text{Br}(K_n)$ d'exposant 2 et d'indice 2^n .

Il existe néanmoins plusieurs classes de corps de dimension cohomologique 2 ([S2]) (ou de dimension cohomologique virtuelle 2) pour lesquels on sait que l'indice coïncide avec l'exposant. Pour chacune des classes suivantes, on sait en outre que toute algèbre à division est cyclique :

- les corps globaux et les corps locaux (Brauer-Hasse-Noether, Albert) ([BHN], [D], [Ro])
- les corps de la forme $k = F((t))$ avec F corps de caractéristique zéro et de dimension cohomologique 1
- les corps de fractions d'anneaux locaux normaux henséliens excellents de dimension 2, à corps résiduel algébriquement clos de caractéristique zéro (Artin [Ar3], Ford-Saltman [FS], [CTOP]).

Soit $r \geq 0$ un entier. Un corps k satisfait la propriété C_r si pour toute extension finie de corps K/k , toute forme homogène à coefficients dans K de degré d en $n > d^r$ variables a un zéro non trivial dans K . Les corps de fonctions de r variables sur un corps algébriquement clos sont des corps C_r (S. Lang). C'est une conséquence du théorème de Merkur'ev et Suslin [MS] que tout corps C_2 est de dimension cohomologique au plus 2.

Problème (M. Artin [Ar2], Appendix) *Soit k un corps C_2 . Pour toute algèbre simple centrale sur k , l'indice est-il égal à l'exposant ?*

Lorsque l'on se limite aux algèbres d'indice 2-primaire ou 3-primaire, la réponse est affirmative¹. Cette conséquence du théorème de Merkur'ev et Suslin [MS] a été remarquée par plusieurs auteurs ([Ar2], [MS]). Le cas particulier des corps de fonctions de deux variables sur les complexes avait été établi plus tôt, par M. Artin et J. Tate ([Ar2], Appendix), qui utilisaient un résultat de S. Bloch (1974).

¹Quitte à remplacer l'hypothèse C_2 par sa variante C'_2 , cf. [CTG], Prop.7.)

Pour les corps de fonctions de deux variables sur les complexes, et les algèbres d'indice arbitraire sur de tels corps, la réponse affirmative à cette question est le théorème de de Jong [dJ] qui fait l'objet de cet exposé. Pour $r \leq 2$, on a donc une réponse affirmative à la question suivante (voir [CT]), où la borne suggérée est, d'après les exemples mentionnés ci-dessus, la meilleure possible.

Problème² *Pour toute algèbre simple centrale sur un corps de fonctions de r variables sur le corps des complexes, l'indice divise-t-il l'exposant à la puissance $r - 1$?*

Les corps de fonctions de r variables sur le corps des complexes sont des corps C_r . On peut se poser la question ci-dessus pour tout corps C_r . Les corps de fonctions de r variables sur un corps fini sont des corps C_{r+1} .

Problème *Pour toute algèbre simple centrale sur un corps de fonctions de r variables sur un corps fini, l'indice divise-t-il l'exposant à la puissance r ?*

Lorsque $r = 1$, c'est un résultat classique (Hasse).

Lorsque $r = 2$, c'est un résultat récent de Lieblich [Lie] que pour les algèbres non ramifiées, d'indice premier à la caractéristique, l'indice est égal à l'exposant³. Pour une algèbre ramifiée, la technique évoquée au début du §4 ci-dessus (méthode de Saltman) permet alors de montrer que l'indice divise l'exposant au cube. Comme on a vu ci-dessus, dans ce cas le mieux que l'on puisse espérer en général est que l'indice divise l'exposant au carré.

3. LE THÉORÈME DE DE JONG

3.1. Deux invariants cohomologiques des algèbres d'Azumaya de degré n

Soient X un schéma et $n > 0$ un entier. L'ensemble $Az_n(X)$ des classes d'isomorphie d'algèbres d'Azumaya de degré n sur X s'identifie à l'ensemble de cohomologie de Čech étale $H^1(X, PGL_n)$.

On a le diagramme commutatif de suites exactes de X -schémas en groupes suivant :

²Soit X une variété de dimension r sur le corps \mathbf{R} des réels, géométriquement intègre et sans point réel. Le corps des fonctions $K = \mathbf{R}(X)$ est de dimension cohomologique r mais on ne sait pas s'il est C_r . On peut pour un tel corps poser le même problème que ci-dessus. Pour les surfaces, i.e. pour $r = 2$, le problème se ramène à l'une quelconque des questions suivantes (Pfister, 1982) :

(a) Toute forme quadratique en au moins 5 variables sur un tel corps admet-elle un zéro non trivial ?
 (b) Le produit tensoriel de deux algèbres de quaternions sur un tel corps est-il semblable à une algèbre de quaternions ?

³Pour obtenir ce résultat, outre des techniques de champs algébriques il utilise les propriétés des espaces de modules de fibrés vectoriels sur les surfaces projectives et lisses [HL].

$$\begin{array}{ccccccc}
& & 1 & & 1 & & \\
& & \downarrow & & \downarrow & & \\
1 & \longrightarrow & \mu_n & \longrightarrow & SL_n & \longrightarrow & PGL_n \longrightarrow 1 \\
& & \downarrow & & \downarrow & & \downarrow = \\
1 & \longrightarrow & \mathbb{G}_m & \longrightarrow & GL_n & \longrightarrow & PGL_n \longrightarrow 1 \\
& & \downarrow x \mapsto x^n & & \downarrow \det & & \\
& & \mathbb{G}_m & \xrightarrow{=} & \mathbb{G}_m & & \\
& & \downarrow & & \downarrow & & \\
& & 1 & & 1 & & .
\end{array}$$

Supposons n inversible sur X . Les suites exactes ci-dessus définissent alors des suites exactes de faisceaux pour la topologie étale sur X . De la suite horizontale médiane on tire la suite exacte d'ensembles pointés de cohomologie de Čech étale ([Mi], Chapitre IV, Thm. 2.5) :

$$H^1(X, \mathbb{G}_m) \rightarrow H^1(X, GL_n) \rightarrow H^1(X, PGL_n) \rightarrow {}_n H^2(X, \mathbb{G}_m). \quad (3.1.1)$$

L'application composée $Az_n(X) = H^1(X, PGL_n) \rightarrow H^2(X, \mathbb{G}_m) = \text{Br}(X)$ associe à une algèbre d'Azumaya A sa classe $[A]$ dans le groupe de Brauer cohomologique $\text{Br}(X)$, classe qui est annulée par n .

La suite exacte supérieure donne naissance à une application

$$\text{cl} : Az_n(X) = H^1(X, PGL_n) \rightarrow H^2(X, \mu_n).$$

La suite exacte verticale de gauche (suite de Kummer) donne naissance à la suite exacte bien connue

$$0 \rightarrow \text{Pic}(X)/n \rightarrow H^2(X, \mu_n) \rightarrow {}_n \text{Br}(X) \rightarrow 0.$$

Pour $A \in Az_n(X)$, la flèche de droite envoie $\text{cl}(A) \in H^2(X, \mu_n)$ sur $[A] \in {}_n \text{Br}(X)$.

LEMME 3.1. — *Soit X un schéma.*

(a) *Soit $A \in Az_n(X)$. Si l'on a $[A] = 0 \in \text{Br}(X)$, alors il existe un fibré vectoriel V sur X de rang n tel que $A = \mathcal{E}nd_X(V)$.*

(b) *Soient V_1 et V_2 deux fibrés vectoriels de rang n sur X ; si l'on a un isomorphisme d'algèbres $\mathcal{E}nd(V_1) \simeq \mathcal{E}nd(V_2)$, alors il existe un fibré inversible \mathcal{L} sur X et un isomorphisme $V_1 \simeq V_2 \otimes \mathcal{L}$.*

(c) *Soit V un fibré vectoriel sur X de rang n , $A = \mathcal{E}nd(V) \in Az_n(X)$. Supposons n inversible sur X . L'image de la classe de $\det(V) = \Lambda^n V$ dans $\text{Pic}(X)$ par la flèche de Kummer $\text{Pic}(X)/n \rightarrow H^2(X, \mu_n)$ coïncide avec l'opposé de $\text{cl}(A) \in H^2(X, \mu_n)$. \square*

PREUVE — Les points (a) et (b) sont des conséquences de la suite exacte (3.1.1). Le point (c) s'établit en considérant les suites exactes de cohomologie de Čech étales déduites du diagramme commutatif de suites exactes de X -schémas en groupes lisses :

$$\begin{array}{ccccccc}
 1 & \longrightarrow & \mu_n & \longrightarrow & SL_n & \longrightarrow & PGL_n \longrightarrow 1 \\
 & & \uparrow x \mapsto x & & \uparrow \text{pr}_1 & & \uparrow \\
 1 & \longrightarrow & \mu_n & \xrightarrow{x \mapsto (x, x^{-1})} & SL_n \times \mathbb{G}_m & \xrightarrow{(u, v) \mapsto uv} & GL_n \longrightarrow 1 \\
 & & \downarrow x \mapsto x^{-1} & & \downarrow \text{pr}_2 & & \downarrow \det \\
 1 & \longrightarrow & \mu_n & \longrightarrow & \mathbb{G}_m & \xrightarrow{x \mapsto x^n} & \mathbb{G}_m \longrightarrow 1.
 \end{array}$$

3.2. Transformations élémentaires d'algèbres d'Azumaya sur une surface

Soient X un schéma, $A \in \text{Az}_n(X)$ et $D \subset X$ un diviseur de Cartier effectif, $I_D \subset \mathcal{O}_X$ l'idéal inversible le définissant. Supposons que la restriction A_D de A à D s'écrive $\mathcal{E}nd(\overline{V})$, avec \overline{V} un fibré vectoriel sur D , et que l'on dispose d'un sous-fibré vectoriel $\overline{F} \subset \overline{V}$ sur D localement facteur direct. On construit alors une autre algèbre d'Azumaya $A' \in \text{Az}_n(X)$, appelée transformée élémentaire de A par rapport à $\overline{F} \subset \overline{V}$, de la façon suivante.

On considère la sous-algèbre B des sections de A qui préservent la filtration $\overline{F} \subset \overline{V}$. Notons $i : D \hookrightarrow X$ l'immersion fermée naturelle. On a la suite exacte de \mathcal{O}_X -modules cohérents

$$0 \rightarrow B \rightarrow A \rightarrow i_* \mathcal{H}om_D(\overline{F}, \overline{Q}) \rightarrow 0, \tag{3.2.1}$$

où $\overline{Q} = \overline{V}/\overline{F}$. Notons que B contient $\mathcal{O}_X \subset A$.

Pour tout point x de X , il existe un schéma affine U et un morphisme étale $U \rightarrow X$ d'image contenant x tel qu'après restriction à U on puisse écrire $A = \mathcal{E}nd(V)$ avec $V_D \simeq \overline{V}$ et que de plus il existe un scindage $V = F \oplus Q$ avec $F_D = \overline{F}$. Sur U , A_U se lit

$$\begin{pmatrix} \mathcal{E}nd(F) & \mathcal{H}om(Q, F) \\ \mathcal{H}om(F, Q) & \mathcal{E}nd(Q) \end{pmatrix}$$

On définit une algèbre d'Azumaya A'_U sur U par

$$\begin{pmatrix} \mathcal{E}nd(F) & I_D^{-1} \otimes \mathcal{H}om(Q, F) \\ I_D \otimes \mathcal{H}om(F, Q) & \mathcal{E}nd(Q) \end{pmatrix}$$

On vérifie que les algèbres A'_U pour divers U se recollent et définissent une algèbre d'Azumaya $A' \in \text{Az}_n(X)$, que l'on appelle la transformée élémentaire de A le long de $\overline{F} \subset \overline{V}$. On vérifie sur la description locale ci-dessus que l'on a la suite exacte de \mathcal{O}_X -modules cohérents

$$0 \rightarrow B \rightarrow A' \rightarrow i_* \mathcal{H}om_D((I_D/I_D^2) \otimes \overline{Q}, \overline{F}) \rightarrow 0. \tag{3.2.2}$$

Ici encore, l'inclusion $B \subset A'$ induit l'identité de $\mathcal{O}_X \subset B$ vers $\mathcal{O}_X \subset A'$.

(C'est une variante de constructions que l'on trouve dans d'autres contextes : transformations élémentaires entre ordres maximaux d'une algèbre simple centrale sur le corps des fractions d'un anneau de valuation discrète $[R]$, transformations élémentaires sur les fibrés vectoriels ([HL]).)

LEMME 3.2. — Soient X un schéma connexe, D un diviseur de Cartier effectif sur X , et $i : D \subset X$ l'inclusion naturelle. Supposons donnée une suite exacte de \mathcal{O}_X -modules cohérents

$$0 \rightarrow V' \rightarrow V \rightarrow i_*Q \rightarrow 0$$

où V' et V sont des fibrés vectoriels sur X et Q est un fibré vectoriel de rang s sur D . L'application déterminant : $\det(V') \rightarrow \det(V)$ identifie le fibré inversible $\det(V')$ avec le sous-fibré inversible $\det(V) \otimes \mathcal{O}_X(-sD) \subset \det(V)$.

PREUVE — On dispose de l'inclusion naturelle de fibrés inversibles $\det(V') \rightarrow \det(V)$. Pour établir l'énoncé, on peut supposer le schéma X local, soit $X = \text{Spec}(R)$ et D défini par un élément non inversible $\pi \in R$. On a une surjection de R/π -modules libres $V/\pi \rightarrow Q$. En scindant celle-ci on trouve une base $\bar{e}_1, \dots, \bar{e}_n$ du R -module libre V/π telle que $\bar{e}_{s+1}, \dots, \bar{e}_n$ soit une base du noyau. Soient (e_1, \dots, e_n) des relevés de $\bar{e}_1, \dots, \bar{e}_n$ dans V . Alors (e_1, \dots, e_n) est une base du R -module libre V et $(\pi e_1, \dots, \pi e_s, e_{s+1}, \dots, e_n)$ une base du R -module libre V' . L'application $\Lambda^n V' \rightarrow \Lambda^n V$ envoie le générateur évident de $\Lambda^n V'$ sur $\pi^s \cdot (e_1 \wedge \dots \wedge e_n)$ qui est une base de $\det(V) \otimes \mathcal{O}_X(-sD)$. \square

PROPOSITION 3.3. — Soient X un schéma connexe, $A \in \text{Az}_n(X)$, avec n inversible sur X , et $D \subset X$ un diviseur de Cartier effectif. Supposons qu'il existe un fibré vectoriel \bar{V} sur D tel que $A_D = \mathcal{E}nd(\bar{V})$, et supposons donné un sous-fibré vectoriel $\bar{F} \subset \bar{V}$ sur D localement facteur direct, de rang constant r . Soit A' le transformé élémentaire de A le long de $\bar{F} \subset \bar{V}$.

On a alors

$$\text{cl}(A') = \text{cl}(A) - r.[D] \in H^2(X, \mu_n),$$

où $[D]$ désigne l'image de D par l'application composée $\text{Div}(X) \rightarrow \text{Pic}(X)/n \hookrightarrow H^2(X, \mu_n)$.

PREUVE (P. Gille) — Soit $p : Y \rightarrow X$ le schéma de Severi-Brauer associé à A . Soit $i : D_Y = D \times_X Y \rightarrow Y$ l'immersion naturelle. Sur Y , il y a une suite exacte

$$0 \rightarrow V' \rightarrow V \rightarrow i_*\bar{Q} \rightarrow 0,$$

avec V' et V fibrés vectoriels sur Y et \bar{Q} le fibré vectoriel de rang $n - r$ sur D_Y image réciproque de \bar{V}/\bar{F} par p . On a $A_Y = \mathcal{E}nd(V)$ et $A'_Y = \mathcal{E}nd(V')$. On a donc (Lemme 3.1 (c)) $\text{cl}(A_Y) = -\det(V) \in H^2(Y, \mu_n)$ et $\text{cl}(A'_Y) = -\det(V') \in H^2(Y, \mu_n)$, où l'on utilise tacitement l'inclusion $\text{Pic}(Y)/n \hookrightarrow H^2(Y, \mu_n)$. D'après le lemme 3.2 on a donc $\text{cl}(A'_Y) = \text{cl}(A_Y) + (n - r).[D]_Y = \text{cl}(A_Y) - r.[D]_Y \in H^2(Y, \mu_n)$. En analysant la suite spectrale de Leray pour le morphisme $p : Y \rightarrow X$ on montre que la flèche naturelle $H^2(X, \mu_n) \rightarrow H^2(Y, \mu_n)$ est injective. Ainsi $\text{cl}(A') = \text{cl}(A) - r.[D] \in H^2(X, \mu_n)$. \square

On peut aussi établir cette formule par un calcul local via des cocycles de Čech ([dJ]).

LEMME 3.4. — *Soit k un corps infini. Soient D/k une courbe projective lisse géométriquement connexe et V un fibré vectoriel de rang n sur D . Supposons donné pour chaque point t dans un ensemble fini T de points de $D(k)$ un sous-espace vectoriel $F_t \subset V_t$ de dimension 1. Pour tout $m > 0$ suffisamment grand il existe un homomorphisme injectif $\mathcal{O}_D(-m) \rightarrow V$ dont l'image est localement facteur direct dans V et tel que l'image de $\mathcal{O}_D(-m)_t$ dans V_t coïncide avec F_t .*

PREUVE — Soit $W \subset V$ le sous- \mathcal{O}_D -module (localement libre) de V dont les sections après évaluation en t donnent des éléments de F_t . Pour $m \gg 0$, le fibré vectoriel $W(m)$ est engendré par ses sections. Soit E le k -vectoriel $H^0(D, W(m))$. Notons encore E l'espace affine défini par E . Soit $Z \subset D \times_k E$ le fermé dont les points géométriques sont les couples (x, e) avec $e \in E$ tel que $e_x = 0 \in W(m)_x$. Comme W est de rang au moins 2, et que $W(m)$ est engendré par ses sections globales, pour x point géométrique fixé, l'ensemble des $e \in E$ satisfaisant $e_x = 0$ est de codimension au moins 2 dans E . La codimension de Z dans $D \times_k E$ est donc au moins 2, et le fermé $Z_1 \subset E$ qui est l'adhérence de la projection de Z dans E est donc de codimension au moins 1. Choisissons un k -point de E dans le complémentaire de Z_1 . Ceci définit un homomorphisme $\mathcal{O}_X \rightarrow W(m)$ et donc un homomorphisme $\mathcal{O}_X(-m) \rightarrow W \subset V$ satisfaisant les propriétés annoncées. \square

3.3. Relèvement des algèbres

Soient X un schéma et $A \in \text{Az}_n(X)$. La trace réduite $\text{Tr}_{\text{red}} : A \rightarrow \mathcal{O}_X$ est une application \mathcal{O}_X -linéaire qui induit sur $\mathcal{O}_X \subset A$ la multiplication par n . On note $A^0 \subset A$ le noyau de la trace réduite. Lorsque n est inversible sur X , l'injection $\mathcal{O}_X \subset A$ est scindée, et le quotient A/\mathcal{O}_X est isomorphe à A^0 .

THÉORÈME 3.5. — *Soit X une surface connexe, projective et lisse sur un corps k algébriquement clos. Soit $A \in \text{Az}_n(X)$ avec $n > 1$ premier à la caractéristique de k . Il existe une algèbre d'Azumaya $A' \in \text{Az}_n(X)$, obtenue par transformation élémentaire de A , telle que $\text{cl}(A') = \text{cl}(A) \in H^2(X, \mu_n)$ et que $H^2(X, A'/\mathcal{O}_X) = 0$.*

PREUVE — Soit \mathcal{L} un faisceau inversible sur X . L'inclusion $\mathcal{O}_X \subset A$ induit une inclusion $\mathcal{L} \subset A \otimes \mathcal{L}$ et donc une inclusion $H^0(X, \mathcal{L}) \subset H^0(X, A \otimes \mathcal{L})$.

1) Pour tout élément s de $H^0(X, A \otimes \mathcal{L})$ n'appartenant pas à $H^0(X, \mathcal{L})$ il existe une infinité de $t \in X(k)$ tels que s_t ne soit pas « scalaire », et donc tels qu'après choix d'un isomorphisme $A_t \simeq \text{End}(V_t)$ il existe un sous-espace vectoriel $F_t \subset V_t$ de dimension 1 tel que $s(F_t)$ ne soit pas contenu dans $F_t \otimes \mathcal{L}_t$. Comme $H^0(X, A \otimes \mathcal{L})$ est de dimension finie, on en déduit :

Il existe un ensemble fini $T \subset X(k)$ et pour chaque $t \in T$ un k -espace vectoriel V_t de dimension n , un sous-espace vectoriel $F_t \subset V_t$ de dimension 1 et un isomorphisme

d'algèbres $A_t \simeq \text{End}(V_t)$ tels que l'inclusion $\mathcal{L} \subset A \otimes \mathcal{L}$ induise une égalité

$$H^0(X, \mathcal{L}) = \{s \in H^0(X, A \otimes \mathcal{L}), s_t(F_t) \subset F_t \otimes \mathcal{L}_t \text{ pour tout } t\}.$$

2) Soit $T \subset X(k)$ un ensemble fini de points. Une variante du théorème de Bertini montre : Il existe une courbe lisse connexe $D \subset X$ d'image nulle dans $\text{Pic}(X)/n\text{Pic}(X)$ telle que $T \subset D$. (Il suffit de prendre une section convenable du faisceau $\mathcal{O}_X(nq)$ pour $q > 0$ assez grand.)

3) Choisissons T comme en 1) et $D \subset X$ comme en 2). D'après le théorème de Tsen, *le groupe de Brauer du corps des fonctions d'une courbe définie sur un corps algébriquement clos est nul. Le théorème 2.4 donne alors* $\text{Br}(D) = 0^4$. D'après le lemme 3.1 il existe donc un fibré vectoriel \bar{V} de rang n sur D tel que $A_D \simeq \mathcal{E}nd_D(\bar{V})$. Pour $t \in T$, on a deux isomorphismes $A_t \simeq \text{End}(V_t)$ et $A_{D,t} \simeq \text{End}_D(\bar{V}_t)$. Ces deux isomorphismes sont déduits l'un de l'autre via des isomorphismes $\eta_t : V_t \simeq \bar{V}_t$. Pour tout $m > 0$ assez grand, le lemme 3.4 assure l'existence d'un homomorphisme $\varphi : \bar{F} \rightarrow \bar{V}$, avec $\bar{F} = \mathcal{O}_D(-m)$, tel qu'en tout $t \in T$ l'image de φ_t soit égale à $\eta_t(F_t)$. Soit $A' \in \text{Az}_n(X)$ le transformé élémentaire de A le long de $\varphi : \bar{F} \rightarrow \bar{V}$.

4) Le choix de D , de classe nulle dans $\text{Pic}(X)/n\text{Pic}(X)$, et la proposition 3.3 impliquent

$$\text{cl}(A') = \text{cl}(A) \in H^2(X, \mu_n).$$

5) La suite exacte (3.2.1) donne après tensorisation avec \mathcal{L} la suite exacte

$$0 \rightarrow B \otimes \mathcal{L} \rightarrow A \otimes \mathcal{L} \rightarrow i_* \mathcal{H}om_D(\bar{F}, \bar{Q}) \otimes \mathcal{L}_D \rightarrow 0 \quad (3.4.1)$$

La flèche composée

$$H^0(X, B \otimes \mathcal{L}) \rightarrow H^0(X, A \otimes \mathcal{L}) \rightarrow H^0(D, i_* \mathcal{H}om_D(\bar{F}, \bar{Q}) \otimes \mathcal{L}_D) \rightarrow \prod_{t \in T} \text{Hom}(\bar{F}_t, \bar{Q}_t) \otimes \mathcal{L}_t$$

est nulle. La première flèche est clairement injective. Le noyau de la deuxième s'identifie par construction à $H^0(X, \mathcal{L}) \subset H^0(X, A \otimes \mathcal{L})$. On en conclut $H^0(X, \mathcal{L}) = H^0(X, B \otimes \mathcal{L})$. En tensorisant la suite exacte (3.2.2) par le fibré inversible \mathcal{L} on obtient la suite exacte

$$0 \rightarrow B \otimes \mathcal{L} \rightarrow A' \otimes \mathcal{L} \rightarrow i_* \mathcal{H}om_D((I_D/I_D^2) \otimes \bar{Q}, \bar{F}) \otimes \mathcal{L}_D \rightarrow 0. \quad (3.4.2)$$

On voit maintenant que le noyau de

$$H^0(X, A' \otimes \mathcal{L}) \rightarrow H^0(D, \mathcal{H}om_D((I_D/I_D^2) \otimes \bar{Q}, \bar{F}) \otimes_D \mathcal{L}_D)$$

s'identifie à $H^0(X, \mathcal{L}) \subset H^0(X, A' \otimes \mathcal{L})$.

Notons $\mathcal{N}_{X/D}$ le faisceau normal de D dans X et E^* le dual d'un \mathcal{O}_D -faisceau localement libre E . Le \mathcal{O}_D -faisceau localement libre $\mathcal{H}om_D((I_D/I_D^2) \otimes \bar{Q}, \bar{F}) \otimes_D \mathcal{L}_D$ peut encore s'écrire

⁴Dans la démonstration du théorème de de Jong, cet argument n'est pas utilisé, car on déforme une algèbre de classe triviale dans le groupe de Brauer, voir le début de la démonstration de la proposition 3.12.

$\mathcal{N}_{X/D} \otimes \overline{Q}^* \otimes \overline{F} \otimes \mathcal{L}_D$. Il s'injecte dans le faisceau $\mathcal{N}_{X/D} \otimes \overline{V}^* \otimes \overline{F} \otimes \mathcal{L}_D$. On a donc aussi une injection

$$H^0(D, \mathcal{H}om_D((I_D/I_D^2) \otimes \overline{Q}, \overline{F}) \otimes_D \mathcal{L}_D) \hookrightarrow H^0(D, \mathcal{N}_{X/D} \otimes \overline{V}^* \otimes \mathcal{L}_D \otimes \overline{F}).$$

Observons alors qu'une fois fixés T, D, \overline{V} on peut choisir m au point 3) aussi grand que l'on veut pour définir $\varphi : \mathcal{O}_D(-m) = \overline{F} \subset \overline{V}$, et donc pour définir A' . Si l'on choisit m tel que

$$H^0(D, \mathcal{N}_{X/D} \otimes \overline{V}^* \otimes \mathcal{L}_D \otimes \mathcal{O}_D(-m)) = 0$$

et φ et donc A' avec un tel m , alors $H^0(X, \mathcal{L}) = H^0(X, A' \otimes \mathcal{L})$. Comme n est inversible sur X , ceci implique $H^0(X, A'^0 \otimes \mathcal{L}) = 0$.

6) Appliquons le résultat obtenu au faisceau inversible canonique $\mathcal{L} = \omega_{X/k} = \Lambda^2(\Omega_X^1)$. Cela donne une algèbre d'Azumaya $A' \in \text{Az}_n(X)$ qui est transformée élémentaire de A , qui satisfait $\text{cl}(A) = \text{cl}(A') \in H^2(X, \mu_n)$ et telle que $H^0(X, A'^0 \otimes \omega_{X/k}) = 0$. L'application $A' \times A' \rightarrow \mathcal{O}_X$ définie par $(x, y) \mapsto \text{Tr}_{\text{red}}(xy)$ induit une dualité parfaite entre les \mathcal{O}_X -modules A'^0 et A'/\mathcal{O}_X . Par dualité de Serre sur la surface projective et lisse X , le k -espace vectoriel $H^2(X, A'/\mathcal{O}_X)$ est dual de $H^0(X, A'^0 \otimes \omega_{X/k})$. Ainsi $H^2(X, A'/\mathcal{O}_X) = 0$. \square

Le théorème 3.5 va permettre d'utiliser les théorèmes de relèvement suivants.

PROPOSITION 3.6. — *Soit $X \rightarrow X'$ une immersion fermée de schémas définie par un idéal $I \subset \mathcal{O}_{X'}$ tel que $I^2 = 0$. L'idéal I est un \mathcal{O}_X -module. Soit A une algèbre d'Azumaya sur X . Si l'on a $H^2(X, (A/\mathcal{O}_X) \otimes I) = 0$, alors il existe une algèbre d'Azumaya A' sur X' induisant A sur X .* \square

C'est un cas particulier d'un résultat général de Giraud ([Gi], Chap. VII, Théorème 1.3.1 et Remarque 1.3.1.2). \square

THÉORÈME 3.7. — *Soit R un anneau local noethérien complet de corps résiduel k . Soit $X/\text{Spec}(R)$ un schéma propre et plat. Soit X_0/k la fibre spéciale. Si $A_0 \in \text{Az}_n(X_0)$ satisfait $H^2(X_0, A_0/\mathcal{O}_{X_0}) = 0$, alors il existe une algèbre d'Azumaya $A \in \text{Az}_n(X)$ qui induit A_0 sur X_0 .*

Cela résulte de la proposition 3.6 et du théorème d'algébrisation des modules formels de Grothendieck (FGA, SGA 1, [EGA] III.5 Théorème 5.1.4, [Gr], III, §3, [I]). \square

THÉORÈME 3.8. — *Soit R un anneau local hensélien de corps résiduel k . Soit $X/\text{Spec}(R)$ un schéma propre et plat. Soient X_0/k la fibre spéciale et $A_0 \in \text{Az}_n(X_0)$. Si l'on a $H^2(X_0, A_0/\mathcal{O}_{X_0}) = 0$, alors il existe une algèbre d'Azumaya $A \in \text{Az}_n(X)$ qui induit A_0 sur X_0 .*

C'est une conséquence formelle du théorème précédent et du théorème d'approximation d'Artin [Ar1]. Pour une situation similaire et quelques détails de plus, voir [CTOP],

Theorem 1.8. On commence par se réduire au cas où R est l'hensélisée d'une \mathbf{Z} -algèbre de type fini en un idéal premier. Le foncteur $S \rightarrow H^1(X \times_R S, PGL_n)$ de la catégorie des R -algèbres commutatives vers la catégorie des ensembles est de présentation finie. Le théorème précédent donne une classe $\hat{\xi} \in H^1(X \times_R \hat{R}, PGL_n)$ d'image $\xi_0 \in H^1(X_0, PGL_n)$. Le théorème d'Artin assure l'existence de $\xi \in H^1(X, PGL_n)$ de même image que $\hat{\xi}$ dans $H^1(X_0, PGL_n)$, c'est-à-dire d'image ξ_0 . \square

3.4. Scindage des algèbres d'Azumaya sur une surface et mise en famille

Soit k un corps. Soit X une k -surface projective, lisse, géométriquement intègre et soit $A \in \text{Az}_m(X)$ une algèbre d'Azumaya sur X . Soit \mathcal{L} un faisceau inversible sur X . Notons $L = \text{Spec}(\text{Sym}(\mathcal{L}^{-1})) \rightarrow X$ le fibré en droites correspondant. À toute section $\sigma \in H^0(X, A \otimes \mathcal{L})$ on associe un idéal caractéristique $I_\sigma \subset \mathcal{O}_L$ ([OP], Lemma 2.10). Localement sur X il est défini de la façon suivante. Soit $U = \text{Spec}(R)$ un ouvert affine de X sur lequel on dispose d'un isomorphisme $\mathcal{O}_U \simeq \mathcal{L}_U$. Soit $f \in \mathcal{L}(U)$ l'image de 1. On a alors $L_U \simeq U[T] = \text{Spec}(R[T])$ et l'idéal I_σ est engendré par le polynôme caractéristique réduit $P_{f,U}[T]$ de $\sigma \cdot f^{-1}$ qui est un élément de l'algèbre d'Azumaya $A(U) \in \text{Az}_m(U)$. Ce polynôme est de degré n et unitaire. Ainsi le fermé $Y_\sigma \subset L$ défini par I_σ est fini et plat sur X , de degré m .

THÉORÈME 3.9 (M. Artin). — *Soient k un corps algébriquement clos, n un entier premier à la caractéristique de k , X une k -surface projective, lisse connexe, $A \in \text{Az}_m(X)$ telle que $A_{k(X)}$ est à division et \mathcal{L} un faisceau inversible. Si \mathcal{L} est suffisamment ample et σ est une section suffisamment générale du fibré vectoriel $A \otimes \mathcal{L}$, alors la surface Y_σ définie ci-dessus est lisse et connexe. Pour tout tel σ , l'image réciproque de A sur Y_σ a une classe triviale dans $\text{Br}(Y_\sigma)$.*

Ce théorème est annoncé par de Jong ([dJ], §8). Une démonstration en caractéristique nulle est donnée par Ojanguren et Parimala ([OP]). Ceux-ci apportent de multiples précisions sur le type de ramification auquel on peut de plus restreindre $Y_\sigma \rightarrow X$. Je ne donnerai pas la démonstration de ce théorème. La première assertion est un énoncé de type Bertini, qui utilise de façon essentielle le fait que la dimension de X est 2. La seconde est une conséquence du théorème 2.4. \square

En utilisant ce théorème Ojanguren et Parimala montrent ([OP]) :

PROPOSITION 3.10. — *Soient k un corps algébriquement clos de caractéristique zéro, X une k -surface projective, lisse, connexe, $\alpha \in \text{Br}(X)$. Il existe une k -variété lisse connexe W de dimension 3, et des morphismes*

$$\begin{array}{ccc} W & \xrightarrow{g} & X \\ f \downarrow & & \\ \mathbb{A}^1 & & \end{array}$$

satisfaisant les propriétés suivantes :

- (i) Le morphisme f est lisse à fibres connexes.
- (ii) Le morphisme $(g, f) : W \rightarrow X \times_k \mathbb{A}_k^1$ est quasi-fini et plat.
- (iii) Il existe un voisinage de $0 \in \mathbb{A}_k^1$ au-dessus duquel f est propre. En particulier la surface $Y = f^{-1}(0)$ est projective, lisse, connexe.
- (iv) On a $g^*(\alpha) = 0 \in \text{Br}(Y)$.
- (v) La fibre $W_1 = f^{-1}(1)$ est non vide, et la restriction de $g : W \rightarrow X$ à W_1 est une immersion ouverte.

PREUVE (ESQUISSE) — D'après le théorème 2.5, il existe une algèbre d'Azumaya $A \in \text{Az}_m(X)$, avec m convenable, de classe $\alpha \in \text{Br}(X)$, telle que $A_{k(X)}$ est à division. Soient \mathcal{L}, L et $\sigma \in H^0(X, \mathcal{L})$ comme au théorème 3.9. On choisit des sections globales distinctes w_1, \dots, w_m de L . On considère le fibré en droites $L \times_k \mathbb{A}_k^1 \rightarrow X \times_k \mathbb{A}_k^1$, où $\mathbb{A}_k^1 = \text{Spec}(k[t])$. Pour $U = \text{Spec}(R) \subset X$ et $f \in \mathcal{L}(U)$ comme ci-dessus, on considère l'idéal de $R[t]$ engendré par

$$Q_{f,U}(t, T) = (1 - t)P_{f,U}(T) + t(T - w_1/f) \dots (T - w_m/f).$$

On vérifie que ces différents idéaux se recollent en un idéal sur le schéma $L \times_k \mathbb{A}_k^1$. Soit $Z \subset L \times_k \mathbb{A}_k^1$ le fermé défini par cet idéal. La projection $Z \rightarrow X \times_k \mathbb{A}_k^1$ est finie et plate de degré m . L'application composée $f : Z \rightarrow X \times_k \mathbb{A}_k^1 \rightarrow \mathbb{A}_k^1$ est donc propre et plate. Sa fibre en $t = 0$ est la surface projective, lisse, connexe Y_σ . La fibre générique de f est donc lisse et géométriquement intègre. Comme $Z \rightarrow X$ est plat, ceci implique que Z est intègre. La fibre au-dessus de $t = 1$ contient n composantes fermées distinctes chacune birationnellement isomorphe à X via la projection sur X . En enlevant un fermé convenable dans Z , on obtient W comme annoncé dans la proposition. \square

REMARQUE 3.11. — Dans [dJ], le théorème 3.9 n'est pas utilisé. L'idée de base, qui est de déformer un polynôme en une variable, unitaire, de degré m , qui définit l'extension $k(Y)/k(X)$, en un polynôme unitaire séparable de même degré avec toutes ses racines dans $k(X)$, est la même, mais la construction d'une bonne famille satisfaisant des propriétés plus faibles que celles de la proposition 3.10, est alors plus délicate ([dJ], §4). De Jong y a en particulier recours au « théorème de la fibre réduite » ([BLR], [dJS2]), qui avait déjà été utilisé dans [dJS1].

3.5. Le théorème de de Jong dans le cas non ramifié

PROPOSITION 3.12. — Soit X une surface projective, lisse, connexe sur un corps k algébriquement clos, $n > 0$ entier inversible dans k . Soit $\alpha \in \text{Br}(X) \subset \text{Br}(k(X))$, d'exposant n .

Supposons qu'il existe une variété intègre W de dimension 3 et des morphismes

$$\begin{array}{ccc} W & \xrightarrow{g} & X \\ f \downarrow & & \\ \mathbb{A}^1 & & \end{array}$$

satisfaisant les propriétés suivantes :

(i) Le morphisme f est lisse à fibres connexes.

(ii) Le morphisme f est propre au voisinage du point $0 \in \mathbb{A}^1$, de fibre $W_0 = f^{-1}(0)$, et g induit un morphisme fini et plat $g_{W_0} : W_0 \rightarrow X$ de surfaces lisses.

(iii) La restriction de g à la fibre $W_1 = f^{-1}(1)$ est une immersion ouverte $W_1 \hookrightarrow X$.

(iv) $g^*(\alpha)_{W_0} = 0 \in \text{Br}(W_0)$.

Alors l'indice de $\alpha_{k(X)} \in \text{Br}(k(X))$ est égal à n .

PREUVE — Soit $\eta \in H^2(X, \mu_n)$ d'image $\alpha \in {}_n\text{Br}(X)$. Notons $Y = W_0$. Soit \mathcal{L}_Y un faisceau inversible sur Y dont la classe par l'application composée $\delta : \text{Pic}(Y)/n \rightarrow H^2(Y, \mu_n)$ est l'opposé de $g^*(\eta)_Y$. Qu'un tel faisceau inversible existe résulte de la suite exacte de Kummer et de l'hypothèse (iv). Considérons l'algèbre d'Azumaya

$$B_0 = \text{End}(\mathcal{L}_Y \oplus (\mathcal{O}_Y)^{n-1}).$$

Elle est de degré n . D'après le lemme 3.1 (c) on a $\text{cl}(B_0) = -\delta(\mathcal{L}_Y) = g^*(\eta)_Y$. D'après le théorème 3.5, une transformation élémentaire convenable de B_0 produit une algèbre $A_0 \in \text{Az}_n(Y)$ satisfaisant $\text{cl}(A_0) = \text{cl}(B_0) = g^*(\eta)_Y \in H^2(Y, \mu_n)$ et $H^2(Y, A_0/\mathcal{O}_Y) = 0$. C'est un point-clé de la démonstration. Cette nullité permet de déformer l'algèbre $A_0 \in \text{Az}_n(Y)$. En utilisant le théorème d'algébrisation 3.7 puis le théorème d'approximation 3.8, on voit qu'il existe un voisinage étale connexe $h : (C', 0) \rightarrow (C, 0)$ du point $0 \in C = \mathbb{A}_k^1$, tel que $W' = W \times_C C' \rightarrow C'$ soit propre et lisse, et une algèbre d'Azumaya $A \in \text{Az}_n(W')$ telle que la fibre de A au-dessus de $0 \in C'$ soit isomorphe à l'algèbre A_0 sur Y . (Notons que l'algèbre A n'a pas de raison d'avoir une classe triviale dans le groupe de Brauer de W' , même après complétion en $0 \in C'$.)

La réduction de $\text{cl}(A) \in H^2(W', \mu_n)$ au-dessus de $0 \in C'$ s'identifie à $\text{cl}(A_0) = g^*(\eta)_Y$, c'est-à-dire à la réduction de l'image au-dessus de $0 \in C'$ de la classe globale obtenue à partir de $\eta \in H^2(X, \mu_n)$ en prenant l'image réciproque par le morphisme composé $g' : W' \rightarrow W \rightarrow X$.

En appliquant le théorème de changement de base propre en cohomologie étale ([Mi], VI.2.7) au morphisme $W' \rightarrow C'$, on voit qu'en remplaçant $(C', 0)$ par un autre voisinage étale, encore noté $(C', 0)$, on peut assurer $\text{cl}(A) = g'^*(\eta) \in H^2(W', \mu_n)$.

On peut étendre le morphisme $C' \rightarrow C$ en un morphisme fini de courbes lisses $D \rightarrow C = \mathbb{A}_k^1$. Le morphisme $W \times_C D \rightarrow D$ est lisse, et comme k est algébriquement clos il existe un point $1 \in D$ au-dessus de 1 , la fibre W'_1 en ce point est lisse, connexe, et la projection composée $W \times_C D \rightarrow W \rightarrow X$ induit sur W'_1 une immersion ouverte $W'_1 \subset X$.

Soit R l'anneau local de $W \times_C D$ au point générique de W'_1 . C'est un anneau de valuation discrète, de corps des fractions K le corps des fonctions de W' , de corps résiduel isomorphe à $k(X)$. De plus, la projection $W \times_C D \rightarrow W \rightarrow X$ induit une inclusion $k(X) \hookrightarrow R$ qui composée avec l'application de réduction modulo l'idéal maximal de R , soit $R \rightarrow k(X)$, est l'identité de $k(X)$. Soit $A_K \in \text{Az}_n(K)$ l'image de $A \in \text{Az}_n(W')$. L'égalité $\text{cl}(A) = g'^*(\eta) \in H^2(W', \mu_n)$ implique que la classe $[A_K] \in \text{Br}(K)$ coïncide avec l'image de $\alpha_{k(X)} \in \text{Br}(k(X))$ via la flèche composée $k(X) \rightarrow R \rightarrow K$. En particulier, $[A_K]$ appartient à $\text{Br}(R) \subset \text{Br}(K)$. L'algèbre A_K est de degré n . La combinaison de ces deux derniers faits et du théorème 2.5 (dans le cas particulier d'un anneau de valuation discrète) assure que la spécialisation $\text{Br}(R) \rightarrow \text{Br}(k(X))$ envoie A_K sur la classe d'une algèbre d'indice divisant n . Comme cette spécialisation est égale à $\alpha_{k(X)}$, et que cette dernière est d'exposant n , on voit que l'indice de $\alpha_{k(X)}$ est n . \square

Combinant les propositions 3.10 et 3.12, on obtient une démonstration en caractéristique nulle du théorème :

THÉORÈME 3.13 (de Jong). — *Soit k un corps algébriquement clos. Soit X/k une surface connexe, projective et lisse sur k . Soit $\alpha \in \text{Br}(X)$ d'exposant $n > 0$ premier à la caractéristique de k . Alors l'indice de $\alpha_{k(X)}$ est égal à n , et α est représenté par une algèbre d'Azumaya A sur X de degré n .*

La dernière assertion est une conséquence du reste du théorème et du théorème 2.5. Une démonstration de la proposition 3.10 en caractéristique positive première à l'exposant de l'algèbre A permettrait d'étendre la démonstration ici décrite. De Jong quant à lui établit le résultat dans cette généralité en utilisant la mise en famille évoquée à la remarque 3.11 ci-dessus.

3.6. Le cas ramifié résulte du cas non ramifié

THÉORÈME 3.14 (de Jong). — *Soit k un corps algébriquement clos. Soit X/k une surface connexe, projective et lisse sur k . Soit $\alpha \in \text{Br}(k(X))$ d'exposant $n > 0$ premier à la caractéristique de k . Alors $\text{ind}_{k(X)}(\alpha) = n$.*

PREUVE — La décomposition des corps gauches en produits tensoriels de corps gauches d'indices premiers entre eux permet de se ramener au cas où $n = l^r$ avec l premier. Montrons le résultat par récurrence sur r , en supposant le résultat connu pour $r = 1$. L'exposant de $\beta = l\alpha \in \text{Br}(k(X))$ est l^{r-1} . Il existe donc une extension de corps $E/k(X)$ de degré l^{r-1} telle que $l\alpha_E = \beta_E = 0 \in \text{Br}(E)$. La résolution des singularités des surfaces algébriques donne l'existence d'une surface projective lisse connexe Y sur k telle que $E = k(Y)$. Le résultat pour $n = l$ premier appliqué à la surface Y assure l'existence d'une extension de corps F/E de degré l telle que $(\alpha_E)_F = 0 \in \text{Br}(F)$. Ainsi $\alpha_F = 0$, et l'extension de corps $F/k(X)$ est de degré l^r . Ce type de réduction est classique ([A], p. 175 ; [CTG], p. 132).

On suppose désormais $n = l$ premier, distinct de la caractéristique de k .

On va construire des morphismes

$$\begin{array}{ccc} W & \xrightarrow{g} & X \\ f \downarrow & & \\ \mathbb{A}^1 & & \end{array}$$

de variétés lisses connexes satisfaisant les propriétés suivantes :

- (i) Le morphisme $f : W \rightarrow \mathbb{A}_k^1$ est lisse.
- (ii) La fibre générique géométrique $W_{\bar{\eta}}$ de $f : W \rightarrow \mathbb{A}_k^1$ est projective et connexe.
- (iii) La fibre $W_0 \subset W$ de f en $0 \in \mathbb{A}_k^1$ est non vide et la restriction de $g : W \rightarrow X$ à $W_0 \subset W$ est un morphisme birationnel $W_0 \rightarrow X$.
- (iv) L'image réciproque $r^*(\alpha)$ de α par l'application composée $r : W_{\bar{\eta}} \rightarrow W_{\eta} \rightarrow W \rightarrow X$, où la dernière flèche est g , est non ramifiée sur la surface $W_{\bar{\eta}}$ (on note η le point générique de \mathbb{A}_k^1 et $\bar{\eta}$ un point générique géométrique).

Supposons ces morphismes construits. De la propriété (iv) et du théorème 3.13 on déduit qu'il existe une extension finie de corps $K = k(\mathbb{A}^1) \hookrightarrow L$ telle que la restriction de α à $W_{\eta} \times_K L$ soit de degré l . Le corps L est le corps des fonctions d'une courbe affine lisse C munie d'un k -morphisme fini $C \rightarrow \mathbb{A}_k^1$. Comme k est algébriquement clos, il existe un k -point $M \in C$ au-dessus de $0 \in \mathbb{A}_k^1$. La variété $Z = W \times_{\mathbb{A}^1} C$ est lisse, la fibre générique de $Z \rightarrow C$ est géométriquement intègre, la variété Z est donc connexe. La fibre de Z_M de $Z \rightarrow C$ au-dessus du point M est lisse, connexe, la restriction de la projection $Z \rightarrow X$ à la fibre Z_M est un morphisme birationnel $Z_M \rightarrow X$.

Soit R l'anneau local de Z au point générique ξ de Z_M . Le corps des fractions de R est le corps des fonctions de Z , le corps résiduel κ est isomorphe à $k(X)$, plus précisément le composé $\text{Spec}(\kappa) \rightarrow \text{Spec}(R) \rightarrow W \rightarrow X$ s'identifie à l'inclusion du point générique de X dans X . L'image réciproque sur le corps des fonctions de Z de la classe $\alpha \in \text{Br}(k(X))$, via la projection $Z \rightarrow X$, est une classe dans $\text{Br}(k(Z))$ non ramifiée au point générique ξ de Z_M , dont la réduction sur κ s'identifie à α . Pour établir le théorème 3.14, il suffit alors d'appliquer le théorème 2.5 dans le cas particulier du spectre d'un anneau de valuation discrète.

Il reste donc à construire un morphisme $h = (g, f) : W \rightarrow X \times \mathbb{A}_k^1$ satisfaisant les propriétés (i) à (iv). Quitte à remplacer la surface X par un éclaté, on peut supposer ([Li]) que le lieu de ramification de la classe $\alpha \in \text{Br}(k(X))$ est un diviseur D strictement à croisements normaux (composantes lisses, par tout point il passe au plus deux composantes, et dans ce cas elles se coupent transversalement).

Soit $\mathbb{A}^1 = \text{Spec}(k[t])$. On cherche un revêtement génériquement cyclique $h = (g, f) : W \rightarrow X \times \mathbb{A}_k^1$ de degré l , donné par la racine l -ième d'une fonction rationnelle f_t sur $X \times \mathbb{A}^1$, de telle sorte qu'au-dessus du point générique géométrique de \mathbb{A}^1 , on ait détruit la

ramification de α , et de telle sorte qu'au-dessus du point $0 \in \mathbb{A}_k^1$ la fibre de f contienne une composante de multiplicité 1 birationnelle à X . En bref, on veut que la spécialisation f_0 de f soit une puissance l -ième, et on veut que le diviseur de f contienne le diviseur D à l'ordre 1 (dans le cas modéré, « la ramification avale la ramification » – lemme d'Abhyankar).

LEMME 3.15. — *Etant donné un diviseur $D \subset X$ strictement à croisements normaux, il existe des diviseurs E et E' effectifs lisses tels que $D + E$ soit linéairement équivalent à $l(D + E')$, que deux de D, E et E' n'aient pas de composante commune, et que $D + E + E'$ soit à croisements normaux.*

PREUVE — Soit \mathcal{M} un fibré inversible ample sur X . Pour r entier suffisamment grand, le faisceau $\mathcal{O}_X((l-1)D) \otimes \mathcal{M}^{\otimes r}$ est très ample, il existe une section de ce faisceau qui définit sur X un diviseur lisse E qui coupe D transversalement (noter que les supports de D et de E se rencontrent nécessairement, le diviseur $D + E$ n'est pas lisse). En outre il existe une section de $\mathcal{M}^{\otimes r}$ dont le lieu des zéros est un diviseur E' lisse transverse à $D + E$. \square

Soit $\mathcal{L} = \mathcal{O}_X(-D - E')$. Soit s_0 une section de $\mathcal{L}^{\otimes -l}$ de lieu des zéros $D + E$. Soit s_1 une section de \mathcal{L}^{-1} de lieu des zéros $D + E'$. On dispose alors de la section $s_1^l = s_1^{\otimes l}$ de $\mathcal{L}^{\otimes -l}$.

Soit $\tilde{\mathcal{L}}$ l'image inverse de \mathcal{L} sur $Y = X \times \mathbb{A}^1$. Soit s_t la section de $\tilde{\mathcal{L}}^{\otimes -l}$ définie par

$$s_t = t^l s_1^l + (1-t)^l s_0,$$

où, par abus de notation, s_0 et s_1 sont les images réciproques sur $X \times \mathbb{A}^1$ de s_0 et s_1 .

On définit alors, de façon classique, un revêtement génériquement cyclique Y de $X \times \mathbb{A}^1$ de la façon suivante. On définit une structure de $\mathcal{O}_{X \times \mathbb{A}^1}$ -algèbre sur le fibré vectoriel

$$\tilde{\mathcal{N}} = \bigoplus_{i=0}^{l-1} \tilde{\mathcal{L}}^{\otimes i}$$

en utilisant la flèche $\tilde{\mathcal{L}}^{\otimes l} \rightarrow \mathcal{O}_{X \times \mathbb{A}^1}$ donnée par la section s_t de $\tilde{\mathcal{L}}^{\otimes -l}$.

Soit $U = \text{Spec}(A)$ un ouvert affine de X sur lequel \mathcal{L} admet une trivialisaton $\mathcal{L} \simeq \mathcal{O}_U$. Au-dessus de U , le revêtement décrit ci-dessus est isomorphe à

$$\text{Spec}(A[t][x]/(x^l - (t^l f_1^l + (1-t)^l f_0))),$$

pour f_0 et f_1 dans A convenables.

On vérifie alors facilement :

- (a) La variété Y est intègre.
- (b) La fibre générique Y_η du morphisme composé $Y \rightarrow X \times \mathbb{A}^1 \rightarrow \mathbb{A}^1$ est géométriquement intègre.
- (c) La fibre Y_1 du morphisme $Y \rightarrow X \times \mathbb{A}^1 \rightarrow \mathbb{A}^1$ au-dessus du point $t = 1$ a l composantes, chacune de multiplicité 1, chacune isomorphe à X , l'isomorphisme étant induit par le morphisme composé $Y_1 \subset Y \rightarrow X$.

La K -variété Y_η ne saurait être lisse, car le lieu des zéros de la section s_t n'est pas lisse : il est somme de $D \times_k K$ et d'un diviseur effectif non nul E_t . Le lieu des zéros de s_0 sur $Y_0 = X$ est $D + E$, qui est strictement à croisements normaux. Pour t général, ceci implique (Bertini) la même propriété pour la décomposition $D + E_t$ du lieu des zéros de s_t . Ceci assure que pour tout tel t , la variété Y_t est normale et n'a que des singularités de type A_{l-1} : une équation locale en un point singulier est analytiquement du type $A[x]/x^l - uv$, où u, v sont des éléments d'un système régulier de paramètres. On considère alors la résolution minimale des singularités de la fibre générique Y_η/K , soit Z_η/K . Cette résolution est lisse. La résolution $Z_\eta \rightarrow Y_\eta$ est donnée par l'éclatement d'un certain idéal cohérent I_η sur Y_η de support les points singuliers. On peut trouver un idéal cohérent I sur Y dont le lieu des zéros est de dimension au plus 1 et dont la restriction à la fibre générique est I_η . Soit $Z \rightarrow Y$ l'éclaté de Y au moyen de l'idéal I . Il existe alors un ouvert W de Z qui satisfait les propriétés (i), (ii) et (iii) de l'énoncé (pour la propriété (iii), il suffit d'observer qu'au-dessus des points génériques des composantes des fibres de $Y \rightarrow \mathbb{A}^1$, l'éclatement via I ne modifie rien ; on utilise alors le résultat pour $Y \rightarrow \mathbb{A}^1$).

Il reste à établir le point (iv). La technique ici est standard ([Ar3], [FS], [CTOP]). Revenons à la situation envisagée dans la démonstration principale. Soit L une clôture algébrique de $K = k(t)$. Le morphisme de L -surfaces projectives, lisses, connexes $Z_L \rightarrow X_L$ est donné au point générique de X_L par l'extraction de la racine l -ième de la fonction rationnelle $f_t = s_t/s_1^l$, dont le diviseur est $D + E_t - l(D + E')$. Le lieu de ramification de $\alpha_K \in \text{Br}(X_K)$ est contenu dans D_K . Le diviseur $D + E_t$ est réduit, strictement à croisements normaux.

Soit M un point de codimension 1 de Z_L . Si son image par $Z_L \rightarrow X_L$ n'est pas contenue dans D_L , alors $\alpha_{L(Z)}$ est non ramifié en M . Si l'image de M est un point générique N d'une composante de D_K , alors $\alpha_{L(Z)}$ est non ramifié en M . On a en effet le diagramme commutatif

$$\begin{array}{ccc} \text{Br}(L(Z)) & \longrightarrow & H^1(\kappa_M, \mathbf{Q}/\mathbf{Z}) \\ \uparrow & & \uparrow \\ \text{Br}(L(X)) & \longrightarrow & H^1(\kappa_N, \mathbf{Q}/\mathbf{Z}) \end{array}$$

où la flèche $H^1(\kappa_N, \mathbf{Q}/\mathbf{Z}) \rightarrow H^1(\kappa_M, \mathbf{Q}/\mathbf{Z})$ est la multiplication par l'indice de ramification en N , qui est l .

Supposons maintenant que l'image de M soit un point fermé N de X_L . Si le point N n'appartient pas à D , alors $\alpha_{L(X)}$ s'étend en une algèbre d'Azumaya au voisinage de N donc $\alpha_{L(Z)}$ est non ramifié en M . Supposons que le point N appartienne à D mais soit sur une unique composante de D . Alors (proposition 2.6) il existe un voisinage affine $\text{Spec}(R)$ de N dans X_L , $u \in R^*$ et $s \in R$ tels que $s = 0$ définisse exactement D sur $\text{Spec}(R)$ et que $\alpha_{L(X)} - (u, s)$ appartienne à $\text{Br}(R)$ (on omet ici l'indice ζ_l dans la notation d'une algèbre cyclique). On a donc $\partial_M(\alpha_{L(Z)}) = \partial_M((u, s)) = \bar{u}^{v_M(s)} \in \kappa_M/\kappa_M^{*l}$.

La classe de u dans κ_M/κ_M^{*l} est l'image de la classe de u dans $\mathcal{O}_{X_L,N}^*/\mathcal{O}_{X_L,N}^{*l}$, et l'application naturelle $\mathcal{O}_{X_L,N}^*/\mathcal{O}_{X_L,N}^{*l} \rightarrow \kappa_M/\kappa_M^{*l}$ se factorise par $\kappa_N^*/\kappa_N^{*l} = 1$ (le corps résiduel κ_N est séparablement clos de caractéristique différente de l .) Supposons maintenant que le point N soit l'intersection de deux composantes de D . Il n'est donc pas sur E_t . Il existe un voisinage affine $\text{Spec}(R)$ de N dans X_L , $s, t \in R$ engendrant l'idéal maximal de N , tels que $st = 0$ définisse exactement D sur $\text{Spec}(R)$ et que E_t ne rencontre pas $\text{Spec}(R)$. Sur $\text{Spec}(R)$ on a donc $f_t = csth^l \in L(X)^*$ avec $c \in R^*$ et $h \in L(X)^*$.

D'après la proposition 2.6, quitte à restreindre $\text{Spec}(R)$, il existe des unités $u, v \in R^*$ et $r \in \mathbf{Z}$ tels que $\alpha + (s, u) + (t, v) + r(s, t)$ appartienne à $\text{Br}(R)$. On a

$$(s, t) = (s, f_t c^{-1} s^{-1} h^{-l}) = (s, f_t) + (s, -c^{-1}) \in \text{Br}(L(X)),$$

où l'on a utilisé la formule $(s, -s) = 0$. On peut donc écrire $\alpha \in \text{Br}(L(X))$ comme la somme d'éléments du type : non ramifié, (s, f_t) , (s, u) et (t, v) avec u et v unités. On a $L(Z) = L(X)(f_t^{1/l})$, donc $(s, f_t)_{L(Z)} = 0$. Le même argument que ci-dessus montre que $(s, u)_{L(Z)}$ et $(t, v)_{L(Z)}$ ont des résidus triviaux en M . Ainsi $\alpha_{L(Z)}$ est non ramifié en M .

Ceci établit le point (iv) et achève la démonstration du théorème de de Jong. \square

REMARQUE 3.16. — Le théorème de de Jong peut se reformuler ainsi :

Soit K un corps de fonctions de deux variables sur un corps algébriquement clos. Soit A une algèbre simple centrale sur K de degré nm et d'exposant n premiers à la caractéristique de K . Alors la K -variété de Severi-Brauer généralisée $SB(A, n)$ possède un K -point.

On peut se demander ce qui fait que la K -variété $SB(A, n)$ a automatiquement un point rationnel. Dans un travail en préparation, de Jong et Starr étudient la notion de 1-connexité rationnelle et donnent des conditions suffisantes pour qu'une K -variété projective et lisse possède automatiquement un point rationnel sur K , corps de fonctions de deux variables sur un corps algébriquement clos. Leur résultat est un analogue du résultat de Graber, Harris, Starr, de Jong sur l'existence de points rationnels sur les variétés (séparablement) rationnellement connexes définies sur un corps de fonctions d'une variable. Les hypothèses mises par de Jong et Starr dans le cas d'un corps de fonctions de deux variables sont assez contraignantes, mais elles s'appliquent aux variétés $SB(A, n)$.

4. CONSÉQUENCES POUR LES GROUPES ALGÈBRIQUES LINÉAIRES

Soit K un corps de fonctions de deux variables sur un corps algébriquement clos de caractéristique zéro. Pour un tel corps, les deux propriétés générales suivantes sont donc satisfaites :

(1) *C'est un corps de dimension cohomologique $\text{cd}(K) \leq 2$ ([S2]).*

(2) *Sur tout corps extension finie de K , indice et exposant des algèbres simples centrales coïncident.*

Ces deux propriétés sont satisfaites par les corps mentionnés au paragraphe 2.4, du moins par ceux qui ne sont pas formellement réels :

- les corps p -adiques
- les corps de nombres totalement imaginaires
- les corps de la forme $F((t))$, où F est un corps de caractéristique zéro de dimension cohomologique 1
- le corps des fractions d'un anneau local intègre, excellent, hensélien, de dimension 2, à corps résiduel algébriquement clos de caractéristique zéro ([Ar3], [FS], [CTOP]).

Un corps de fonctions de deux variables sur un corps algébriquement clos de caractéristique zéro, est un corps C'_2 (théorème de Tsen-Lang), ce qui, d'après Merkur'ev et Suslin est une propriété plus forte que la propriété (1). Comme mentionné au §2, une autre application du théorème de Merkur'ev et Suslin montre que la propriété C'_2 implique qu'exposant et indice coïncident pour les algèbres 2-primaires et 3-primaires. Introduisons la condition suivante, plus faible que (2).

(2') *Sur tout corps L extension finie de K , indice et exposant des L -algèbres simples centrales 2-primaires ou 3-primaires coïncident.*

Dans [CTGP], qui repose sur les travaux antérieurs de nombreux auteurs, on a étudié de façon systématique les propriétés des corps de caractéristique zéro satisfaisant les hypothèses (1) et (2), ou (1) et (2'). Les énoncés généraux suivants, extraits de [CTGP], s'appliquent donc aux corps de fonctions de deux variables sur un corps algébriquement clos de caractéristique zéro.

L'énoncé suivant (cf. [CTGP], Thm. 1.2) est un cas particulier de la conjecture II de Serre ([S2], [S3]). Il rassemble des travaux successifs de Merkur'ev-Suslin, Suslin, Bayer-Fluckiger et Parimala, P. Gille, Chernousov.

THÉORÈME 4.1. — *Soit K un corps de caractéristique zéro satisfaisant (1) et (2'). Soit G un K -groupe algébrique semi-simple simplement connexe. Si G contient un facteur de type E_8 , supposons en outre $\text{cd}(K^{\text{ab}}) \leq 1$. Alors $H^1(K, G) = 0$.*

On note ici K^{ab} l'extension abélienne maximale de K . La question de savoir si l'on a $\text{cd}(K^{\text{ab}}) \leq 1$ pour un corps K de fonctions de deux variables sur les complexes est ouverte.

Dire que sur un corps K (de caractéristique zéro) exposant et indice coïncident pour les K -algèbres simples centrales équivaut à dire que pour tout entier $n \geq 2$ l'application

bord

$$H^1(K, PGL_n) \rightarrow H^2(K, \mu_n)$$

déduite de la suite exacte

$$1 \rightarrow \mu_n \rightarrow SL_n \rightarrow PGL_n \rightarrow 1$$

est surjective. On peut donc se poser des questions analogues avec d'autres isogénies. Une longue analyse par type de groupe (le cas 2A_n étant particulièrement délicat) mène au résultat suivant, qui requiert toute l'hypothèse (2), et vaut donc pour les corps de fonctions de deux variables sur les complexes.

THÉORÈME 4.2 ([CTGP], Thm. 2.1). — *Soit K un corps de caractéristique zéro satisfaisant (1) et (2). Soit G un K -groupe semi-simple simplement connexe, de centre μ , de groupe adjoint G^{ad} . Si G contient un facteur de type E_8 , supposons en outre $\text{cd}(K^{\text{ab}}) \leq 1$.*

(i) *L'application bord $H^1(K, G^{\text{ad}}) \rightarrow H^2(K, \mu)$ associée à la suite exacte*

$$1 \rightarrow \mu \rightarrow G \rightarrow G^{\text{ad}} \rightarrow 1$$

est une bijection.

(ii) *Si le groupe G n'est pas purement de type A , alors il est isotrope.*

(Ce théorème généralise des résultats classiques de Kneser sur les corps locaux et sur les corps globaux.)

THÉORÈME 4.3 (cf. [CTGP], Thm. 4.5). — *Soit K un corps de caractéristique zéro satisfaisant (1) et (2'). Soit G un K -groupe semi-simple simplement connexe. Si G ne contient pas de facteur de type E_8 , ou si l'on a en outre $\text{cd}(K^{\text{ab}}) \leq 1$, alors $G(K)/R = 1$.*

(Pour la notion de R -équivalence, voir [G1].)

La démonstration se fait en discutant chaque type simple. Il y a une différence marquée entre le cas des groupes de type A_n et les autres types simples. Pour les autres types simples, la K -variété sous-jacente au K -groupe G est K -birationnelle à un espace affine (défini sur K), ce qui n'est pas le cas en général pour les groupes de type A_n , comme le montre un exemple de Merkur'ev.

Le théorème 4.3 est utilisé dans la démonstration du théorème suivant, analogue d'un résultat de P. Gille ([G1], [G2]).

THÉORÈME 4.4 ([CTGP], Thm. 4.12). — *Soient K un corps de fonctions de deux variables sur un corps algébriquement clos de caractéristique zéro et G un K -groupe linéaire connexe sans facteur de type E_8 . Le groupe $G(K)/R$ est un groupe abélien fini.*

(Ici encore la condition sur E_8 pourrait être omise si l'on savait établir $\text{cd}(K^{\text{ab}}) \leq 1$.)

La technique des résolutions flasques de groupes linéaires connexes permet de donner une formule pour le groupe abélien fini $G(K)/R$. Je renvoie pour cela à [CTGP], §4.

5. QUELQUES RÉSULTATS SUR LES CORPS DE FONCTIONS D'UNE VARIABLE SUR UN CORPS p -ADIQUE

THÉORÈME 5.1 (Saltman [Sa1]). — *Soient k un corps p -adique et K un corps de fonctions d'une variable sur k . Soit l premier, $l \neq p$. Supposons $\mu_l \subset k$. Etant donné un ensemble fini d'algèbres simples centrales $A_i, i \in I$, chacune d'exposant l dans le groupe de Brauer de K , il existe f et g dans K^* tels que l'extension $K(f^{1/l}, g^{1/l})$ déploie chacune des algèbres $A_i, i \in I$.*

L'outil de base de la démonstration est le théorème suivant.

THÉORÈME 5.2 (Tate, Lichtenbaum, Grothendieck). — *Soit O l'anneau des entiers d'un corps p -adique et soit $Y/\text{Spec}(O)$ un schéma connexe, régulier, propre et plat sur $\text{Spec}(O)$, de dimension relative 1. Alors le groupe de Brauer de Y est trivial.*

En utilisant la proposition 2.3 on voit alors que le « groupe de Brauer non ramifié » du corps des fonctions K d'une telle surface arithmétique Y est trivial. En particulier si un élément $A \in {}_l\text{Br}(K)$ a la propriété que pour tout anneau de valuation discrète R de corps des fractions K , de corps résiduel κ , le résidu $\delta(A) \in \kappa^*/\kappa^{*l}$ est trivial, alors $A = 0 \in \text{Br}(K)$. On notera que l'image du point fermé s_R de $\text{Spec}(R)$ dans Y peut être soit le point générique d'une courbe soit un point fermé de Y .

La démonstration du théorème 5.1 est analogue dans son principe à celle de [Ar3], [FS], [CTOP] (où là on n'extrait qu'une seule racine l -ième). On commence par écrire le corps K comme le corps des fonctions d'un schéma X connexe, régulier, propre et plat sur $\text{Spec}(O)$, de dimension relative 1, choisi de telle sorte que la réunion des supports des diviseurs de ramification des A_i soit à croisements normaux stricts sur X ; par éclatements convenables, on peut assurer cela sur un schéma régulier excellent de dimension 2 (Lipman [Li]). On détermine ensuite des fonctions $f, g \in K^*$ telles que les classes A_i deviennent non ramifiées sur le corps $L = K(f^{1/l}, g^{1/l})$, qui est le corps des fonctions d'un schéma Y connexe, régulier, propre et plat sur $\text{Spec}(O)$, de dimension relative 1, muni d'un morphisme génériquement fini $Y \rightarrow X$. Le théorème 5.2 assure alors la nullité de ces classes dans $\text{Br}(L)$. Pour choisir les fonctions f et g , on discute les images possibles dans X des morphismes composés $s_R \rightarrow \text{Spec}(R) \rightarrow Y \rightarrow X$, pour R comme ci-dessus. Pour le choix (délicat) des fonctions f et g , je renvoie à [Sa1] et à [HVG].

Un argument algébrique simple permet de déduire du théorème 5.1 l'énoncé suivant :

COROLLAIRE 5.3 (Saltman [Sa1]). — *Etant donné un corps K comme ci-dessus et une algèbre simple centrale A d'exposant n premier à p , l'indice de A divise n^2 .*

Des exemples dus à Jacob et Tignol ([Sa1], voir aussi [KRTY]) montrent que cette borne est en général la meilleure. Sur le corps $K = \mathbf{Q}_p(x)$, avec $p \neq 2$, si $a \in \mathbf{Z}_p$ est une unité qui

n'est pas un carré, alors le produit tensoriel d'algèbres de quaternions $(x, a) \otimes_K (x + 1, p)$ est une algèbre à division, donc d'indice 4 mais d'exposant 2 (comparer avec l'exemple donné après la proposition 2.7).

Saltman vient récemment d'analyser ce type d'exemple. Cela lui a permis de montrer :

THÉORÈME 5.4 (Saltman [Sa2]). — *Soient k un corps p -adique et K un corps de fonctions d'une variable sur k . Soit D une algèbre à division sur K d'indice l premier, $l \neq p$. Une telle algèbre est cyclique : D contient un sous-corps commutatif maximal cyclique sur K .*

La démonstration passe par une caractérisation des algèbres de degré l en termes de leur ramification, sur un modèle régulier propre convenable de K sur l'anneau des entiers du corps p -adique k . Elle utilise le théorème 5.2.

Un corps de fonctions d'une variable sur un corps local d'égale caractéristique $\mathbf{F}_l((t))$ est un corps C_3 . En particulier, toute forme quadratique en au moins 9 variables sur un tel corps possède un zéro non trivial. Il est donc naturel de poser la question : si K est un corps de fonctions d'une variable sur un corps p -adique, toute forme quadratique en au moins 9 variables possède-t-elle un zéro non trivial ?

Merkur'ev remarqua le premier qu'en combinant le théorème 5.1 (pour $l = 2$) sur les corps de fonctions d'une variable sur un corps p -adique (avec $p \neq 2$) avec les résultats généraux sur les formes quadratiques obtenus grâce à la K -théorie algébrique (en fait uniquement le théorème de Merkur'ev [M1]), on obtient au moins une borne supérieure pour la dimension d'une forme quadratique anisotrope définie sur un tel corps. Son résultat fut amélioré par Hoffmann et Van Geel [HVG] puis par Parimala et Suresh :

THÉORÈME 5.5 (Parimala/Suresh [PS1]). — *Soit K un corps de fonctions d'une variable sur un corps p -adique, avec $p \neq 2$. Toute forme quadratique en au moins 11 variables sur K possède un zéro non trivial.*

Avant d'indiquer le principe de la démonstration, commençons par quelques rappels. Le groupe de Witt WK d'un corps K de caractéristique différente de 2 est par définition l'ensemble des classes d'isomorphie de formes quadratiques non dégénérées, muni de la somme directe orthogonale et quotienté par la classe du plan hyperbolique standard $x^2 - y^2$. Le produit tensoriel des formes quadratiques lui donne une structure d'anneau. L'idéal des classes de formes de rang pair est noté $IK \subset WK$. L'idéal $I^n K$ est engendré additivement par les n -formes de Pfister. L'intersection de tous les $I^n K$ est réduit à zéro. Plus précisément, toute forme anisotrope appartenant à $I^n K$ a une dimension au moins égale à 2^n (Arason-Pfister). La conjecture de Milnor établie par Voevodsky implique (Orlov-Vishik-Voevodsky) l'existence d'isomorphismes $I^n K / I^{n+1} K \simeq H^n(K, \mathbf{Z}/2)$. Dans ces isomorphismes, une n -forme de Pfister correspond à un n -symbole.

On peut alors facilement établir l'énoncé général suivant.

PROPOSITION 5.6 ([Kh], [PS2]). — Soit K un corps de caractéristique différente de 2. Soit $N > 0$ un entier tel que $H^{N+1}(K, \mathbf{Z}/2) = 0$; on suppose que pour tout entier n avec $1 \leq n \leq N$ il existe un entier $\lambda_n(K)$ tel que tout élément de $H^n(K, \mathbf{Z}/2)$ est somme d'au plus $\lambda_n(K)$ symboles. Alors il existe un entier positif $u(K)$ tel que toute forme quadratique en strictement plus de $u(K)$ variables ait un zéro non trivial.

PREUVE — L'hypothèse et les rappels ci-dessus impliquent d'une part que $I^{N+1}K = 0$ (et $H^m(K, \mathbf{Z}/2) = 0$ pour tout $m > N$), d'autre part que tout élément de $I^n K$ peut s'écrire comme une somme orthogonale de $\lambda_n(K)$ n -formes de Pfister et d'une forme appartenant à $I^{n+1}K$. Ainsi dans WK toute forme quadratique est représentée par une forme de rang au plus $1 + \sum_{n=1}^N 2^n \lambda_n(K)$. Ceci implique que toute forme quadratique de rang strictement plus grand que $1 + \sum_{n=1}^N 2^n \lambda_n(K)$ est isotrope. \square

La majoration obtenue peut être améliorée ([Kh] Prop. 1.2.d) :

$$u(K) \leq 2 + \sum_{n=2}^N (2^n - 2) \lambda_n(K).$$

Dans le cas qui nous intéresse ici, le corps K est de dimension cohomologique 3 par des arguments généraux, en particulier $H^4(K, \mathbf{Z}/2) = 0$. On a la borne évidente $\lambda_1(K) = 1$ (valable sur tout corps). Le théorème 5.1 pour $l = 2$ et un résultat d'Albert ([A], Chap. XI, Thm. 2; [Ar2], Thm. 5.5) impliquent que tout élément de $H^2(K, \mathbf{Z}/2)$ est la classe d'un produit tensoriel de deux algèbres de quaternions, i.e. est la somme de deux symboles. On a donc $\lambda_2(K) = 2$. Il reste à majorer $\lambda_3(K)$. Partant du résultat de Saltman, par des manipulations algébriques, Hoffmann et Van Geel [HVG] établissent $\lambda_3(K) \leq 4$, puis $u(K) \leq 22$. Par un argument de géométrie arithmétique, Parimala et Suresh montrent $\lambda_3(K) = 1$:

THÉORÈME 5.7 ([PS1]). — Pour K comme ci-dessus, toute classe dans $H^3(K, \mathbf{Z}/2)$ est représentable par un seul symbole $(a) \cup (b) \cup (c)$.

La majoration générale ci-dessus donne alors immédiatement $u(K) \leq 12$. Un travail arithmétique plus fin permet à Parimala et Suresh d'obtenir la borne $u(K) \leq 10$. L'outil fondamental pour la démonstration de 5.7 est le résultat suivant (analogue pour le groupe H^3 du résultat 5.2), qui est un cas particulier d'un théorème de K. Kato :

THÉORÈME 5.8 ([Kt]). — Soient k un corps p -adique avec $p \neq 2$ et K un corps de fonctions d'une variable sur k . Le groupe de cohomologie non ramifié $H_{\text{nr}}^3(K, \mathbf{Z}/2)$ est nul.

Parimala et Suresh partent d'un élément quelconque $\alpha \in H^3(K, \mathbf{Z}/2)$. En considérant un modèle régulier propre de K au-dessus de l'anneau des entiers de k , ils montrent comment trouver un élément $f \in K^*$ tel que α devienne non ramifié dans $H^3(K(\sqrt{f}), \mathbf{Z}/2)$, donc nul dans $H^3(K(\sqrt{f}), \mathbf{Z}/2)$ par le théorème de Kato appliqué à $K(\sqrt{f})$. Le théorème

de Saltman assure alors que α est une somme d'au plus 2 symboles. Un travail arithmétique plus précis montre que α est représenté par un seul symbole. Le théorème 5.7 répond à une question de Serre [S3] : pour K comme ci-dessus et G le K -groupe simple déployé de type G_2 , la flèche naturelle $H^1(K, G) \rightarrow H^3(K, \mathbf{Z}/2)$ est une bijection.

RÉFÉRENCES

- [A] A. A. ALBERT – *Structure of Algebras*, American Mathematical Society Colloquium Publications **24**, American Mathematical Society, New York, 1939.
- [Ar1] M. ARTIN – Algebraic approximation of structures over complete local rings. *Inst. Hautes Études Sci. Publ. Math.* **36** (1969) 23–58.
- [Ar2] M. ARTIN – Brauer-Severi varieties, in *Brauer groups in ring theory and algebraic geometry* (Wilrijk, 1981), *Lecture Notes in Math.* **917**, pp. 194–210, Springer, Berlin-New York, 1982.
- [Ar3] M. ARTIN – Two-dimensional orders of finite representation type, *Manuscripta Math.* **58** (1987), 445–471.
- [AG] M. AUSLANDER et O. GOLDMAN – Maximal orders, *Trans. Amer. Math. Soc.* **97** (1960) 1–24.
- [Bld] A. BLANCHARD – *Les corps non commutatifs*, Presses Universitaires de France, 1972.
- [Blt] A. BLANCHET – Function fields of generalized Brauer-Severi varieties, *Comm. Algebra* **19** (1991) 97–118.
- [BLR] S. BOSCH, W. LÜTKEBOHMERT et M. RAYNAUD – Formal and rigid geometry, IV : The reduced fibre theorem, *Invent. math.* **119** (1995) 361–398.
- [Bki] N. BOURBAKI – *Algèbre*, Chapitre 8, Modules et anneaux semi-simples, Hermann, Paris, 1958.
- [BHN] R. BRAUER, H. HASSE et E. NOETHER, Beweis eines Hauptsatzes in der Theorie der Algebren, *J. reine angew. Math. (Crelle)* **167** (1932) 399–404.
- [CT] J.-L. COLLIOT-THÉLÈNE – Die Brauersche Gruppe ; ihre Verallgemeinerungen und Anwendungen in der Arithmetischen Geometrie (Vortragsnotizen, Brauer Tagung, Stuttgart, 22.-24. März 2001), tapuscrit.
- [CTG] J.-L. COLLIOT-THÉLÈNE – Exposant et indice d'algèbres simples centrales non ramifiées, avec un appendice par O. GABBER, *Enseign. Math.* **48** (2002) 127–146.

- [CTGP] J.-L. COLLIOT-THÉLÈNE, P. GILLE et R. PARIMALA – Arithmetic of linear algebraic groups over 2-dimensional geometric fields, *Duke Math. J.* **121** (2004) 285–341.
- [CTOP] J.-L. COLLIOT-THÉLÈNE, M. OJANGUREN et R. PARIMALA – Quadratic forms over fraction fields of two-dimensional Henselian rings and Brauer groups of related schemes, in *Arithmetic and Geometry*, Part II (Mumbai, 2000) ed. R. Parimala, *Tata Inst. Fund. Res. Stud. Math.* **16**, Tata Inst. Fund. Res., pp. 185–217, Mumbai, 2002.
- [CTS] J.-L. COLLIOT-THÉLÈNE et J.-J. SANSUC, Fibrés quadratiques et composantes connexes réelles, *Math. Annalen* **244** (1979) 105–134.
- [D] M. DEURING – *Algebren*, Zweite, korrigierte Auflage. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **41**, Springer-Verlag, Berlin-New York 1968.
- [FS] T. J. FORD et D. J. SALTMAN – Division algebras over Henselian surfaces, in *Ring Theory* (Ramat Gan and Jerusalem, 1988/1989), ed. L. Rowen, *Israel Math. Conf. Proc.* **1**, pp. 320–336, Weizmann, Jerusalem, 1989.
- [G1] P. GILLE – La R-équivalence sur les groupes algébriques réductifs définis sur un corps global, *Inst. Hautes Études Sci. Publ. Math.* **86** (1997) 199–235.
- [G2] P. GILLE – Cohomologie galoisienne des groupes quasi-déployés sur des corps de dimension cohomologique ≤ 2 , *Compos. Math.* **125** (2001) 283–325.
- [GSz] P. GILLE et T. SZAMUELY – *Central simple algebras and Galois cohomology*, *Cambridge Studies in Advanced Mathematics* **101**, Cambridge University Press (2006).
- [Gi] J. GIRAUD – *Cohomologie non abélienne*, *Grundlehren der mathematischen Wissenschaften* **179**. Springer-Verlag, Berlin-New York, 1971.
- [GHS] T. GRABER, J. HARRIS et J. STARR – Families of rationally connected varieties. *J. Amer. Math. Soc.* **16** (2003) 57–67.
- [Gr] A. GROTHENDIECK – Le groupe de Brauer I, II, III, in *Dix exposés sur la cohomologie des schémas*, Masson et North-Holland, Paris et Amsterdam, 1968.
- [EGA] A. GROTHENDIECK et J. DIEUDONNÉ – Éléments de géométrie algébrique, III : Étude cohomologique des faisceaux cohérents, I, *Inst. Hautes Études Sci. Publ. Math.* **11** (1961).
- [HVG] D. W. HOFFMANN et J. VAN GEEL – Zeros and norm groups of quadratic forms over function fields in one variable over a local non-dyadic field, *J. Ramanujan Math. Soc.* **13** (1998) 85–110.
- [HL] D. HUYBRECHTS et M. LEHN – *The geometry of moduli spaces of sheaves*, *Aspects of Math.* **E31**, Vieweg, Braunschweig, 1997.

- [I] L. ILLUSIE – Grothendieck’s existence theorem in formal geometry (with a letter of J-P. Serre), in *Advanced School in Basic Algebraic Geometry*, ICTP Trieste, 2003, soumis à Contemporary Math.
- [dJ] A. J. de JONG – The period-index problem for the Brauer group of an algebraic surface, *Duke Math. J.* **123** (2004), 71–94.
- [dJ2] A. J. de JONG – A result of Gabber, tapuscrit.
- [dJS1] A. J. DE JONG et J. STARR – Every rationally connected variety over the function field of a curve has a rational point, *Amer. J. Math.* **125** (2003), 567–580.
- [dJS2] A. J. DE JONG et J. STARR – Obtaining reduced fibres after normalized base change, addendum to [dJS1], tapuscrit.
- [Kh] B. KAHN – On « horizontal » invariants attached to quadratic forms, in *Algebra and Number Theory*, Proceedings of the Silver Jubilee Conference, University of Hyderabad, Rajat Tandon ed., pp. 21–33, Hindustan Book Agency, 2005.
- [Kt] K. KATO – A Hasse principle for 2-dimensional global fields, *J. reine angew. Math. (Crelle)* **366** (1986) 142–181.
- [Kr] A. KRESCH – Hodge-theoretic obstruction to the existence of quaternion algebras. *Bull. London Math. Soc.* **35** (2003) 109–116.
- [KRTY] B. È. KUNYAVSKIĬ, L. H. ROWEN, S. V. TIKHONOV et V. I. YANCHEVSKIĬ, Bicyclic algebras of prime exponent over function fields, à paraître dans *Trans. Amer. Math. Soc.*
- [Lie] M. LIEBLICH – Twisted sheaves and the period-index problem, tapuscrit.
- [Li] J. LIPMAN – Introduction to resolution of singularities, *Proc. Symp. Pure Math.* **29** (1975) 187–230.
- [M1] A. S. MERKUR’EV – Sur le symbole de norme résiduelle de degré 2 (en russe), *Dokl. Akad. Nauk SSSR* **261** (1981) 542–547.
- [M2] A. S. MERKUR’EV – Algèbres simples centrales et formes quadratiques (en russe), *Izv. Akad. Nauk SSSR* **55** (1991) 218–224.
- [MS] A. S. MERKUR’EV et A. A. SUSLIN – K -cohomologie des variétés de Severi-Brauer et homomorphisme de norme résiduelle (en russe), *Izv. Akad. Nauk SSSR Ser. Mat.* **46** (1982) 1011–1046, 1135–1136.
- [Mi] J. S. MILNE – *Étale cohomology*, Mathematical Series vol. **33**, Princeton University Press 1980.
- [OP] M. OJANGUREN et R. PARIMALA – Algebras of prime degree on function fields of surfaces, tapuscrit.
- [PS1] R. PARIMALA et V. SURESH – Isotropy of quadratic forms over function fields of p -adic fields, *Inst. Hautes Études Sci. Publ. Math.* **88** (1998) 129–150.

- [PS2] R. PARIMALA et V. SURESH – On the length of a quadratic form, in *Algebra and Number Theory*, Proceedings of the Silver Jubilee Conference, University of Hyderabad, Rajat Tandon ed., pp. 147–157, Hindustan Book Agency, 2005.
- [R] I. REINER – *Maximal orders*, Corrected reprint of the 1975 original. With a foreword by M. J. Taylor. London Mathematical Society Monographs. New Series **28**, The Clarendon Press, Oxford University Press, Oxford, 2003.
- [Ro] P. ROQUETTE – The Brauer-Hasse-Noether theorem in historical perspective. *Schriften der Math.-Phys. Klasse der Heidelberger Akademie der Wissenschaften* **15** (2004).
- [Sa1] D. J. SALTMAN – Division algebras over p -adic curves, *J. Ramanujan Math. Soc.* **12** (1997) 25–47. Corrections : *Zentralblatt für Mathematik*, 0902.16021 et *J. Ramanujan Math. Soc.* **13** 125–129.
- [Sa2] D. J. SALTMAN – Cyclic algebras over p -adic surfaces, tapuscrit.
- [S1] J-P. SERRE – *Corps locaux*, Publications de l’Institut de Mathématique de l’Université de Nancago, VIII. *Actualités Sci. Indust.* **1296**, Hermann, Paris, 1962.
- [S2] J-P. SERRE – *Cohomologie galoisienne*, Springer Lecture Notes **5**, Cinquième édition, Springer-Verlag, Berlin, 1964.
- [S3] J-P. SERRE – Cohomologie galoisienne : progrès et problèmes. Séminaire Bourbaki, Exposé **783**, Vol. 1993/94. *Astérisque* **227** (1995) 229–257.
- [T] J.-P. TIGNOL – Algèbres indécomposables d’exposant premier, *Advances in Math.* **65** (1987) 205–228.
- [TW] J.-P. TIGNOL et A.R. WADSWORTH – Totally ramified valuations on finite-dimensional division algebras, *Trans. Amer. Math. Soc.* **302** (1987) 223–250.

Jean-Louis COLLIOT-THÉLÈNE

Université Paris XI

UMR 8628 du CNRS

Département de Mathématiques

Bâtiment 425

F-91405 ORSAY Cédex

courriel : colliot@math.u-psud.fr

**GÉOMÉTRIE CONFORME EN DIMENSION 4 :
CE QUE L'ANALYSE NOUS APPREND**

par **Christophe MARGERIN**

INTRODUCTION

Si toute variété différentielle admet une structure riemannienne – les métriques sur une variété forment un cône de dimension infinie –, on sait que certaines propriétés algébriques de la courbure de la connexion riemannienne se traduisent dans la topologie sous-jacente : restrictions sur le type homologique (théorèmes d'annulation), homotopique, topologique, voire différentiel. Toute variété riemannienne complète de courbure sectionnelle négative ou nulle est ainsi revêtue par \mathbb{R}^n (théorème de Cartan-Hadamard) et toute variété orientable de dimension paire compacte et de courbure sectionnelle strictement positive est aussi simplement connexe (théorème de Synge). Dans une logique de classification topologique par « géométrisation », on cherche à affaiblir la caractérisation métrique obtenue : s'il est facile de se convaincre qu'une variété – que l'on supposera simplement connexe en dimension impaire – de courbure sectionnelle constante et strictement positive est une sphère, le fait qu'il en aille de même en dimension 3 de toute variété de courbure de Ricci strictement positive – un résultat aujourd'hui classique, dû à R. Hamilton – doit être considéré comme un « vrai » résultat de géométrisation.

Une autre façon d'affaiblir une hypothèse de courbure consiste, au lieu de prendre une trace « algébrique » de l'invariant, comme dans l'exemple précédent où l'on passe de la courbure de Riemann à la courbure de Ricci, à considérer l'hypothèse « en moyenne » sur la variété : grâce à la formule de Gauss-Bonnet on peut en dimension 2 remplacer dans la caractérisation précédente de la sphère le signe de la courbure (de Gauss dans ce cas) par celui de son intégrale (pour la mesure canoniquement associée à la métrique).

Le formalisme de Chern-Weil généralise à la dimension supérieure cette belle formule en donnant des expressions des nombres caractéristiques en termes d'intégrales de traces de polynômes en la courbure, mais les caractérisations de types topologiques ou différentiels en termes de propriétés « en moyenne » de la courbure sont rares. D'où l'intérêt de l'énoncé suivant, particulièrement satisfaisant.

THÉORÈME 1 ([CGY0]). — *Toute variété différentiable de dimension 4, compacte et sans bord, admettant une métrique de courbure scalaire strictement positive et dont la norme L^2*

de la courbure de Weyl et la caractéristique d'Euler sont reliées par la relation

$$\int_M |W|^2 d\text{vol} < 16\pi^2 \chi(M)$$

est difféomorphe à S^4 ou à $\mathbb{P}_{\mathbb{R}}^4$.

Comme nous l'expliquons dans le préliminaire, cet énoncé est une version L^2 de la caractérisation de la sphère standard en terme du pincement faible – un invariant suffisamment « faible » pour autoriser en tout point des courbures sectionnelles négatives – établie dans [M].

La preuve de Chang S.-Y. A., M. Gursky et Yang P. consiste d'ailleurs à réduire leur énoncé à celle-ci en construisant dans la classe conforme d'une métrique vérifiant les hypothèses du théorème 1 une métrique $1/6$ - faiblement pincée, l'hypothèse de [M].

La caractérisation donnée dans [M] est optimale, et on y établit la classification des géométries limites. Ce théorème de rigidité admet lui aussi une version L^2 .

THÉORÈME 2 ([CGY0]). — *Toute variété différentiable de dimension 4, compacte et sans bord, admettant une métrique de courbure scalaire strictement positive et dont la norme L^2 de la courbure de Weyl et la caractéristique d'Euler sont reliées par la relation*

$$\int |W|^2 d\text{vol} \leq 16\pi^2 \chi(M)$$

est soit difféomorphe à S^4 ou à $\mathbb{P}_{\mathbb{R}}^4$, soit conformétement équivalente au plan projectif complexe ($\mathbb{P}_{\mathbb{C}}^2$, F - S) ou à un quotient du produit $\mathbb{R} \times (S^3, \text{can})$ par un sous-groupe d'isométries.

Dans cet exposé nous ne reviendrons pas sur la preuve de l'énoncé « géométrique », [M], obtenue par l'étude de l'invariant « pincement faible » le long des courbes intégrales de la courbure de Ricci considérée comme champ de vecteurs sur l'espace des métriques. Ces idées ont été remises au goût du jour par le travail de G. Perelman en dimension 3 ; elles joueront d'ailleurs un rôle en un point crucial de l'argument, mais sous le mode mineur de flot de Yamabe.

Nous allons plutôt présenter l'ensemble des résultats de géométrie conforme de la dimension 4 qui ont permis aux auteurs de réduire leurs énoncés à ceux établis dans [M] : un travail technique de longue haleine, des premiers papiers sur les métriques extrémales pour les déterminants régularisés – en particulier [CY] – à ceux, plus récents, où ils étudient un analogue du problème de Yamabe pour un invariant scalaire *quadratique* en la courbure de Ricci ([CGY1], [CGY2]), en passant par une étude systématique des « paires conformes » – et en particulier de « l'opérateur de Paneitz » et de sa « Q -courbure ».

1. PRÉLIMINAIRE : UNE PREMIÈRE RÉDUCTION

La courbure riemannienne est un 4-tenseur présentant un certain nombre de symétries qui font qu'elle peut être considérée comme une section du fibré des endomorphismes symétriques de la puissance extérieure seconde du cotangent. Elle vérifie de plus la première identité de Bianchi, qui exprime son orthogonalité à la puissance extérieure quatrième du cotangent. Aux deux (seules) traces de la courbure de Riemann, la courbure de Ricci, $\text{ric} = \text{tr}_{24}R$, et la courbure scalaire, $\text{scal} = \text{tr ric}$, correspondent deux composantes irréductibles de l'algèbre de courbure sous l'action du groupe orthogonal, de dimensions respectives $(n^2 + n - 2)/2$ ($= \dim S^2T^*M - 1$) et 1; la projection sur la première est donnée par $\sigma = 1/2n(n - 1) \text{scal} g \otimes g$, et l'autre par $\rho_0 = 1/(n - 2) \text{ric}_0 \otimes g$ où $\text{ric}_0 = \text{ric} - 1/n \text{scal} g$ représente la partie sans trace de la courbure de Ricci et \otimes une suspension algébrique de S^2T^*M dans $S^2\Lambda^2T^*M$ parfois appelée produit de Kulkarni. Ce qui reste, $W := R - \frac{\text{scal}}{2n(n-1)}g \otimes g - \frac{\text{ric}_0 \otimes g}{n-2}$ est appelé courbure de Weyl et représente la projection de la courbure de Riemann, R , sur la dernière composante irréductible, celle des tenseurs de courbure dont toutes les traces s'annulent. Réduite à 0 en dimension 2 et 3, c'est la plus grande composante (en terme de dimension) dès la dimension 4.

Cette composante admet une décomposition exceptionnelle sous l'action de $SO(n)$ en dimension $n = 4$, associée à l'action de l'opérateur de Hodge et correspondant à la décomposition $\underline{so}(4) = \underline{so}(3) \oplus \underline{so}(3)$. Ce raffinement joue un rôle crucial dans [M], et interviendra ici dans la discussion de la classification des métriques plates au sens de Bach, (cf. le paragraphe 4.2.2, en particulier l'identité (4.27) du Lemme 4.3), par laquelle passe la démonstration du résultat de rigidité énoncé dans le Théorème 2.

On vérifie facilement que la courbure de Weyl est un covariant conforme : $W(e^{2f}g) = e^{2f}W(g)$. L'intégrale $\int_M |W|^2 d\text{vol}$ est en particulier un invariant, et le théorème 1 donne donc une caractérisation *conforme* et *intégrale* de la sphère standard.

Rappelons que la généralisation due à Chern S.S. de la formule de Gauss-Bonnet s'énonçant, en dimension 4,

$$(1.1) \quad 32\pi^2 \chi(M) = \int_M (|\sigma|^2 - |\rho_0|^2 + |W|^2) d\text{vol} ,$$

les hypothèses des théorèmes 1 (2) se lisent donc, respectivement,

$$(1.2) \quad \int_M (|\sigma|^2 - |\rho_0|^2 - |W|^2) d\text{vol} > (\geq) 0.$$

Le pincement faible étant défini par (cf. [M])

$$PF = \sup_M \frac{|R - \sigma|^2}{\text{scal}^2} ,$$

et la décomposition de la courbure rappelée ci-dessus étant orthogonale, l'hypothèse de [M], $PF < (\leq) 1/6$ s'écrit donc

$$|R - \sigma|^2 = |\rho_0|^2 + |W|^2 < (\leq) \frac{\text{scal}^2}{6} = |\sigma|^2 ,$$

ce qui revient à la positivité de l'intégrand dans l'intégrale (1.2) : les théorèmes 1 et 2 sont bien une version « L^2 » de [M], et il suffira pour les établir de démontrer l'existence d'une métrique dont le polynôme de courbure $|\sigma|^2 - |\rho_0|^2 - |W|^2 = \frac{\text{scal}^2}{6} - 2|\text{ric}_0|^2 - |W|^2$ est partout (strictement) positif dans la classe conforme de toute métrique de courbure scalaire strictement positive satisfaisant l'hypothèse intégrale (1.2).

Il existe une combinaison de la courbure de Ricci et de la courbure scalaire particulièrement pertinente en dimension 4, $A := \text{ric} - \frac{\text{scal}}{6}g = \text{ric}_0 + \frac{\text{scal}}{12}g$, appelée *courbure de Schouten*, et en terme de laquelle la décomposition précédente du tenseur de Riemann s'écrit $R = \frac{1}{2}A \otimes g + W$. Exprimé avec la courbure de Schouten, la positivité du polynôme $|\sigma|^2 - |\rho_0|^2 - |W|^2$ est encore équivalente à celle du polynôme $4\sigma_2(A) - |W|^2$, où $\sigma_2(A)$ représente la seconde fonction symétrique élémentaire de l'endomorphisme symétrique A . Notons qu'en terme de cette fonction la formule de Chern-Gauss-Bonnet (1.1) admet la forme simple suivante

$$(1.3) \quad 32\pi^2 \mathcal{X}(M) = \int_M (4\sigma_2(A) + |W|^2) d\text{vol} ,$$

qui nous permet de déduire de l'invariance conforme de $\int_M |W|^2 d\text{vol}$ celle de l'intégrale $\int_M \sigma_2(A) d\text{vol}$, bien que $\sigma_2(A)$ ne soit pas lui-même un covariant conforme. Cette remarque élémentaire est essentielle à la compréhension de la stratégie adoptée.

Dans une première partie, (le chapitre 2), nous établissons la réduction annoncée dans le cas où la norme L^2 de la courbure de Weyl est *strictement* majorée par $4\pi \sqrt{\mathcal{X}(M)}$ – et donc le théorème 1. Nous commençons par une preuve dans l'esprit de celle proposée par les auteurs, qui a le mérite de préciser comment des préoccupations essentiellement analytiques, comme l'étude de courbures du *quatrième* degré, objets insolites de la géométrie « classique », les y ont conduits.

Dans une seconde partie, (le chapitre 3), nous en donnons une preuve inédite, plus naturelle, plus simple et nettement plus rapide qui s'affranchit du recours à des opérateurs différentiels du quatrième degré ; elle s'inspire très largement des travaux de M. Gursky et J. Viaclovsky (en particulier de [GV]) sur la seconde fonction symétrique du tenseur de Schouten, $\sigma_2(A)$. À l'issue nous discutons un corollaire intéressant de ces travaux qui donne un critère très général d'existence de métriques de Q -courbure *constante*, et explicitons quelques familles d'exemples.

Dans une dernière partie, (le chapitre 4), nous abordons l'étude du cas limite où la norme L^2 de la courbure conforme est égale à $4\pi \sqrt{\mathcal{X}(M)}$ et où les deux preuves précédentes de la réduction proposée s'effondrent par dégénérescence de l'ellipticité des équations considérées. En suivant [CGY0] nous commençons par résoudre un « problème du type Yamabe » pour les invariants quadratiques $|\sigma|^2 - |\rho_0|^2 - \alpha|W|^2 = 4\sigma_2(A) - \alpha|W|^2$, $\alpha < 1$, avant de construire la solution de l'équation $|\sigma|^2 - |\rho_0|^2 - |W|^2 \equiv 0$ comme limite en $\alpha = 1$ – un argument délicat qui passe par la classification des variétés plates, au sens de Bach, dont la norme L^2 de la courbure de Weyl est égale à $4\pi \sqrt{\mathcal{X}(M)}$.

2. LE CAS $\int(4\sigma_2(A) - |W|^2) d\text{vol} > 0$: DÉMONSTRATION DU THÉORÈME 1

2.1. Déterminants régularisés et paires conformes

Le point de départ est une étude variationnelle plus ou moins systématique du déterminant régularisé d'opérateurs différentiels intrinsèques *conformément covariants* de poids (a, b) , $(a, b) \in \mathbb{R}^2$, c'est-à-dire tels que $L_{e^{2f}g} = e^{-bf} L e^{af}$ pour toute fonction infiniment différentiable f . Si L est un opérateur différentiel intrinsèque de degré d sur une variété compacte sans bord de dimension n , qui est formellement auto-adjoint et de symbole principal défini positif, le spectre de cet opérateur L est réel, discret, minoré et tend vers l'infini comme $\lambda_i \underset{i \rightarrow +\infty}{\sim} C i^{d/n}$, $i \in \mathbb{N}$. On introduit alors classiquement la fonction ζ spectrale de L , $\zeta_L(s) = \sum_{\lambda_j \neq 0} |\lambda_j|^{-s}$, définie pour les points s de \mathbb{C} dont la partie réelle est assez grande, et son extension méromorphe, à pôles simples isolés, que l'on obtient par prolongement analytique. On appelle *déterminant régularisé*, et l'on note $\det L$, la valeur $e^{-\zeta'_L(0)}$.

En supposant de plus l'opérateur L conformément covariant et homogène – i.e. satisfaisant $L_{e^{2c}g} = e^{-dc} L_g$ pour tout réel c –, et en s'appuyant sur l'asymptotique en temps petit de la trace du noyau de la chaleur, T. Branson et B. Ørsted, ([B-Ø], §2, en particulier (2.7)), explicitent l'expression de la variation conforme du déterminant régularisé d'un tel opérateur sur une variété compacte sans bord de dimension 4 en termes de la géométrie et d'opérateurs «classiques», obtenant ainsi une généralisation de la formule dérivée par Polyakov dans le cas particulier du laplacien sur une surface de Riemann. En posant $F(f) := \log(\det L_{e^{2f}g} / \det L_g)$ – au sens régularisé précédent – ils établissent que F se décompose en une combinaison linéaire de trois fonctionnelles universelles $(I_i)_{i=1}^3$, la dépendance en l'opérateur L n'affectant que les coefficients $(\gamma_i)_{i=1}^3$ de la décomposition $F = \sum_{i=1}^3 \gamma_i I_i$.

Dans cette décomposition $I_1(f)$ représente une moyenne normalisée de f , relative à la densité $|W|^2$:

$$I_1(f) = 4 \int_M |W|^2 f d\text{vol} - \int_M |W|^2 d\text{vol} \log \int e^{4f} d\text{vol}.$$

L'expression de la fonctionnelle I_2 implique l'opérateur introduit par Paneitz pour étudier l'interaction entre le groupe conforme et le groupe de jauge des équations de Maxwell, que nous représenterons par la lettre P , et la « Q -courbure» qui lui est associée : en posant

$$Q := \frac{1}{4} \left(\frac{(n-2)(n+2)}{4(n-1)^2 n} \text{scal}^2 - 4 \frac{|\text{ric}_0|^2}{(n-2)^2} \right) + \frac{\Delta \text{scal}}{4(n-1)},$$

où d représente la différentielle extérieure, d^* son adjoint formel L^2 , et $\Delta = d^*d (+dd^*)$ le laplacien riemannien (agissant ici sur les fonctions), P est l'opérateur différentiel d'ordre 4 d'expression

$$P := \Delta^2 + d^* \left(-\frac{4\text{ric}_0}{n-2} + \frac{(n^2 - 2n - 4)}{2n(n-1)} \text{scal } g \right) d + (n-4) Q.$$

Il vérifie la relation de covariance conforme

$$P(e^{2f}g) = e^{-\frac{(n+4)}{2}f} P(g) e^{\frac{(n-4)}{2}f}.$$

En faisant opérer les deux membres de cette identité sur la fonction constante égale à 1, un élément du noyau de $P_0 := P - (n-4)Q$, nous trouvons

$$(n-4)Q(e^{2f}g) = e^{-\frac{(n+4)}{2}f} P_0(g) (e^{\frac{(n-4)}{2}f} - 1) + (n-4)e^{-4f}Q(g);$$

en nous autorisant à prolonger formellement cette expression à $n \in \mathbb{R}$, à la dériver par rapport à la variable n et à spécialiser en $n = 4$, c'est-à-dire en considérant la limite en $n = 4$ du quotient par $(n-4)$ de cette expression, nous dérivons l'importante relation suivante (de la dimension 4, donc)

$$(2.1) \quad Q(e^{2f}g) = e^{-4f}(Q(g) + \frac{1}{2}P(g)f).$$

Cette identité peut évidemment être vérifiée de façon conventionnelle au prix d'un calcul plus laborieux. Un tel couple (P, Q) est appelé « *paire conforme* », et Q , la « *Q-courbure* » associée à l'opérateur covariant conforme P . Un exemple plus élémentaire de telle paire est fourni en dimension 2 par le couple (Δ, κ) , où κ représente la courbure de Gauss : la variation de la courbure dans une classe conforme est en effet régie par l'équation $\kappa(e^{2f}g) = e^{-2f}(\kappa(g) + \Delta f)$, qui joue le rôle de (2.1) dans ce cas.

La construction systématique de telles paires en toute dimension, leur calcul explicite, et l'étude de leurs propriétés est un sujet en pleine expansion (voir par exemple [F-G1], [F-G2], [F-H], [G-J-M-S], [Be], [G-P], [G-Z]... , et la dernière partie du chapitre 3 où nous présentons une construction systématique de métriques de Q-courbure constante).

En dimension 4, les expressions précédentes se spécialisent en

$$(2.2) \quad P = \Delta^2 + d^*(-2\text{ric} + \frac{2}{3}\text{scal}g)d,$$

et

$$(2.3) \quad Q = \frac{1}{12}(\Delta \text{scal} - 3|\text{ric}|^2 + \text{scal}^2) = \frac{1}{12}(\Delta \text{scal} - 3|\text{ric}_0|^2 + \frac{1}{4}\text{scal}^2),$$

en termes desquelles nous pouvons maintenant exprimer le second générateur de [B-Ø] :

$$(2.4) \quad I_2(f) := \int_M fPf \, d\text{vol} + 4 \int_M Qf \, d\text{vol} - \int_M Q \, d\text{vol} \log \int e^{4f} \, d\text{vol}.$$

Le gradient de I_2 étant alors clairement donné par la formule

$$(2.5) \quad \nabla I_2(f) = 2Pf + 4Q - 4e^{4f} \int_M Q \, d\text{vol} / \int_M e^{4f} \, d\text{vol},$$

et l'équation (2.1) exprimant la Q-courbure dans une classe conforme s'écrivant en un point critique f

$$(2.6) \quad Q(e^{2f}g) = e^{-4f}(Q + \frac{1}{2}Pf) = \int_M Q \, d\text{vol} / \int_M e^{4f} \, d\text{vol},$$

les points critiques de la fonctionnelle I_2 sont donc les métriques de *Q-courbure constante*.

Le dernier générateur, I_3 , est une variante quadratique de la fonctionnelle de Yamabe :

$$I_3(f) := 12 Y(f) + 4 \int f \Delta \text{scal} \, d\text{vol} ,$$

où

$$Y(f) := \int_M (\Delta f - |df|^2)^2 \, d\text{vol} - \frac{1}{3} \int_M \text{scal} |df|^2 \, d\text{vol} .$$

En invoquant la formule régissant l'expression de la courbure scalaire dans une classe conforme

$$(2.7) \quad \text{scal}(e^{2f}g) = e^{-2f}(\text{scal} + 6(\Delta f - |df|^2)) ,$$

on obtient pour I_3 l'expression plus transparente

$$I_3(f) = \frac{1}{3} \int_{(M, e^{2f}g)} \text{scal}^2 \, d\text{vol} - \frac{1}{3} \int_{(M, g)} \text{scal}^2 \, d\text{vol} ,$$

qui justifie l'appellation « fonctionnelle de Yamabe quadratique ». On peut se convaincre que les points critiques de I_3 sont les métriques de courbure scalaire constante en calculant le gradient, égal en f à : $4 e^{4f}(\Delta \text{scal})(e^{2f}g)$.

Le travail initial de Chang S.-Y. A. et Yang P. consiste en une extension à la dimension 4 du théorème de compacité de B. Osgood, R. Philipps et P. Sarnak (cf. [OPS1], et [OPS2]) qui établit, en s'appuyant sur l'inégalité de Moser-Trudinger, que le maximum du logarithme du déterminant régularisé du laplacien sur une surface de Riemann est atteint en « la » métrique de courbure constante. Dans [CY], Chang S.-Y. A. et Yang P. démontrent que sur une variété compacte sans bord de dimension 4 la fonctionnelle $F = \sum_{i=1}^3 \gamma_i I_i$ atteint son minimum dès que γ_2 et γ_3 sont strictement positifs et que $\kappa := \gamma_1 \int_M |W|^2 \, d\text{vol} + \gamma_2 \int_M Q \, d\text{vol}$ est strictement majoré par $8\gamma_2 \pi^2$. (On notera que κ est un invariant conforme puisque, d'après (2.3), $\int_M Q \, d\text{vol} = \frac{1}{2} \int_M \sigma_2(A) \, d\text{vol}$.)

Reportons pour l'instant la discussion de la preuve de cet énoncé, de toute façon encore insuffisant pour notre objectif, et notons qu'en regroupant les expressions des gradients, ∇I_i , calculés précédemment nous obtenons pour celui de la fonctionnelle F évalué en une fonction f

$$4 e^{4f}(\gamma_1 |W|^2 + \gamma_2 Q + \gamma_3 \Delta \text{scal} - \frac{\kappa}{\text{vol}(e^{2f}g)}) .$$

Il reste à choisir les poids γ_i de façon à annuler κ , soit $\gamma_2 = 1$, et

$$(2.8) \quad \gamma_1 = - \int_M Q \, d\text{vol} / \int_M |W|^2 \, d\text{vol} = -\frac{1}{2} \int_M \sigma_2(A) \, d\text{vol} / \int_M |W|^2 \, d\text{vol} ,$$

et à introduire $\delta := 8\gamma_3 + 2/3$, pour que l'équation d'Euler de la fonctionnelle F devienne

$$0 = \gamma_1 |W|^2 + Q + \frac{1}{24}(3\delta - 2) \Delta \text{scal} = \gamma_1 |W|^2 + \frac{1}{2} \sigma_2(A) + \frac{1}{8} \delta \Delta \text{scal} .$$

Nous tenons ici la clé de l'approche de Chang S.-Y. A. et Yang P. ; imaginons que nous souhaitions, sous l'hypothèse $\int_M \sigma_2(A) \, d\text{vol} > 0$, établir l'existence d'une métrique satisfaisant

$\sigma_2(A) > 0$ (ce qui est l’objet de l’article [CGY1]) : tout minimum de la fonctionnelle F est une solution de l’équation d’Euler,

$$(2.9) \quad \sigma_2(A) = -\delta/4 \Delta \text{scal} - 2\gamma_1 |W|^2 ,$$

qui pourra être utilisée comme *régularisation* de l’opérateur géométrique pertinent, $\sigma_2(A)$.

2.2. Un premier résultat d’existence

2.2.1. *L’énoncé.* — Trois problèmes se posent ici :

1) le théorème d’existence de [CY] requiert la positivité de γ_3 et ne s’applique donc que dans le cas où $\delta > 2/3$;

2) en supposant que l’on réussisse à résoudre l’équation régularisée $(2.9)_\delta$ pour toutes les valeurs strictement positives de δ et à passer à la limite en zéro, l’hypothèse $\gamma_1 = -\frac{1}{2} \int_M \sigma_2(A) d\text{vol} / \int_M |W|^2 d\text{vol} < 0$ entraînerait $\sigma_2(A) = -2\gamma_1 |W|^2 \geq 0$, mais pas nécessairement la positivité *stricte* de $\sigma_2(A)$;

3) c’est la positivité de $\sigma_2(A) - |W|^2/4 = 1/4 (|\sigma|^2 - |\rho_0|^2 - |W|^2)$ – et non celle de $\sigma_2(A)$, objet de [CGY0] – qui correspond à la propriété $PF(g) < 1/6$ que nous prétendons établir (cf. la discussion du préliminaire).

Concernant le second point il suffit, pour garantir la positivité *stricte* de $\sigma_2(A)$, de translater l’équation d’Euler par le carré de la norme d’une 2-forme symétrique ne s’annulant nulle part et fixée une fois pour toute, η : pour cela nous modifions la fonctionnelle F en lui ajoutant un multiple de la fonctionnelle

$$I_0(f) := 4 \int_M |\eta|^2 f d\text{vol} - \left(\int_M |\eta|^2 d\text{vol} \right) \log \int_M e^{4f} d\text{vol} .$$

Le paramètre γ_0 que nous introduisons ainsi, poids du générateur I_0 dans l’expression de la nouvelle fonctionnelle, permet de plus de relâcher la contrainte (2.8) et de régler ainsi la troisième difficulté relevée ci-dessus : nous avons de fait la généralisation suivante du résultat de compacité de [CY].

THÉORÈME 2.1. — *Sur une variété riemannienne compacte sans bord de dimension 4 le minimum de la fonctionnelle $H = \sum_{i=0}^3 \gamma_i I_i(f)$ est atteint par une fonction infiniment différentiable, f , dès que γ_2 et γ_3 sont strictement positifs et que $\kappa' = \kappa + \gamma_0 \int_M |\eta|^2 d\text{vol}$ est strictement majoré par $8\pi^2 \gamma_2$. La métrique $e^{2f} g$ vérifie de plus l’identité*

$$\gamma_0 |\eta|^2 + \gamma_1 |W|^2 + \gamma_2 Q + \gamma_3 \Delta \text{scal} = \frac{\kappa'}{\text{vol}(e^{2f} g)} .$$

Cet énoncé regroupe un résultat de *compacité*, l’existence d’un minimum f dans $L^{2,2}(M)$ – l’espace de Sobolev des fonctions qui sont, ainsi que leurs dérivées premières et secondes, de carré intégrables –, et un résultat de *régularité*, le fait que ce minimum soit effectivement infiniment différentiable.

2.2.2. *Compacité.* — Puisque $H(0) = 0$, le minimum $\lim_{\ell} H(f_{\ell})$, où $(f_{\ell})_{\ell \in \mathbb{N}}$ représente une suite minimisante de H , est majoré par 0. Pour dériver une borne a priori sur la norme $L^{2,2}$ d'une telle suite, nous commençons par minorer $H(f_{\ell})$: l'outil déterminant est une version fine de l'inégalité de Moser-Trudinger due à D. Adams (cf. [A]) d'après laquelle

$$(2.10) \quad \log \int e^{4(f-\bar{f})} d\text{vol} \leq \frac{1}{8\pi^2} \int (\Delta f)^2 d\text{vol} + C \quad ;$$

cette majoration est uniforme sur $L^{2,2}(M)$, la constante C ne dépendant que de (M, g) . En posant $\mathcal{E}(g) := \gamma_0 |\eta|^2 + \gamma_1 |W|^2 + \gamma_2 Q + \gamma_3 \Delta \text{scal}$, nous pouvons écrire

$$H(f) = \int_M (4(f - \bar{f})\mathcal{E} + \gamma_2 f P f) d\text{vol} + 12 \gamma_3 Y(f) - \kappa' \log \int_M e^{4(f-\bar{f})} d\text{vol},$$

où $\kappa' = \int_M \mathcal{E} d\text{vol}$. Par convexité de l'application exponentielle $\int_M e^{4(f-\bar{f})} d\text{vol} \geq 1$, et nous obtenons, dans le cas où $\kappa' \leq 0$, la minoration banale

$$H(f) \geq \int_M (4(f - \bar{f})\mathcal{E} + \gamma_2 f P f) d\text{vol} + 12 \gamma_3 Y(f),$$

tandis que, dans le cas où $\kappa' \geq 0$, nous invoquons l'inégalité d'Adams (2.10) pour obtenir alors :

$$H(f) \geq -C\kappa' - \frac{\kappa'}{8\pi^2} \int_M (\Delta f)^2 d\text{vol} + \int_M (4(f - \bar{f})\mathcal{E} + \gamma_2 f P f) d\text{vol} + 12 \gamma_3 Y(f).$$

Dans tous les cas, et en posant $\kappa'' = \max(0, \kappa')$, pour tout $\varepsilon > 0$, et tout entier ℓ assez grand, $\ell \geq \ell(\varepsilon)$, nous avons établi la minoration

$$\begin{aligned} \varepsilon > H(f_{\ell}) &\geq -C\kappa'' + \left(-\frac{\kappa''}{8\pi^2} + \gamma_2 + 12 \gamma_3\right) \int_M |\Delta f_{\ell}|^2 d\text{vol} - 2\gamma_2 \int_M \text{ric}(df_{\ell}, df_{\ell}) d\text{vol} \\ &+ \left(\frac{2}{3} \gamma_2 - 4 \gamma_3\right) \int_M |df_{\ell}|^2 \text{scal} d\text{vol} + 12 \gamma_3 \left(\int_M |df_{\ell}|^4 d\text{vol} - 2 \int_M \Delta f_{\ell} |df_{\ell}|^2 d\text{vol}\right) + 4 \int_M (f_{\ell} - \bar{f}_{\ell}) \mathcal{E} d\text{vol}. \end{aligned}$$

Il reste à minorer $2 \int_M \Delta f_{\ell} |df_{\ell}|^2 d\text{vol}$ par $x \int_M |\Delta f_{\ell}|^2 d\text{vol} + \frac{1}{x} \int_M |df_{\ell}|^4 d\text{vol}$, $x \in \mathbb{R}_+^*$, à invoquer l'inégalité de Poincaré pour majorer $\int_M (f_{\ell} - \bar{f}_{\ell}) \mathcal{E} d\text{vol}$ par $C(g) \sqrt{\int_M |\mathcal{E}|^2 d\text{vol}} \sqrt{\int_M |df_{\ell}|^2 d\text{vol}}$ et à conclure à l'existence d'un majorant uniforme des normes $\int_M ((\Delta f_{\ell})^2 + |df_{\ell}|^4) d\text{vol}$ ne dépendant que de la variété riemannienne (M, g) , et des paramètres γ_2, γ_3 et κ' .

La fonctionnelle H étant clairement invariante par translation par une constante – c'est-à-dire par homothétie sur les métriques –, on supposera sans restriction la suite minimisante f_{ℓ} normalisée par la condition $\int e^{4f_{\ell}} d\text{vol} = 1$, ce qui entraîne que la moyenne \bar{f}_{ℓ} est négative ou nulle. Un corollaire immédiat de l'inégalité d'Adams (2.10) et de la majoration uniforme de $\int (\Delta f_{\ell})^2 d\text{vol}$ que nous venons d'établir est l'existence d'un majorant uniforme de la suite $-\bar{f}_{\ell}$; la compacité faible de $L^{2,2}$ fournit alors le minimum recherché.

2.2.3. *Régularité.* — L'étude de la régularité des minima des fonctionnelles F dont l'existence est établie par [CY] est précisément l'objet de [CGY3]. Le résultat, qui s'énonce comme suit, s'applique aussi aux minima de H .

THÉORÈME 2.2 ([CGY3]). — Soient (M, g) une variété compacte sans bord de dimension 4, a' et a'' deux réels, φ une fonction (de la variable réelle) à croissance au plus exponentielle, $|\varphi(x)| \leq a_1 \exp a_2|x|$, $(a_2, a_2) \in \mathbb{R}^2$, ainsi que sa première dérivée, et b une forme bilinéaire symétrique bornée, $b \in S^2T^*M$, $|b(x, x)| \leq a_3|x|^2$, $a_3 \in \mathbb{R}$. Tout minimum d'une fonctionnelle du type $\mathcal{F}(f) := \int (\Delta f)^2 + (a' \Delta f + a'' |df|^2)^2 + b(df, df) + \varphi(f - \bar{f}) d\text{vol}$ est nécessairement infiniment différentiable.

L'originalité du résultat tient au caractère de la non-linéarité de l'équation d'Euler associée.

2.2.3.1. *Régularité höldérienne.* — La continuité höldérienne d'un minimum est établie en utilisant un critère de Morrey : sur une variété de dimension 4, une fonction est β -höldérienne dès que l'intégrale de son gradient sur toute boule de rayon r est majorée par un multiple uniforme de $r^{3+\beta}$; par l'inégalité de Hölder il suffit donc de majorer la norme L^4 du gradient sur toute boule de rayon r par un multiple uniforme de r^β , ce qu'on établit en démontrant que la fonction $d_{f,P}(r) := \int_{B(P,r)} (|\text{Hess}f|^2 + |df|^4 + f^2 + |df|^2/d(P, \cdot)^2) d\text{vol}$ satisfait, pour r petit, l'inéquation différentielle

$$d - C_1 r d' \leq C_2 r^\gamma$$

où γ est un réel strictement positif et β et C_1 sont reliés par l'identité $\beta = \frac{1}{4C_1}$.

C'est ici qu'intervient l'hypothèse de minimisation : pour une suite minimisante (f_ℓ) de la fonctionnelle H , nous avons expliqué comment obtenir une borne uniforme sur $\int_M ((\Delta f_\ell)^2 + |df_\ell|^4) d\text{vol}$; en procédant de même avec le prolongement biharmonique, h , de f à l'intérieur de la boule $B(P, r)$ – défini par les identités $h \equiv f$ sur $M \setminus B(P, r)$, $\Delta^2 h \equiv 0$ sur $B(P, r)$, $\frac{\partial h}{\partial n} = \frac{\partial f}{\partial n}$ et $h = f$ sur $\partial B(P, r)$ –, nous déduisons de la majoration $H(h) \geq H(f)$, (f est un minimum de H par hypothèse) :

$$\int_{B(P,r)} (|\Delta f|^2 + |df|^4) d\text{vol} \leq C \int_{B(P,r)} (|\Delta h|^2 + |dh|^4) d\text{vol} + C r^\gamma, \quad \gamma > 0.$$

Pour conclure, il reste à majorer le terme de droite par $C_1 r d' + C_2 r^\gamma$, où $d = d_{f,P}(r)$ est la fonction introduite ci-dessus : il s'agit de contrôler la norme $\int_{B(P,r)} (|\Delta h|^2 + |dh|^4) d\text{vol}$ à partir de données au bord, sur lequel h coïncide avec f à l'ordre 1. C'est là une partie substantielle de [CGY3] (en particulier §2, « Preliminary estimates for biharmonic functions ») qui utilise une représentation – établie dans [CQ] – des dérivées troisièmes le long du bord comme image par un opérateur pseudo-différentiel des seules données $h|_{\partial B(P,r)}$ et $\frac{\partial h}{\partial n}|_{\partial B(P,r)}$.

2.2.3.2. *Infinie régularité.* — Pour passer de la régularité höldérienne d'un minimum à sa différentiabilité à tous les ordres, les auteurs adaptent les arguments développés par R. Schoen ([S]), et K. Uhlenbeck ([SU]) dans le cadre des applications harmoniques :

PROPOSITION 2.3. — *Pour toute solution $L^{2,2}$, f , de l'équation d'Euler de \mathcal{F} satisfaisant, pour tout r positif, la majoration $d_{f,P}(r) < Kr^{4\beta}$, $0 < \beta < 8$, la fonction*

$$D_{f,P}(r) := \frac{1}{r^4} \int_{B(P,r)} (r^{2-\beta/4} (\Delta f)^2 + |df|^2 + 1) \, d\text{vol}$$

vérifie la majoration

$$D_{f,P}\left(\frac{r}{2}\right) \leq (1 + C r^{\beta/8}) D_{f,P}\left(2^{\frac{8-\beta}{8+\beta}} r\right).$$

L'hypothèse est automatiquement satisfaite par un minimum de \mathcal{F} d'après le paragraphe précédent. Par itération de la proposition 2.3 nous établissons alors facilement que la fonction $D_{f,P}$ est majorée, uniformément en r et en P . Le critère de Morrey établit ensuite que la norme sup de $|df|$ est bornée et un argument d'*amorçage* (« bootstrap » outremanche) relativement aisé permet d'en déduire une borne uniforme pour le hessien de f . L'équation d'Euler s'écrivant $\Delta^2 f = E(f, \nabla f, \text{Hess} f)$, où les coefficients de E sont des fonctions infiniment différentiables, la théorie de la régularité elliptique « classique » nous enseigne finalement que la fonction f est infiniment différentiable, ce qui conclut la preuve du théorème 2.2.

2.3. Résolution de la régularisation

Reprenons la stratégie exposée avant l'énoncé du théorème 2.1 et l'ayant motivé. Si cet énoncé répond aux points 2) et 3), il n'établit cependant l'existence et la régularité d'une solution de la régularisation $(\gamma_0 |\eta|^2 + \gamma_1 |W|^2 + \gamma_2 Q + \gamma_3 \Delta \text{scal}) \text{vol}(e^{2f} g) = \kappa'$ que dans le cas où les paramètres γ_2 et γ_3 sont strictement positifs. En posant $\gamma_1 = -\alpha/8$, $\gamma_2 = 1$ et $\gamma_3 = \delta/8 - 1/12$, l'annulation de κ' nous dicte la valeur de γ_0 ,

$$(2.11) \quad \gamma_0 = -1/2 \int_M (\sigma_2(A) - \alpha/4 |W|^2) \, d\text{vol} / \int_M |\eta|^2 \, d\text{vol},$$

le théorème 2.1 établissant alors, pour tout $\delta > 2/3$, l'existence d'une solution infiniment différentiable de l'équation δ -régularisée

$$(2.12) \quad \sigma_2(A) - \frac{\alpha}{4} |W|^2 + \frac{\delta}{4} \Delta \text{scal} = -\frac{\alpha}{4} |W|^2 + 2 Q + 2 \gamma_3 \Delta \text{scal} = -2 \gamma_0 |\eta|^2.$$

Dans le but de considérer ensuite la limite en $\delta = 0$, nous commençons par discuter l'existence de solutions de cette équation pour des valeurs de δ arbitrairement petites. En suivant [CGY1], qui traite du cas particulier $\alpha = 0$, nous introduisons pour tout réel strictement positif δ_0 , l'ensemble $S := S_{\delta_0} := \{\delta \in [\delta_0, 1], \text{ tel que l'équation (2.12)}_\delta \text{ admet une solution de courbure scalaire strictement positive}\}$.

2.3.1. *S n'est pas vide.* — Pour $\delta > 2/3$ nous disposons d'une solution de l'équation $(2.12)_\delta$: sous l'hypothèse que α et $\int_M (\sigma_2 - \frac{\alpha}{4} |W|^2) \, d\text{vol}$ sont positifs, $\delta \Delta \text{scal} + 1/6 \text{scal}^2$ l'est aussi. On démontre facilement à l'aide du principe du maximum (voir par exemple [G1]), que cette inéquation différentielle entraîne, pour $\delta = 1$, que la courbure scalaire est strictement positive pourvu que l'invariant de Yamabe, $\mu(g)$, soit positif. Cette condition est ici assurée par l'existence d'une métrique de courbure scalaire strictement positive dans la classe conforme.

2.3.2. *S est ouvert.* — Pour démontrer l’ouverture de l’ensemble S sous l’hypothèse $\int_M (\sigma_2(A) - \frac{\alpha}{4} |W|^2) d\text{vol} > 0$, on s’appuie sur le théorème d’Agmon, Douglis et Nirenberg qui garantit l’existence d’une (essentiellement) unique solution infiniment différentiable de l’équation $(2.12)_\delta$ pour des valeurs de δ suffisamment proches d’un point δ_1 de S dès que le noyau de la linéarisation \mathcal{L}_{δ_1} de l’équation $(2.12)_{\delta_1}$ en la solution f_{δ_1} est réduit aux constantes. Cette propriété découle ici de la majoration

$$\int_M (\mathcal{L}_{\delta_1} \varphi, \varphi) d\text{vol} \geq \frac{3}{13} \delta^2 \int_M |\Delta \varphi|^2 d\text{vol} + \frac{7}{64} \delta \int_M \text{scal} |d\varphi|^2 d\text{vol},$$

que l’on dérive assez facilement de l’explicitation de la linéarisation \mathcal{L}_{δ_1} sous l’hypothèse que l’intégrale $\int_M (\sigma^2(A) - \frac{\alpha}{4} |W|^2) d\text{vol}$ est strictement positive. On peut d’ailleurs interpréter cette minoration comme une généralisation d’estimations spectrales pour l’opérateur de Paneitz dérivées antérieurement par M. Gursky (cf. [G2], et, pour aller au-delà, la discussion de la partie 3.4).

2.3.3. *S est fermé.* — De l’équation $(2.12)_\delta$ on déduit sans trop de difficulté l’estimation a priori $\int_M (\delta |\Delta f_\delta|^2 + |df_\delta|^4) d\text{vol} \leq C_0$, valable pour toute solution f_δ de $(2.12)_\delta$ sous la normalisation $\int_M f_\delta d\text{vol} = 0$. C’est suffisant pour assurer la compacité faible d’une suite de solutions $f_{\delta_k}, \delta_k \in S, \delta_k \xrightarrow{k \rightarrow +\infty} \delta$, et donc l’existence d’une solution faible, f_δ , dans $L^{2,2}(M)$.

Pour ce qui est de la régularité de f_δ , nous pourrions répéter l’argument utilisé pour établir celle de $f_\delta, \delta > 2/3$, dès que nous disposerons de la majoration

$$\int_{B(P,r)} (|\Delta f|^2 + |df|^4) d\text{vol} \leq C_1 \int_{B(P,r)} (|\Delta h|^2 + |dh|^4) d\text{vol} + C_2 r^2,$$

où h représente comme précédemment le prolongement biharmonique de f à la boule $B(P, r)$, et où les constantes $C_i, i \in \{1, 2\}$, ne dépendent que de (M, g) .

Cette estimation peut être déduite de la contraction de l’équation $(2.12)_\delta$ contre la différence $f - h$, en procédant aux majorations idoines dans l’expression obtenue ; on trouvera les détails dans la quatrième partie de [CGY1], en particulier dans la preuve du *Lemma 4.4*.

Il reste, pour conclure, à établir que la courbure scalaire de $e^{2f_\delta} g$ est strictement positive. Puisque $\delta < 1$, on ne peut plus invoquer le principe du maximum de M. Gursky pour le laplacien conforme, utilisé en 2.3.1 pour le cas $\delta = 1$. On vérifie néanmoins facilement que $\text{scal}(e^{2f_\delta} g)$ est positif ou nul : la fonction f_δ étant par construction une limite faible au sens $L^{2,2}$ de fonctions $f_{\delta_k}, \delta_k \in S$, et la courbure scalaire dans une classe conforme étant donnée par l’identité (cf. (2.7))

$$(2.13) \quad -\Delta_g f_{\delta_k} + |df_{\delta_k}|_g^2 + \frac{1}{6} \text{scal}(e^{2f_{\delta_k}} g) e^{2f_{\delta_k}} = \frac{1}{6} \text{scal}(g),$$

la limite f_δ satisfait l’inéquation différentielle : $-\Delta f_\delta + |df_\delta|^2 \leq 1/6 \text{scal}(g)$; nous en déduisons la positivité annoncée : $\text{scal}(e^{2f_\delta} g) = e^{-2f_\delta} (\text{scal}(g) + 6 \Delta f_\delta - 6 |df_\delta|^2) \geq 0$. La limite f_δ satisfaisant par ailleurs l’équation $(2.12)_\delta$, notons la minoration $\delta \Delta \text{scal} = -8 \gamma_0 |\eta|^2 + \alpha |W|^2 + 2 |\text{ric}_0|^2 - \text{scal}^2/6 > -\text{scal}^2/6$; le principe du maximum (ordinaire) appliqué à cette

inéquation différentielle démontre alors que la courbure scalaire de la métrique limite, $e^{2f_\delta}g$, est partout strictement positive sur M .

2.4. L'équation : $\sigma_2(A) - \frac{\alpha}{4}|W|^2 + \gamma_0|\eta|^2 = 0$

Sous l'hypothèse $\int_M(\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} > 0$, nous disposons à ce point de solutions infiniment différentiables de l'équation $(2.12)_\delta$ pour tout δ strictement positif. Nous cherchons dans cette partie à résoudre l'équation $(2.12)_{\delta=0}$ en établissant des estimations a priori *uniformes en δ* des solutions f_δ des équations $(2.12)_\delta$ dans une norme suffisamment forte pour pouvoir ensuite passer à la limite $\delta \rightarrow 0$ dans l'équation $(2.12)_\delta$. Nous verrons qu'il semble difficile de faire mieux que $L^{2,5}$, ce qui, au vu de la non-linéarité de l'équation considérée, est insuffisant. Nous ne pourrions conclure qu'au prix d'une ultime régularisation par un flot parabolique, discutée en (2.4.3).

2.4.1. Estimation a priori $L^{2,3}$. — Les arguments développés dans la preuve du *Theorem 5.1* de [CGY1] permettent de démontrer, sous l'hypothèse $\int_M(\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} > 0$, $\alpha \geq 0$, l'estimation a priori, uniforme en $\delta \in (0, \delta_0)$, δ_0 assez petit,

$$(2.14) \quad \int_M |\text{Hess } f_\delta|_g^3 d\text{vol}(g) + \int_M |df_\delta|_g^{12} d\text{vol}(g) < C .$$

Ceci peut sembler mesquin, mais si, pour cette majoration, le passage du cas particulier $\alpha = 0$, objet de [CGY1], au cas général est aisé dès que l'on a remarqué que, comme opérateur différentiel, $f \mapsto |W|^2(e^{2f}g)$ est du même type que $f \mapsto |\eta|^2(e^{2f}g)$, tous les deux étant des covariants conformes de même poids -4 , la démonstration du cas particulier n'en requiert pas moins de trente pages d'estimations intégrales ingénieuses, et difficiles à résumer. On y retrouve des idées « classiques » conduisant aux estimations C^2 pour les équations de Monge-Ampère, bien que le passage à une équation du quatrième degré interdise dans notre cas le recours systématique au principe du maximum et force celui à des estimations « en moyenne ». Au cœur de cette discussion on trouve la minoration suivante, valable sous la même hypothèse dès que δ est assez petit :

$$(2.15) \quad \int_M \left(\frac{\text{scal}}{6}\right)^3 d\text{vol} \leq (1 + C_1\delta) \int_M |df_\delta|^6 d\text{vol} + C_2 \int_M \text{scal}^2 d\text{vol} + C_3 .$$

Pour comprendre cet énoncé, notons que le *tenseur gravitationnel* $G := -\text{ric} + \frac{\text{scal}}{2}g$ étant de divergence nulle – un corollaire immédiat et classique de la seconde identité de Bianchi – et M sans bord, pour toute fonction infiniment différentiable φ , $\int_M (G, \text{Hess } \varphi) d\text{vol} = 0$. Appliquée à la courbure scalaire, $\varphi = \text{scal}$, cette identité s'écrit alors, en utilisant l'équation $(2.12)_\delta$ et sa dérivée

$$\int_M (6 \text{tr ric}_0^3 + 1/12 \text{scal}^3 + \text{ordres inférieurs}) d\text{vol} \leq \int_M (G, \text{Hess scal}) d\text{vol} \leq 0 ;$$

en l'appliquant à $\varphi = |df_\delta|^2$, elle devient

$$-\frac{1}{12} \int_M (-6 \text{tr ric}_0^3 + \frac{1}{12} \text{scal}^3 - 6 \text{scal} |df_\delta|^4 + \text{ordres inf.}) d\text{vol} \leq \int_M (G, \text{Hess } |df_\delta|^2) d\text{vol} \leq 0 .$$

Une combinaison linéaire adéquate de ces deux inégalités donne la majoration

$$\int_M \left(\frac{\text{scal}}{6}\right)^3 d\text{vol} \leq \int_M \left(\frac{\text{scal}}{6} |df_\delta|^4 + \text{ordres inférieurs}\right) d\text{vol},$$

qui « ressemble » à la majoration annoncée (2.15), dont on peut démontrer qu'elle s'en déduit. Les étapes suivantes s'écrivent

$$\int_M |\text{Hess } f_\delta|^2 |df_\delta|^2 d\text{vol} \leq C \int_M (\delta |df_\delta| + \text{scal}^2 + 1) d\text{vol};$$

et

$$\int_M |df_\delta|^{12} d\text{vol} \leq C \left(\int_M |df_\delta|^6 d\text{vol} + 1 \right)^4.$$

Un important corollaire de la majoration a priori $L^{2,3}$ (2.14) et de sa démonstration que nous venons seulement d'esquisser assure la majoration suivante

$$(2.16) \quad \delta \int_{(M, e^{2f_\delta} g)} \left(\frac{\Delta \text{scal}}{\text{scal}} \right)^2 d\text{vol} \leq C.$$

Pour l'établir, il est important de disposer d'un minorant uniforme en δ de la courbure scalaire des métriques $e^{2f_\delta} g$: par définition de l'ensemble S , $\text{scal}(e^{2f_\delta} g) > 0$, et il suffit de minorer scal^2 . En un minimum de la courbure scalaire, l'équation $(2.12)_\delta$ entraîne la minoration : $0 \geq \delta \Delta \text{scal} \geq -8 \gamma_0 |\eta|^2 - 1/6 \text{scal}_{\min}^2$; on a alors

$$(2.17) \quad \text{scal}_{\min}^2(e^{2f_\delta} g) \geq -48 \gamma_0 \min(e^{-4f_\delta} |\eta|_g^2) \geq C > 0,$$

puisque les fonctions f_δ sont bornées uniformément en norme $L^{2,3}$, et donc en norme C^β , $\beta < 2/3$, d'après les inclusions de Sobolev classiques.

2.4.2. Estimation a priori $L^{2,s}$, $s > 5$. — La dégénérescence de l'équation $(2.12)_\delta$ en $\delta = 0$ est forte : c'est le terme différentiel d'ordre principal Δscal qui disparaît. Ceci explique en partie le prix à payer pour dériver des bornes a priori sur les solutions qui soient uniformes au voisinage de $\delta = 0$. Dans ce paragraphe, nous expliquons comment passer de la borne $L^{2,3}$ discutée ci-dessus à une borne $L^{2,s}$, pour tout $s < 5$.

La démarche est proche de celle que nous venons de présenter et les détails techniques plus lourds encore (dix nouvelles pages d'estimations intégrales sauvages). On applique l'identité $\int_M (G, \text{Hess } \varphi) d\text{vol} = 0$ aux puissances $1 + p$, $p > 0$, de la courbure scalaire et pour cela on introduit

$$I_p := \int_M (G, \text{Hess } \text{scal}^{p+1}) d\text{vol}.$$

Pour préciser l'estimation obtenue en l'appliquant à $|df_\delta|^2$ nous introduisons symétriquement

$$II_p := \frac{1}{2} \int_M (G, D(\text{scal}^p d|df_\delta|^2)) d\text{vol},$$

de telle sorte que, comme précédemment dans le cas $p = 0$, nous avons $I_p = II_p = 0$. I_p se laisse alors décomposer en la somme $I'_p + I''_p$, où

$$I'_p := (p+1) \int_M \text{scal}^p (G, \text{Hess } \text{scal}) d\text{vol}, \text{ et } I''_p := p(p+1) \int_M \text{scal}^{p-1} G(d\text{scal}^\#, d\text{scal}^\#) d\text{vol};$$

nous écrivons de même $II_p = II'_p + II''_p$, où

$$II'_p := \frac{p}{2} \int_M \text{scal}^{p-1} G(d\text{scal}^\sharp, (d|df_\delta|^2)^\sharp) d\text{vol}, \quad \text{et} \quad II''_p := \frac{1}{2} \int_M \text{scal}^p (G, \text{Hess} |df_\delta|^2) d\text{vol}.$$

En reportant l'équation (2.12) $_\delta$ dans l'expression de I'_p , on vérifie alors facilement, dès que γ_0 et γ_1 sont négatifs, la minoration suivante

$$I'_p \geq I'_{p,\delta} + (p+1) \int_M (6 \text{scal}^p \text{tr ric}_0^3 + \text{scal}^{p+1} |\text{ric}_0|^2) d\text{vol} \\ - C \int_M \text{scal}^{p+2} d\text{vol} - C \int_M \text{scal}^p (|d\eta|^2 + |d|W|^2) d\text{vol},$$

où

$$I'_{p,\delta} := \frac{3}{4} \delta (p+1) \int_M (\Delta \text{scal}^p \Delta \text{scal} + 2 \text{scal}^{p-1} (\Delta \text{scal})^2 - 2p \text{scal}^{p-2} |d\text{scal}|^2 \Delta \text{scal}) d\text{vol}.$$

(Rappelons que $\gamma_1 = -\alpha/8$ est négatif ou nul par hypothèse et que l'identité (2.11) et l'hypothèse $\int_M (4\sigma_2(A) - |W|^2) d\text{vol} > 0$ entraînent que γ_0 est lui aussi négatif.)

Une application directe de l'inégalité de Hölder donne alors

$$\int_M \text{scal}^p |d\eta|^2 d\text{vol} \leq \left(\int_M \text{scal}^{p+2} d\text{vol} \right)^{p/p+2} \left(\int_M |d\eta|^{p+2} d\text{vol} \right)^{2/p+2}.$$

De l'identité $|\eta| = e^{-2f_\delta} |\eta|_g$ nous déduisons, les fonctions f_δ étant bornées dans $L^{2,3}(M) \subset L^{1,12}(M)$ uniformément en δ , que les intégrales $(\int |d\eta|^{p+2} d\text{vol}_g)^{2/p+2}$ le sont aussi dès que $p \leq 10$. Le même argument vaut évidemment pour $|W|$ et $\int_M \text{scal}^p (|d\eta|^2 + |d|W|^2) d\text{vol}$ est donc majoré par $C (\int_M \text{scal}^{p+2} d\text{vol} + 1)$ uniformément en δ . La minoration précédente de I'_p s'écrit maintenant :

$$I'_p \geq I'_{p,\delta} + (p+1) \int_M (6 \text{scal}^p \text{tr ric}_0^3 + \text{scal}^{p+1} |\text{ric}_0|^2) d\text{vol} - C \int_M \text{scal}^{p+2} d\text{vol} - C.$$

Pour II'_p , nous utilisons l'équation (2.12) $_\delta$ pour minorer le tenseur gravitationnel : la courbure scalaire de $e^{2f_\delta} g$ étant strictement positive,

$$(2.18) \quad G \geq 3 \frac{\sigma_2(A)}{\text{scal}} = \frac{3}{\text{scal}} \left(-\frac{\delta}{4} \Delta \text{scal} - 2\gamma_0 |\eta|^2 + \frac{\alpha}{4} |W|^2 \right).$$

On établit alors sans trop de mal, pour tout ε , $\varepsilon > 0$, et tout η , $\eta > 0$, la minoration

$$II'_p \geq -\frac{p}{2} \varepsilon^2 \int_M \text{scal}^{p-1} G(d\text{scal}^\sharp, d\text{scal}^\sharp) d\text{vol} - C \delta \varepsilon^2 \eta \int_M \text{scal}^{p-1} (\Delta \text{scal})^2 d\text{vol} \\ - C \delta \varepsilon^2 \eta^{-1} \int_M \text{scal}^{p-3} |d\text{scal}|^4 d\text{vol} - C \varepsilon^{-6} \eta^{-1} \left(\int_M \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+1}{p+3}} - C p \varepsilon^{-2} \left(\int_M \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+2}{p+3}}.$$

Concernant II''_p , il faut un peu de ténacité pour dériver la majoration suivante, valable pour tout $\gamma \in \mathbb{R}_+^*$,

$$II''_p \geq \int_M \frac{\text{scal}^p}{4} \left(-\text{tr ric}_0^3 + \frac{\text{scal}^3}{72} \right) d\text{vol} - C \gamma \delta \int_M \text{scal}^{p-1} (\Delta \text{scal})^2 d\text{vol} \\ - C \gamma \delta \int_M \text{scal}^{p-3} |d\text{scal}|^4 d\text{vol} - C \gamma^{-1} \delta \int_M \text{scal}^{p+3} d\text{vol} - C \int_M \text{scal}^{p+2} d\text{vol} - C.$$

Comme dans le cas $p = 0$ du paragraphe précédent, nous considérons la combinaison $I_p + 24(p + 1)II_p$, qui annule le coefficient du terme $\int_M \text{scal}^p \text{tr ric}_0^3 d\text{vol}$, apparaissant dans l'expression de la minoration de I'_p avec le poids $6(p + 1)$ et dans celle de II''_p avec le poids $-1/4$.

De l'identité riemannienne universelle

$$3p \int_M \text{scal}^{p-2} |d\text{scal}|^2 \Delta \text{scal} d\text{vol} = -4 \int_M \text{scal}^{p-1} |\text{Hess}_0 \text{scal}|^2 d\text{vol} + 3 \int_M \text{scal}^{p-1} (\Delta \text{scal})^2 d\text{vol} \\ + 2(p-2) \int_M \text{scal}^{p-3} |d\text{scal}|^4 d\text{vol} - 4(p-2) \int_M \text{scal}^{p-2} \text{Hess}_0(\text{scal})(d\text{scal}^\#, d\text{scal}^\#) d\text{vol} \\ - 4 \int_M \text{ric}(d\text{scal}^\#, d\text{scal}^\#) d\text{vol},$$

et de l'équation (2.12) $_\delta$, en rappelant de plus que γ_0 est négatif d'après l'identité (2.11) et l'hypothèse $\int_M (\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} > 0$, nous déduisons la majoration suivante, valable pour tout réel p , $p < 2$:

$$3(p - \delta) \int_M \text{scal}^{p-2} |d\text{scal}|^2 \Delta \text{scal} d\text{vol} \\ \leq 3 \int_M \text{scal}^{p-1} (\Delta \text{scal})^2 d\text{vol} - (2 - p) \left(\frac{1}{2} + \frac{3}{4}p \right) \int_M \text{scal}^{p-3} |d\text{scal}|^4 d\text{vol}.$$

(C'est ici qu'apparaît de façon essentielle la restriction $p < 2$, que l'on voit mal comment relâcher). En invoquant à nouveau la minoration de G par $-\frac{3\delta}{4} \frac{\Delta \text{scal}}{\text{scal}}$ (cf. (2.18)), nous démontrons par un choix adéquat des paramètres η et ε dans les relations précédentes la minoration suivante, valable pour tout δ assez petit,

$$I'_{p,\delta} + I''_p + 24(p + 1)II'_p \geq \\ C\delta \int_M \text{scal}^{p-1} (\Delta \text{scal})^2 d\text{vol} + C\delta \int_M \text{scal}^{p-3} |d\text{scal}|^4 d\text{vol} - C \left(\int_M \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+2}{p+3}} - C.$$

Choisissant alors γ suffisamment petit dans l'expression précédente de la minoration de II''_p , nous établissons

$$I_p + 24(p+1)II_p \geq \left(\frac{p+1}{6} - C\delta \right) \int_M \text{scal}^{p+3} d\text{vol} - C \left(\int_M \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+2}{p+3}} - C \int_M \text{scal}^{p+2} d\text{vol} - C,$$

et donc, en majorant $\int_M \text{scal}^{p+2} d\text{vol}$ par $C(g) \left(\int_M \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+2}{p+3}}$, la majoration recherchée de $\int \text{scal}^{p+3} d\text{vol}$, puisque nous obtenons :

$$\left(\int \text{scal}^{p+3} d\text{vol} \right)^{\frac{p+2}{p+3}} \geq C_1 \int \text{scal}^{p+3} - C_2.$$

En rappelant que les solutions f_δ de la régularisation sont uniformément bornées par l'estimation $L^{2,3}$ du paragraphe précédent, nous déduisons alors de l'expression (2.7) de la courbure scalaire dans une classe conforme une majoration uniforme de $\Delta_g f_\delta$ dans les normes L^{p+3} , pour tout réel p , $0 \leq p < 2$, et donc la proposition suivante :

PROPOSITION 2.4. — *Pour tout $s \in [0, 5)$, les fonctions f_δ , solutions de moyenne nulle des équations régularisées $(2.12)_\delta$, $0 < \delta < \delta_0$, sont uniformément bornées en norme $L^{2,s}$. Elles le sont donc aussi pour les normes höldériennes $C^{1,\beta}$, $\beta < 1/5$.*

2.4.3. *Flot de Yamabe.* — La restriction $p < 2$, i.e. $s < 5$, est intervenue cruciallement dans les majorations a priori précédente, et la borne en norme $L^{2,s < 5}$ à laquelle elle conduit est pourtant encore trop faible pour permettre de passer à la limite $\delta = 0$ dans les équations $(2.12)_\delta$. Il faut une idée...

C'est l'endroit où rappeler la majoration L^2 uniforme (2.16) que nous pouvons écrire, en invoquant $(2.12)_\delta$ et en posant $\xi(f_\delta) := (\sigma_2(A) - \frac{\alpha}{4}|W|^2 + 2\gamma_0|\eta|^2)(e^{2f_\delta}g)/\text{scal}(e^{2f_\delta}g)$,

$$(2.19) \quad \int_{(M, e^{2f_\delta}g)} \xi^2(f_\delta) d\text{vol} = \frac{\delta^2}{16} \int_{(M, e^{2f_\delta}g)} \left(\frac{\Delta \text{scal}}{\text{scal}}\right)^2 d\text{vol} \leq C\delta.$$

Il reste à invoquer l'outil régularisant universel – un *flot parabolique* – pour passer d'une métrique $e^{2f_\delta}g$ satisfaisant $\|\xi(e^{2f_\delta})\|_{L^2}^2 < C\delta$ à une métrique $h_\delta = e^{2\hat{f}_\delta}g$ satisfaisant partout sur la variété M la majoration $|\xi(e^{2\hat{f}_\delta})| < C(\delta)$, pour une fonction $C(\delta)$ tendant vers zéro avec δ ; nous en déduisons que le polynôme $(\sigma_2(A) - \frac{\alpha}{4}|W|^2)(e^{2\hat{f}_\delta}g)$ est strictement positif partout sur M dès que le paramètre δ est assez petit. Nous détaillons ces arguments dans ce paragraphe.

THÉORÈME 2.5. — *Soit (M, g) une variété riemannienne compacte sans bord de dimension 4 de courbure scalaire strictement positive et telle que $\int_M (4\sigma_2(A) - |W|^2) d\text{vol} > 0$. Pour tout δ suffisamment petit, il existe dans la classe conforme de toute solution $e^{2f_\delta}g$ de courbure scalaire positive de l'équation régularisée $(2.12)_\delta$ une métrique (infiniment différentiable) h_δ telle que le polynôme en la courbure de Ricci $\sigma_2(A(h)) - \alpha/4|W(h)|^2$ est partout strictement positif.*

L'intégrabilité local du champ de vecteurs

$$(2.20) \quad X(h) = -\frac{\text{scal}}{3}h$$

est un résultat classique (« flot de Yamabe ») ; les propriétés régularisantes de ce flot qui nous seront utiles sont résumées dans l'énoncé suivant, dont nous n'utiliserons que la version de dimension 4.

PROPOSITION 2.6 ([Y]). — *Soient M une variété compacte sans bord et $h(t), t \in [0, T]$, où T est un réel strictement positif, une famille de métriques sur M dont les constantes de Sobolev sont uniformément majorées i.e. telles que toute fonction infiniment différentiable φ sur M satisfait la majoration*

$$(2.21) \quad \|\varphi\|_{L^{2n/n-2}(M, h(t))} \leq C \|\varphi\|_{L^{1,2}(M, h(t))},$$

où la constante C est indépendante de φ et de $t \in [0, T]$. Pour toute fonction positive ou nulle sur $M \times [0, T]$, ρ , satisfaisant l'inéquation différentielle $\frac{d}{dt} \log d\text{vol}(h(t)) \leq \rho$, et pour

tous réels $p_0, p, q, p_0 > 1, p \geq p_0$, et $q > n$, toute solution de l'inéquation différentielle $\frac{\partial \varphi}{\partial t} + \Delta \varphi \leq \rho \varphi$ vérifie les majorations a priori

$$1) \quad \sup_M |\varphi(t, \cdot)| \leq C_1 e^{c_1 t} t^{-n/2p_0} \|\varphi(0, \cdot)\|_{L^{p_0}},$$

$$2) \quad \forall t \in [0, T], \quad \frac{d}{dt} \int_M \varphi^p d\text{vol} + \int_M |d\varphi^{\frac{p}{2}}|^2 d\text{vol} \leq C_2 p^{\frac{2n}{q-n}} \int_M \varphi^p d\text{vol},$$

où la constante C_2 dépend uniquement de n, q, p_0 et C , la constante C_1 dépendant de plus de la norme $\sup_{[0, T]} \|\rho\|_{L^{4/2}(M, h_t)}$.

Nous appliquerons cette proposition aux courbes intégrales du « flot de Yamabe » issues des métriques $e^{2f_\delta} g$; notons pour cela que le « flot de Yamabe » préserve – par définition – les classes conformes (cf. (2.21)) et que toutes les métriques que nous considérons sont donc conformes à la métrique de référence, g .

Par définition de l'invariant de Yamabe nous avons la relation suivante

$$\mu(g) \int_M \varphi^4 d\text{vol}(h) \leq \int_M |d\varphi|^2 d\text{vol}(h) + \int_M \text{scal}(h) \varphi^2 d\text{vol}(h);$$

en nous restreignant à un intervalle $[0, T_0(h_0 = e^{2f_\delta} g)]$ sur lequel $\int_M \text{scal}^s(h_t) d\text{vol}(h_t) \leq 2 \int_M \text{scal}^s(h_0) d\text{vol}(h_0)$, en rappelant la borne uniforme sur les intégrales $\int_M \text{scal}^s(h_0) d\text{vol}(h_0)$, $h_0 = e^{2f_\delta} g$, établie à la proposition 2.4, et en supposant $s > 2$, nous en déduisons la majoration

$$\mu(g) \left(\int_M \varphi^4 d\text{vol}(h) \right)^{1/2} \leq C(g) \int_M \varphi^2 d\text{vol}(h) + \frac{1}{2} \mu(g) \left(\int_M \varphi^4 d\text{vol}(h) \right)^{1/2} + 6 \int_M |d\varphi|^2 d\text{vol}(h),$$

c'est-à-dire la majoration uniforme des constantes de Sobolev (2.21), requise par l'énoncé de la proposition 2.6.

2.4.3.1. *La courbure scalaire.* — Une première application de la proposition 2.6 avec $\rho = \text{scal}$, $p = p_0 = s$ et $q = 2s > 4$ établit l'inéquation différentielle

$$\frac{d}{dt} \int_M \text{scal}^s d\text{vol} \leq C(g, s) \int_M \text{scal}^s d\text{vol},$$

et donc la majoration $\int_M \text{scal}^s(h_t) d\text{vol}(h_t) \leq e^{Ct} \int_M \text{scal}^s(h_0) d\text{vol}(h_0)$. D'après les estimations a priori du paragraphe 2.4.2 nous disposons, pour $s < 5$, d'une borne uniforme sur les intégrales $\int_{(M, e^{2f_\delta} g)} \text{scal}^s d\text{vol}$, et donc, par la majoration précédente, d'un minorant pour $T_0(e^{2f_\delta} g)$ ne dépendant que de s et de g – et uniforme, en particulier, en δ , $\delta \in (0, 1] -$, que nous noterons $T_1, T_1 > 0$.

L'autre conclusion de la proposition 2.6 nous enseigne alors que la courbure scalaire admet la majoration

$$(2.22) \quad \text{scal} \leq C(s, g) t^{-2/s}, \quad t \in [0, T_1],$$

tandis que le principe du maximum parabolique appliqué à l'équation régissant l'évolution de la courbure scalaire le long des courbes intégrales,

$$(2.23) \quad \frac{\partial}{\partial t} \text{scal} + \Delta \text{scal} = \frac{1}{3} \text{scal}^3 ,$$

garantit que le minimum de la courbure scalaire le long du « flot de Yamabe » est une fonction monotone croissante. De la minoration uniforme (2.17) des courbures scalaires des métriques $e^{2f_\delta} g$ nous déduisons alors l'existence d'un minorant uniforme – ne dépendant que de g – des courbures scalaires des métriques h_t , $t \in [0, T_1]$, le long des courbes intégrales issues des métriques $e^{2f_\delta} g$, $\delta \in (0, 1]$,

$$(2.24) \quad \text{scal}(h_t) \geq C > 0 .$$

Remarquons finalement qu'en écrivant une courbe intégrale sous la forme $h(t) = e^{2x(t)} h(0)$ l'équation du flot devient

$$(2.25) \quad \frac{\partial x}{\partial t} = -\frac{1}{6} \text{scal}(h_t), \quad x(0, \cdot) = 0 .$$

De la borne uniforme (2.22) pour la courbure scalaire, nous déduisons immédiatement la borne uniforme suivante pour le facteur conforme $x(t)$:

$$(2.26) \quad \sup_M |x| \leq C(s, g) T_1(s, g)^{1-2/s} \leq C_1(s, g) .$$

2.4.3.2. *La courbure de Ricci.* — L'équation régissant la courbure de Ricci le long du « flot de Yamabe » s'écrit

$$(2.27) \quad \left(\frac{\partial}{\partial t} + \Delta \right) \text{ric} = -2 \text{ric} \circ \text{ric} + \frac{1}{2} |\text{ric}|^2 g + \frac{2}{3} \text{scal} \text{ric} - \frac{1}{6} \text{scal}^2 g + 2 W(\text{ric}) + 2 B ,$$

où $B := -\text{tr}_{13} \text{tr}_{25} D^2 W - 1/2 W(\text{ric})$ représente le *tenseur de Bach*, un covariant conforme fondamental que nous retrouverons dans l'étude du cas limite comme gradient de la fonctionnelle $\|W\|_{L^2}^2$: notons simplement pour l'instant la relation $B(e^{2f} g) = e^{-2f} B(g)$. De l'inéquation différentielle immédiate

$$\left(\frac{\partial}{\partial t} + \Delta \right) |\text{ric}|^2 \leq -2 |D \text{ric}|^2 + C |\text{ric}|^3 + 4 |W| |\text{ric}|^2 + 4 |B| |\text{ric}| ,$$

– où C représente une constante universelle –, de la covariance conforme des courbures de Weyl et de Bach, et de la borne uniforme sur les modules conformes (2.26) nous déduisons l'existence d'une constante $C_1 = C(g)$ telle que le long d'une courbe intégrale du « flot de Yamabe » issue d'une métrique $e^{2f_\delta} g$ solution de la régularisation (2.12) $_\delta$ nous avons la majoration différentielle uniforme

$$\left(\frac{\partial}{\partial t} + \Delta \right) |\text{ric}| \leq C_1 |\text{ric}|^2 , \quad \forall t \in [0, T_1] .$$

Nous pouvons donc encore invoquer la proposition 2.6, avec cette fois $\varphi = \rho/C_1 = |\text{ric}|$, $q = 2s > 4$, et $p = p_0 = s$, pour établir l'inéquation différentielle

$$\frac{d}{dt} \int_M |\text{ric}|^s d\text{vol} \leq C(s, g) \int_M |\text{ric}|^s d\text{vol} .$$

Elle s'intègre en la majoration $\int_M |\text{ric}|^s d\text{vol} \leq e^{ct} \int_M |\text{ric}(h_0)|^s d\text{vol}(h_0)$. Comme précédemment pour la courbure scalaire, nous en déduisons l'existence d'un minorant uniforme, $T_2 := T_2(s, g)$, $0 < T_2 \leq T_1$, de la longueur du plus grand intervalle sur lequel $\int_M |\text{ric}|^s d\text{vol} \leq 2 \int_M |\text{ric}(h_0)|^s d\text{vol}(h_0)$. Par ailleurs, la courbure de Ricci de la métrique $e^{2f_\delta} g$ étant donnée par l'identité

$$(2.28) \quad \text{ric}(e^{2f} g) = \text{ric}(g) - 2 \text{Hess} f - \Delta f g + 2 df \otimes df - 2 |df|^2 g,$$

nous déduisons de la borne uniforme $L^{2,s}$, $s < 5$, sur les solutions f_δ de l'équation régularisée l'existence d'un majorant uniforme des normes $\int_{(M, e^{2f_\delta} g)} |\text{ric}|^s d\text{vol}$; la majoration uniforme

$$(2.29) \quad \sup_M |\text{ric}(h_t)| \leq C(g) t^{-2/s}$$

est alors un corollaire immédiat de celle-ci et du point 1) de la proposition 2.6.

2.4.3.3. *Conclusion.* — À ce point, et pour les raisons évoquées au début de cette partie 2.4, il est naturel d'essayer d'appliquer la proposition 2.6 à l'invariant ξ ,

$$(2.30) \quad \xi(f) = \text{scal}^{-1} (\sigma_2(A) - \frac{\alpha}{4} |W|^2 + 2 \gamma_0 |\eta|^2) (e^{2f} g).$$

Puisque c'est la positivité de ξ que nous cherchons à établir, il sera plus simple de travailler avec $\hat{\xi} = \max\{-\xi, 0\}$. Nous devons alors vérifier les hypothèses de la proposition 2.6, et en particulier établir pour $\hat{\xi}$ une inéquation différentielle de la forme $(\frac{\partial}{\partial t} + \Delta) \hat{\xi} < \rho \hat{\xi}$, pour une solution ρ de l'inéquation différentielle $(\frac{\partial}{\partial t} + \Delta) \log d\text{vol} \leq \rho$.

PROPOSITION 2.7. — *Pour tout $s \in (2, 5)$, la fonction $\hat{\xi}$ vérifie sur l'intervalle $[0, T_2(s, g)]$ l'inéquation différentielle $(\frac{\partial}{\partial t} + \Delta) \hat{\xi} \leq C(s, g) |\text{ric}| (1 + \hat{\xi})$, où la constante $C(s, g)$ est en particulier indépendante de la condition initiale $e^{2f_\delta} g$, $\delta > 0$, du « flot de Yamabe ».*

Des équations d'évolution de la courbure scalaire (2.23), de la courbure de Ricci (2.27) et du facteur conforme (2.25) nous déduisons celle du produit ξscal :

$$\begin{aligned} (\frac{\partial}{\partial t} + \Delta) (\xi \text{scal}) &= \frac{2}{3} \xi \text{scal} + |D \text{ric}_0|^2 - \frac{1}{12} |d\text{scal}|^2 + 2 \text{tr ric}_0^3 + \frac{1}{3} |\text{ric}_0|^2 \text{scal} \\ &\quad - 2 W(\text{ric}_0, \text{ric}_0) + 2 (B, \text{ric}_0) + 2 \gamma_0 \Delta |\eta|^2 - \frac{\alpha}{4} \Delta |W|^2. \end{aligned}$$

On vérifie alors les quatre majorations élémentaires suivantes

$$(1.) \quad |D \text{ric}_0|^2 - \frac{1}{12} |d\text{scal}|^2 \geq -\frac{2}{\text{scal}} (d(\xi \text{scal}), d\text{scal}) + 2 \xi \text{scal} \left| \frac{d\text{scal}}{\text{scal}} \right|^2 + 4 \gamma_0 |d|\eta||^2 - \frac{\alpha}{4} |d|W||^2;$$

$$(2.) \quad 2 \gamma_0 \Delta |\eta|^2 - \frac{\alpha}{4} \Delta |W|^2 + 4 \gamma_0 |d|\eta||^2 - \frac{\alpha}{2} |d|W||^2 \geq -C(g);$$

$$(3.) \quad 2 \text{scal}^{-1} \text{tr ric}_0^3 + \frac{1}{3} |\text{ric}_0|^2 \geq \frac{8 |\text{ric}_0|^2 \xi}{2 \sqrt{3} |\text{ric}_0| + \text{scal}};$$

et, en invoquant la covariance conforme de W et de B , et la borne uniforme (2.26) sur la norme sup du facteur conforme,

$$(4.) \quad -2 \operatorname{scal}^{-1} W(\operatorname{ric}_0, \operatorname{ric}_0) - 2 \operatorname{scal}^{-1} (B, \operatorname{ric}_0) \geq C_1(g) \xi - C_2(g) - C_3(g) \operatorname{scal} .$$

Ces majorations sont uniformes en $\delta \in (0, \delta_0]$ et valables sur $[0, T_2(s, g)]$. Elles nous permettent de dériver de l'équation d'évolution de ξ l'inéquation différentielle suivante

$$\left(\frac{\partial}{\partial t} + \Delta\right) \xi \geq \frac{2}{\operatorname{scal}} \operatorname{tr} \operatorname{ric}_0^3 + \frac{1}{3} |\operatorname{ric}_0|^2 + \frac{1}{3} \xi \operatorname{scal} - 2 W(\operatorname{ric}_0, \operatorname{ric}_0) \operatorname{scal}^{-1} - 2 (B, \operatorname{ric}_0) \operatorname{scal}^{-1} - C(g) \geq \frac{8 |\operatorname{ric}_0|^2}{2 \sqrt{3} |\operatorname{ric}_0| + \operatorname{scal}} \xi + \left(\frac{\operatorname{scal}}{3} + C_1(g)\right) \xi - C_2(g) - C_3(g) \operatorname{scal} .$$

En utilisant une fois encore la minoration uniforme de la courbure scalaire (2.24) : $|\operatorname{ric}| \geq \frac{\operatorname{scal}}{2} \geq C(g) > 0$, nous en déduisons que $\hat{\xi} = \max(-\xi, 0)$ satisfait l'inéquation différentielle suivante

$$(2.31) \quad \left(\frac{\partial}{\partial t} + \Delta\right) \hat{\xi} \leq c_1(g) \left(|\operatorname{ric}_0| + \operatorname{scal} + 1\right) \hat{\xi} + 1 \leq c_2(g) |\operatorname{ric}| \left(\hat{\xi} + 1\right).$$

Pour nous débarrasser du terme constant de l'inéquation différentielle (2.31), il suffit de soustraire de $\hat{\xi}$ une fonction adaptée de la variable réelle t , $\hat{\xi}_0$: la fonction $\hat{\xi}_0$ est par définition constante sur M et la différence $\tau = \hat{\xi} - \hat{\xi}_0$ vérifie l'inéquation

$$\left(\frac{\partial}{\partial t} + \Delta\right) \tau \leq C(g) |\operatorname{ric}| \tau + C(g) |\operatorname{ric}| (1 + \hat{\xi}_0) - \frac{d\hat{\xi}_0}{dt} ,$$

où, d'après (2.29), $\sup_M |\operatorname{ric}| \leq C_1(g) t^{-2/s}$. Si l'on pose $\hat{\xi}_0(t) = e^{C(s)t^{\frac{s-2}{s}}} - 1$, avec $C(s) = CC_1 s/(s-2)$, la fonction τ satisfait l'inéquation différentielle $\left(\frac{\partial}{\partial t} + \Delta\right) \tau \leq C(g) |\operatorname{ric}| \tau$; on peut alors invoquer la proposition 2.6, en posant $\varphi = \tau$, $\rho = C(g) |\operatorname{ric}|$, $p_0 = 2$ et $q = 2s$, et conclure :

$$\sup_M |\tau| \leq \frac{C}{t} \|\tau(0, \cdot)\|_{L^2} = \frac{C}{t} \|\hat{\xi}(0, \cdot)\|_{L^2} \leq \frac{C}{t} \|\xi\|_{L^2} .$$

De la majoration fondamentale (2.19) nous déduisons ensuite $\xi \geq -\hat{\xi}_0(t) - \frac{C}{t} \sqrt{\delta}$, et donc, en rappelant la majoration (2.22) de la courbure scalaire,

$$\sigma_2(A) - \frac{\alpha}{4} |W|^2 \geq -2 \gamma_0 |\eta|^2 - C(s, g) (t^{1-4/s} + \sqrt{\delta} t^{-(1+2/s)}) .$$

Ces deux dernières minoration sont valables pour tout t de l'intervalle $[0, T_2(s, g)]$. Remarquons encore, en rappelant la majoration (2.26) du facteur conforme, et la définition (2.10) de γ_0 , la minoration

$$-2 \gamma_0 |\eta|^2 \geq -2 C_1(s, g) \gamma_0 |\eta|_g^2 \geq C_2(s, g) \int_{(M, g)} \left(\sigma_2(A) - \frac{\alpha}{4} |W|^2\right) d\operatorname{vol} = C_3(s, g, \alpha) > 0 ,$$

de telle sorte que, finalement,

$$\sigma_2(A) - \frac{\alpha}{4} |W|^2 \geq c_1(s, g, \alpha) - c_2(s, g) t^{1-4/s} - c_3(s, g) \sqrt{\delta} t^{-(1+2/s)} ,$$

où $c_1(s, g, \alpha)$ est strictement positif. Pour tout point s de l'intervalle ouvert (4, 5), il reste à considérer un point t_1 de $[0, T_2(s, g)]$ tel que $c_1 - c_2 t_1^{1-4/s}$ est minoré par $c_1/2$ pour conclure

aisément à l'existence d'un réel strictement positif δ_0 tel que pour tout point δ de $(0, \delta_0]$ la courbe intégrale du flot de Yamabe issue de $e^{2f\delta}g$ contient des métriques – par exemple $h(t_1)$ – pour lesquelles le polynôme $\sigma_2(A) - \frac{\alpha}{4}|W|^2$ est, partout sur M , strictement positif. Ceci conclut la preuve du théorème 2.5 et donc celle du théorème 1.

3. UNE PREUVE PLUS SATISFAISANTE DU THÉORÈME 1

Oubliant les motivations initiales de Chang S.-Y. A., M. Gursky et Yang P. et la stratégie de preuve qui en découle, largement discutées dans le chapitre 2, nous proposons ici une preuve beaucoup plus courte du théorème 1. Dans des travaux contemporains de [CGY0] et [CGY1], Guan P. et Wang G. dérivent les estimations C^2 a priori des solutions des équations $\sigma_k(A(e^{2f}g)) = \varphi$, $k \in \{1, \dots, n\}$, (cf. [GW]). Même si celles-ci s'avèrent insuffisantes pour notre propos, elles ouvrent la voie à une preuve directe du théorème 1, qui s'affranchit de la régularisation par le terme du quatrième ordre différentiel, $\delta \Delta \text{scal}(e^{2f}g)$, et des délicates estimations a priori que requiert le passage à la limite $\delta = 0$. Dans la démonstration que nous donnons ici nous suivons l'approche qui a permis à M. Gursky et J. Viaclovsky de donner une preuve « simple » de [CGY1] (l'existence de métriques avec $\sigma_2(A) > 0$) et la généralisons à l'invariant $\sigma_2(A) - \frac{\alpha}{4}|W|^2$, $\alpha \geq 0$.

3.1. Une nouvelle déformation de l'opérateur $\sigma_2(A) - \frac{\alpha}{4}|W|^2$, $\alpha \geq 0$

Commençons par utiliser la métrique de référence, g – par opposition à $e^{2f}g$ – pour identifier formes quadratiques et endomorphismes symétriques partout dans la classe conforme de g , quelque chose comme la « jauge de Piola-Kirchoff » des mécaniciens, un outil souvent efficace pour simplifier l'apparence des équations, mais rarement déterminant : l'équation $(\sigma_2(A) - \alpha/4|W|^2)(e^{2f}g) = \varphi$ devient ainsi, en utilisant la covariance conforme de la courbure de Weyl, $\sigma_2(g^{-1}A(e^{2f}g)) - \alpha/4|W(g)|_g^2 = \varphi e^{4f}$. Pour des raisons techniques (cf. la discussion de l'ellipticité par exemple) on cherchera plutôt à résoudre l'équation

$$(3.1) \quad \sigma_2(g^{-1}A) - \frac{\alpha}{4}|W(g)|^2 = \psi e^{-4f}, \quad \psi > 0,$$

ce qui, dans la « jauge » initiale, reviendrait à résoudre l'équation $(\sigma_2(A) - \frac{\alpha}{4}|W|^2)(e^{2f}g) = \psi e^{-8f}$, $\psi > 0$, tout aussi naturelle, dans notre perspective, que l'équation initiale $(\sigma_2(A) - \alpha/4|W|^2)(e^{2f}g) = \varphi$.

En suivant la démarche de M. Gursky et J. Viaclovsky dans leur étude du cas $\alpha = 0$ (voir [GV], et, pour une variante, [LL]), nous introduisons la déformation

$$(3.2) \quad A_x = \text{ric} - \frac{x}{6} \text{scal} g, \quad x \in \mathbb{R}.$$

De l'hypothèse $\text{scal}(g) > 0$ nous déduisons que la forme bilinéaire symétrique A_x est définie positive et strictement minorée par $\sqrt{6\alpha}/12 |W(g)|_g^2 g$ pour x suffisamment petit (et négatif si nécessaire), de telle sorte que la fonction

$$\psi_g := \sigma_2(g^{-1}A_{x_0}) - \frac{\alpha}{4}|W(g)|_g^2$$

est strictement positive pour un tel choix du réel x_0 . La fonction identiquement nulle fournit alors pour $x = x_0$ une solution de courbure scalaire strictement positive de l'équation

$$(3.3) \quad \sigma_2(g^{-1}A_x) - \frac{\alpha}{4} |W(g)|_g^2 = \psi_g e^{-4f}, \alpha \geq 0.$$

Par un argument de connexité dans l'esprit de celui développé en 2.3 pour la résolution des régularisations considérées dans [CGY1] et connu en analyse sous l'appellation « *méthode de la continuité* », nous allons établir l'existence d'une solution de l'équation (3.3)_x dans la classe conforme de la métrique g pour tout x , $x_0 \leq x \leq x_1 \leq 1$, dès que l'invariant conforme

$$(3.4) \quad C(g, x_1) := \int_M (\sigma_2(A) - \frac{\alpha}{4} |W|^2) d\text{vol} + \frac{1}{6} (1 - x_1)(2 - x_1) \mu^2(g)$$

est strictement positif. (Nous rappelons que nous représentons par

$$(3.5) \quad \mu(g) := \inf_{h \in [g]} \int_M \text{scal}(h) d\text{vol}(h) / \sqrt{\text{vol}(h)}$$

l'invariant de Yamabe de la classe conforme de la métrique g .)

Posons pour cela $\mathcal{S} = \{x \in [x_0, x_1] \text{ tel que l'équation (3.3)}_x \text{ admet une solution } f_x \text{ de régularité } C^{2,\beta}, \beta > 0, \text{ pour laquelle la courbure scalaire de la métrique } e^{2f_x}g \text{ est strictement positive}\}$.

\mathcal{S} n'est pas vide : nous avons déjà observé que $x_0 \in \mathcal{S}$.

3.2. Ellipticité et ouverture

L'ouverture de l'ensemble \mathcal{S} ainsi défini résultera du théorème d'inversion locale appliqué à l'opérateur

$$\begin{pmatrix} C^{2,\beta} & \rightarrow & C^\beta \\ f & \mapsto & \sigma_2(g^{-1}A_x(e^{2f}g)) - \frac{\alpha}{4} |W_g|_g^2 - \psi_g e^{-4f} \end{pmatrix},$$

où $\beta \in (0, 1)$, dès que nous aurons vérifié que sa linéarisation en toute solution f_x , $x \in \mathcal{S}$, est inversible.

De l'expression des courbures scalaire (2.13) et de Ricci (2.28) dans une classe conforme nous déduisons sans mal l'identité

$$(3.6) \quad A_x(e^{2f}g) = A_x(g) + 2(-\text{Hess}f + \frac{(1-x)}{2}\Delta f g - df \otimes df + \frac{(2-x)}{2}|df|^2 g).$$

En remarquant que le gradient de la fonction $g \rightarrow \sigma_2(g^{-1}A(g))$ est donné par le tenseur

$$(3.7) \quad T(g^{-1}A_x) = \text{tr}(g^{-1}A_x)1d - g^{-1}A_x,$$

(une généralisation du tenseur gravitationnel, le cas $x = 1$, exprimée dans la jauge de Piola-Kirchoff), nous dérivons alors l'expression suivante pour la linéarisation $\mathcal{L}_{x,f}$ en une solution f de l'opérateur $\sigma_2(g^{-1}A_x(e^{2f}g)) - \frac{\alpha}{4} |W(g)|_g^2 - \psi(x) e^{-4f}$

$$\mathcal{L}_{x,f}(u) = (T(g^{-1}A(e^{2f}g)), -\text{Hess}u + (\frac{1-x}{2})\Delta u g + (2-x)(df, du)g - 2df \otimes du)_g + 4\psi e^{-4f}u.$$

En introduisant la combinaison linéaire suivante de $T(a)$ et de sa trace,

$$(3.8) \quad T_x(a) := T(a) + \frac{1-x}{2} \operatorname{tr}(T(a)) \operatorname{Id},$$

nous obtenons pour $\mathcal{L}_{x,f}(u)$ l'expression

$$\begin{aligned} -\mathcal{L}_{x,f}(u) &= \operatorname{tr}(T_x(g^{-1}A(e^{2f}g)) \circ (g^{-1}\operatorname{Hess} u)) - (2-x)(df, du)_g \operatorname{tr} T(g^{-1}A(e^{2f}g)) \\ &\quad + 2T(g^{-1}A(e^{2f}g))(df^{\sharp_g})(du) - 4\psi_g e^{-4f}u, \end{aligned}$$

soit encore, en posant pour une métrique $h = e^{2f}g$

$$T(A(h)) = \operatorname{tr}_h A(h)h - A(h), \quad \text{et} \quad T_x(A(h)) = T(A(h)) + \frac{1-x}{2} \operatorname{tr}_h T(A(h))h,$$

l'expression équivalente

$$(3.9) \quad \begin{aligned} -\mathcal{L}_{x,f}(u) &= (T_x(A(e^{2f}g)), \operatorname{Hess} u)_g - (2-x)(df, du)_g \operatorname{tr}_g T(A(e^{2f}g)) \\ &\quad + 2(T(A(e^{2f}g)), df \otimes du)_g - 4\psi_g e^{-4f}u. \end{aligned}$$

L'endomorphisme symétrique $T(g^{-1}A(e^{2f}g))$ – et donc l'endomorphisme $T_x(g^{-1}A(e^{2f}g))$ pour $x \leq 1$ – est défini positif dès que $\sigma_2(g^{-1}A_x(e^{2f}g))$ et $\operatorname{tr} g^{-1}A_x(e^{2f}g) = (1 - 3/2 x) e^{2f} \operatorname{scal}(e^{2f}g)$ sont strictement positifs. Ces deux conditions étant satisfaites pour une solution f_x , $x \in \mathcal{S}$, puisque

$$(3.10) \quad \sigma_2(e^{2f_x}g) = \frac{\alpha}{4} |W(g)|_g^2 + \psi_g e^{-4f_x},$$

où $\alpha \geq 0$ par hypothèse, $-\mathcal{L}_{x,f}$ est elliptique ; le terme d'ordre différentiel zéro $-4\psi e^{-4f}$ étant strictement négatif, $\mathcal{L}_{x,f}$ est même inversible en toute solution f_x de régularité C^β : l'ensemble \mathcal{S} est donc ouvert.

3.3. Estimations C^2 a priori à la Guan-Wang-Li-Li et fermeture

3.3.1. *Minoration.* — De l'inégalité $3(\operatorname{tr} a)^2 - 8\sigma_2(a) \geq 0$ valable pour tout endomorphisme symétrique positif ou nul, nous déduisons la majoration

$$8\psi e^{-4f_x} \leq 3(\operatorname{tr}_g A_x(e^{2f_x}g))^2 = 3(\operatorname{tr}_g A_x(g) + (3-2x)\Delta f_x)^2;$$

puisque $x \leq 1 < \frac{3}{2}$, en un minimum P de f_x , nous concluons : $8\psi e^{-4f_x} \leq 3(\operatorname{tr} A_x(g))^2$, c'est-à-dire à une minoration de f_x ne dépendant que de g (et indépendante, en particulier, de x , $x \geq x_0$, puisque $\operatorname{tr}_g A_x(g) = \operatorname{scal}(g)(1 - \frac{2x}{3})$ par définition, et que la courbure scalaire $\operatorname{scal}(g)$ est strictement positive par hypothèse).

Comme souvent, l'inégalité de Harnack passe par une estimation a priori du gradient.

3.3.2. *Gradient.* — Dans ce paragraphe nous établissons que toute solution C^3 de l'équation (3.3)_x vérifie l'estimation a priori

$$(3.11) \quad \sup_M |df_x| < C ,$$

où la constante ne dépend que de la métrique g et d'un minorant uniforme de f_x , dont nous venons d'établir l'existence.

Ces estimations reprennent celles dérivées dans [GW] (pour le cas $x = 1$) et [LL] (pour l'extension au cas $x \leq 1$). En un maximum P de la fonction $\gamma = |df_x|^2$,

$$(3.12) \quad d\gamma(P) = 2(Ddf_x, df_x)(P) = 0 ,$$

et le hessien de γ est négatif ou nul : $\text{Hess } \gamma(P) = 2 \text{tr}_{34}(D^2df_x \otimes df_x)(P) + 2 \text{tr}_{24}(Ddf_x \otimes Ddf_x)(P) \leq 0$. Nous avons déjà observé que la positivité stricte de $\text{tr } A_x$ et $\sigma_2(A_x)$ implique que les opérateurs $T_x(A_x)$ sont, pour $x \leq 1$, définis positifs ; nous en déduisons que la fonction

$$(3.13) \quad (T_x(A_x), Dd\gamma)_g = 2(T_x(A_x), \text{tr}_{34}(D^2df_x \otimes df_x) + \text{tr}_{24}(Ddf_x \otimes Ddf_x))_g$$

est négative ou nulle en P . Pour évaluer le terme d'ordre différentiel 3, nous dérivons l'équation (3.3) pour obtenir, à partir de l'identité (3.6) et de l'expression (3.7) du gradient,

$$\begin{aligned} & \text{tr}_{14}^g \text{tr}_{25}^g T(A_x(e^{2f_x}g)) \otimes (D^2df_x - \frac{(1-x)}{2} d\Delta f_x \otimes g) = \text{tr}_{14}^g \text{tr}_{25}^g T_x(A_x(e^{2f_x}g)) \otimes D^2df_x \\ & = \text{tr}_{14}^g \text{tr}_{25}^g T(A_x(e^{2f_x}g)) \otimes (D(A_x(g)) + 2 \text{Hess}f_x \otimes df_x - (2-x) \text{tr}_{23}(Ddf_x \otimes df_x) \otimes g) \\ & \quad - \frac{\alpha}{4} d|W|^2 - e^{-4f_x} d\psi + 4\psi e^{-4f_x} df_x . \end{aligned}$$

En contractant cette identité contre df et en remarquant qu'au maximum P de γ (cf. (3.12)) $\text{tr}_{23}^g(Ddf \otimes df)(P) = 0$ et que, $\text{Hess}f$ étant un tenseur symétrique, $\text{tr}_{13}^g(Ddf \otimes df)(P) = \text{tr}_{23}^g(Ddf \otimes df)(P)$ s'annule aussi, nous obtenons finalement l'identité suivante

$$\begin{aligned} & (df \otimes T_x(A_x(e^{2f_x}g)), D^2df_x)_g = \\ & (T(A_x(e^{2f_x}g)), df_x \otimes DA_x(g))_g - \frac{\alpha}{4} (d|W|^2, df_x)_g - e^{-4f_x} (d\psi, df_x)_g + 4\psi e^{-4f_x} |df_x|_g^2 . \end{aligned}$$

Elle permet, en rappelant que la fonction ψ est par hypothèse positive, d'écrire la majoration (3.13) sous la forme

$$\begin{aligned} 0 \geq & \frac{1}{2} (T_x(A_x(e^{2f_x}g)), Dd\gamma)_g = (T(A_x(e^{2f_x}g)), df_x \otimes DA_x(g))_g - \frac{\alpha}{4} (d|W|^2, df_x)_g - e^{-4f_x} (d\psi, df_x)_g \\ & + (T_x(A_x(e^{2f_x}g)), \text{tr}_{34}^g(D^2df_x \otimes df_x) - \text{tr}_{14}^g(D^2df_x \otimes df_x) + \text{tr}_{24}^g(Ddf_x \otimes Ddf_x)) . \end{aligned}$$

Par définition de la courbure (de la métrique g) et par symétrie du hessien

$$D^2df = \sigma_{23} D^2df = \sigma_{12} \circ \sigma_{23} D^2df - \sigma_{23} R^{(T^*M, g^{-1})}(df) ,$$

ce qui permet d'écrire la majoration précédente sous la forme

$$\begin{aligned} (3.14) \quad 0 \geq & (T_x(A_x(e^{2f_x}g)), R_g(\cdot, df_x^\sharp, df_x^\sharp))_g + (\frac{1-x}{2} \text{ric}_g(df_x^\sharp, df_x^\sharp) \text{tr}_g T_x(A_x(e^{2f_x}g))) \\ & + (T_x(A_x(e^{2f_x}g)), \text{tr}_{24} Ddf_x \otimes Ddf_x)_g + (T(A_x(e^{2f_x}g)), df_x \otimes DA_x(g))_g \end{aligned}$$

$$-\frac{\alpha}{4} (d|W|^2, df_x)_g - e^{-4f_x} (d\psi, df_x)_g .$$

Notons la minoration suivante du facteur quadratique en le hessien :

LEMME 3.1. — *Il existe une constante strictement positive ne dépendant que de la métrique g , $\epsilon(g)$, telle que pour tout x , $x \in [x_0, 1]$, on a*

$$(T_x(A_x(e^{2f_x}g)), \operatorname{tr}_{24}^g Ddf_x \otimes Ddf_x)_g \geq \epsilon \operatorname{tr}_g T(A(e^{2f_x}g)) |df_x|_g^4 .$$

La *proposition 1.19* de [LL] établit la majoration de l'énoncé, uniformément en x , $x_0 \leq x \leq 1$, dès qu'elle est satisfaite pour $x = 1$, ce dernier cas correspondant au *Lemma 2.4* de [GW] (avec $k = 2$). Ce lemme de Guan-Wang-Li-Li permet de réécrire la majoration (3.14) sous la forme

$$(3.15) \quad 0 \geq (T_x(A_x(e^{2f_x}g)), R_g(\cdot, df_x^{\sharp g}, \cdot, df_x^{\sharp g}))_g + \frac{1-x}{2} \operatorname{tr}_g T_x(A_x(e^{2f_x}g)) \operatorname{ric}_g(df_x^{\sharp g}, df_x^{\sharp g}) \\ + (T(A_x(e^{2f_x}g)), df_x \otimes DA_x(g))_g - \frac{\alpha}{4} (d|W|^2, df_x)_g - e^{-4f_x} (d\psi, df_x)_g + \epsilon \operatorname{tr}_g T(A_x(e^{2f_x}g)) |df_x|_g^4 .$$

Il nous reste à faire les remarques élémentaires suivantes :

1.— la trace de l'identité (3.8) s'écrivant

$$(3.16) \quad \operatorname{tr}_g T_x(A_x(e^{2f_x}g)) = (3 - 2x) \operatorname{tr}_g T(A_x(e^{2f_x}g)) ,$$

nous avons pour tout x , $x \leq 1$,

$$|T_x(A_x(e^{2f_x}g))| \leq |T(A_x(e^{2f_x}g))| + (1 - x) \operatorname{tr}_g T(A_x(e^{2f_x}g)) ,$$

où nous rappelons que $T(A)$ est donné par l'identité (3.7) ;

2.— l'identité de Newton appliquée à $\sigma_2(A)$ s'écrit

$$(3.17) \quad (\operatorname{tr}_g A_x(e^{2f_x}g))^2 - |A_x(e^{2f_x}g)|_g^2 = 2 \sigma_2(g^{-1}A_x(e^{2f_x}g)) = \frac{\alpha}{2} |W(g)|_g^2 + 2 e^{-4f_x} \psi_g > 0 ;$$

sous l'hypothèse $\sigma_2(A_x) > 0$, nous en déduisons la relation

$$(3.18) \quad |T(A_x(e^{2f_x}g))|_g \leq 3 \operatorname{tr}_g A_x(e^{2f_x}g) = \operatorname{tr}_g T(A_x(e^{2f_x}g)) .$$

La majoration dérivée au point 1 s'écrit donc sous cette hypothèse :

$$(3.19) \quad |T_x(A_x(e^{2f_x}g))|_g \leq (2 - x) \operatorname{tr}_g T(A_x(e^{2f_x}g)) \leq (2 - x_0) \operatorname{tr}_g T(A_x(e^{2f_x}g)) .$$

3.— en traçant la définition (3.2) de la courbure A_x nous obtenons la minoration

$$(3.20) \quad 0 < \operatorname{scal}(e^{2f_x}g) = \frac{3}{3 - 2x} e^{-2f_x} \operatorname{tr}_g A_x(e^{2f_x}g) ;$$

la minoration (3.17) entraînant la relation

$$(\operatorname{tr}_g A_x(e^{2f_x}g))^2 \geq 2 \psi_g e^{-4f_x} + \frac{\alpha}{2} |W(g)|_g^2 ,$$

nous en déduisons alors

$$(3.21) \quad \operatorname{tr}_g T(A_x(e^{2f_x}g)) = 3 \operatorname{tr}_g A_x(e^{2f_x}g) \geq 3 \sqrt{2} \max\left(-\frac{\sqrt{\alpha}}{2} |W(g)|_g, \sqrt{\psi_g} e^{-2f_x}\right) ;$$

4.— il suffit de différencier l'identité (3.2) relativement à la connexion canonique D pour obtenir la majoration banale suivante

$$(3.22) \quad |DA_x(g)|_g \leq |Dric(g)|_g + \frac{|x|}{3} |dscal(g)|_g \leq C(x_0) |Dric(g)|_g .$$

Les inégalités (3.18), (3.19) et (3.22) permettent ensuite de déduire de la relation (3.15) la majoration

$$\begin{aligned} & \epsilon(g) \operatorname{tr}_g T(A_x(e^{2f_x}g)) |df_x|_g^4 \\ & \leq \operatorname{tr}_g T(A_x(e^{2f_x}g)) (C(x, |R(g)|) |df_x|_g^2 + C(x, |DR(g)|) |df_x|_g) + \left(\frac{\alpha}{2} |W| |DW|_g + |d\psi|_g e^{-4f_x}\right) |df_x|_g . \end{aligned}$$

D'après la minoration (3.21) nous pouvons majorer le dernier facteur en terme de la trace $\operatorname{tr}_g T(A_x(e^{2f_x}g))$ pour établir l'inégalité suivante

$$\epsilon(g) |df_x|_g^4 \leq C(g) (|df_x|_g^2 + |df_x|_g) + \frac{1}{3} (\sqrt{\alpha} |DW|_g + \frac{|d\psi|}{\sqrt{\psi}} e^{-2f_x}) |df_x|_g .$$

La majoration uniforme de e^{-2f_x} sur $M \times [x_0, 1]$ établie au paragraphe précédent nous autorise finalement à conclure avec la majoration annoncée du $\sup_M |df_x|_g$, uniforme en $x \in [x_0, 1]$:

$$|df_x|_g^4 \leq C(g) (|df_x|_g^2 + |df_x|_g) \leq C_1(g) (|df_x|_g^2 + 1) \leq \frac{1}{2} |df_x|_g^4 + C_2(g) .$$

3.3.3. Majoration, borne C^2 uniforme et conclusion.— De la minoration uniforme en x , $x_0 \leq x < 3/2$ des solutions f_x des équations (3.3)_x et de la majoration uniforme en x , $x_0 \leq x \leq 1$, des différentielles df_x nous dérivons maintenant une estimation de Harnack uniforme pour les solutions f_x , $x_0 \leq x \leq x_1$, sous l'hypothèse – qui intervient ici pour la première fois, et de façon essentielle – que l'invariant conforme $C(g, x_1)$ (cf. (3.4)) est strictement positif.

L'identité élémentaire $\sigma_2(g^{-1}A_x) = \sigma_2(g^{-1}A) + \frac{3}{2}(1-x)(2-x)(\operatorname{tr}(g^{-1}A))^2$ permet, en invoquant l'équation (3.3)_x, d'écrire le terme $e^{-4f_x} \psi$ sous la forme

$$\begin{aligned} \sigma_2(g^{-1}A_x(e^{2f_x}g)) - \frac{\alpha}{4} |W(g)|_g^2 &= \sigma_2(g^{-1}A(e^{2f_x}g)) + \frac{3}{2}(1-x)(2-x) (\operatorname{tr}(g^{-1}A(e^{2f_x}g)))^2 - \frac{\alpha}{4} |W(g)|_g^2 \\ &= e^{4f_x} \left((\sigma_2(A(e^{2f_x}g)) + \frac{1}{6}(1-x)(2-x) \operatorname{scal}^2(e^{2f_x}g) - \frac{\alpha}{4} |W(e^{2f_x}g)|_{e^{2f_x}g}^2) \right) ; \end{aligned}$$

après intégration contre la forme volume de la métrique g , nous obtenons ainsi

$$(3.23) \quad \sup_M \psi \int_{(M, g)} e^{-4f_x} d\operatorname{vol} \geq \int_{(M, e^{2f_x}g)} \left(\sigma_2(A) - \frac{\alpha}{4} |W|^2 + \frac{(1-x)(2-x)}{6} \operatorname{scal}^2 \right) d\operatorname{vol} .$$

En rappelant la définition (3.5) de l'invariant de Yamabe $\mu(g)$, nous pouvons observer l'inégalité

$$0 < \mu(g) \sqrt{\operatorname{vol}(e^{2f_x}g)} \leq \int_{(M, e^{2f_x}g)} \operatorname{scal} d\operatorname{vol} \leq \sqrt{\operatorname{vol}(e^{2f_x}g)} \sqrt{\int_{(M, e^{2f_x}g)} \operatorname{scal}^2 d\operatorname{vol}}$$

que nous reformulons en la minoration suivante

$$(3.24) \quad \int_{(M, e^{2f_x}g)} \operatorname{scal}^2 d\operatorname{vol} \geq \mu(g)^2 .$$

Celle-ci permet de déduire de la minoration (3.23) la majoration suivante du minimum de f_x :

$$e^{-4 \inf f_x} \geq \frac{1}{\sup_M \psi_g \text{vol}(g)} \left(\int_M (\sigma_2(A) - \frac{\alpha}{4} |W|^2) d\text{vol} + \frac{(1-x)(2-x)}{6} \mu^2(g) \right),$$

et donc, sous la condition que l'invariant conforme C introduit précédemment (voir (3.4)) est strictement positif et en rappelant la minoration (3.11) du paragraphe précédent, uniforme en x , $x \in [x_0, 1]$, de conclure à la *majoration* suivante des solutions f_x de (3.3)_x, *uniforme* elle aussi en x , $x \in [x_0, 1]$:

$$(3.25) \quad \sup_M f_x \leq \inf_M f_x + \sup_M |df_x| \text{diam}(g) \leq C(g, \alpha).$$

Il reste à invoquer les estimations C^2 de [LL] (ou alternativement de [GV2]) pour déduire des estimations C^1 uniformes précédentes une majoration, uniforme en x , $x \in [x_0, x_1]$, $x_1 \leq 1$, du $\sup_M |\text{Hess} f_x|$; ces estimations reposent – dans [LL] comme dans [GV2] – sur la *concavité* de la fonctionnelle $\sqrt{\sigma_2(g^{-1}A_x(e^{2f_x}g))}$ considérée comme fonction du hessien de f , que l'on déduit facilement de la concavité bien connue de la fonction $\sqrt{\sigma_2}$ opérant sur le cône des endomorphismes symétriques vérifiant $\text{tr} a > 0$ et $\sigma_2(a) > 0$.

De l'équation (3.3)_x et de la majoration uniforme (3.25) des solutions f_x des équations (3.3)_x nous déduisons que $\sigma_2(g^{-1}A_x(e^{2f_x}g))$ est uniformément minoré (cf. (3.10)). De l'identité $2\sigma_2(g^{-1}A) = (\text{tr}(g^{-1}A))^2 - |g^{-1}A|^2 = (\text{tr}(g^{-1}A) - |g^{-1}A|)(\text{tr}(g^{-1}A) + |g^{-1}A|)$, et de la majoration uniforme de $|A(e^{2f_x}g)|$, conséquence des estimations C^2 uniformes et de l'identité (3.6), nous déduisons que l'expression $\text{tr}(g^{-1}A) - |g^{-1}A|$ admet un minorant strictement positif uniforme.

L'expression (3.7) définissant l'endomorphisme $T(A_x)$ assure qu'il est alors uniformément minoré par un multiple strictement positif de l'identité qui minore aussi, pour $x \leq 1$, les endomorphismes $T_x(A_x)$ d'après l'identité (3.8) les définissant, puisque la trace

$$\text{tr}_g T(A_x(e^{2f_x}g)) = e^{2f_x} \text{tr}_{e^{2f_x}g} T(A_x(e^{2f_x}g)) = 3 e^{2f_x} \text{tr}_{e^{2f_x}g} A_x(e^{2f_x}g) = (3 - 2x) e^{2f_x} \text{scal}(e^{2f_x}g)$$

est positive ou nulle (et même strictement positive) par hypothèse.

Au vu de l'expression (3.9) de la linéarisation de l'équation (3.3)_x, ceci établit l'*uniforme ellipticité* de ces équations, pour $x \in [x_0, x_1]$. La théorie classique de N.V. Krylov et C. Evans pour les équations concaves uniformément elliptiques assure alors l'existence d'une borne uniforme $C^{2,\beta}$, pour tout $\beta \in [0, 1)$.

La minoration uniforme de $\sigma_2(A_x(e^{2f_x}g))$ passe à la limite pour une convergence C^k , $k \geq 2$ et entraîne, par les identités banales $\text{scal}(e^{2f_x}g) = \frac{3}{3-2x} \text{tr}_{e^{2f_x}g} A_x(e^{2f_x}g)$, obtenue en traçant (3.2), et $\sigma_2(A_x(e^{2f_x}g)) = (\text{tr}_{e^{2f_x}g} A_x(e^{2f_x}g))^2 - |A_x(e^{2f_x}g)|_{e^{2f_x}g}^2$, l'existence d'un minorant strictement positif uniforme pour les courbures scalaires $\text{scal}(e^{2f_x}g)$, $x \in [x_0, x_1]$, qui vaut encore à la limite pour une convergence C^k , $k \geq 2$, établissant ainsi la fermeture de \mathcal{S} .

L'ensemble \mathcal{S} , qui est ouvert, fermé et non vide dans l'intervalle $[x_0, x_1]$ lui est donc égal, ce qui établit l'énoncé suivant

THÉORÈME 3.2. — Soit (M, g) une variété riemannienne compacte sans bord de dimension 4 dont l'invariant de Yamabe $\mu(g)$ (cf. (3.5)) est strictement positif. Pour tous réels α , $\alpha \in \mathbb{R}_+$, et x_1 , $x_1 \in (-\infty, 1]$, pour lesquels l'invariant conforme

$$\int_{(M, g)} \left(\sigma_2(A) - \frac{\alpha}{4} |W|^2 \right) d\text{vol} + \frac{1}{6} (1 - x_1) (2 - x_1) \mu(g)^2$$

est strictement positif, et pour tout réel x , $x \in (-\infty, x_1]$, il existe dans la classe conforme de g une métrique de courbure scalaire strictement positive pour laquelle le polynôme $\sigma_2(A_x) - \frac{\alpha}{4} |W|^2$ est (partout) strictement positif.

Le cas $\alpha = 0$ de cet énoncé est le théorème principal de [GV].

Le cas $\alpha = 1$ et $x = x_1 = 1$ établit la réduction à [M] du théorème 1 proposée dans le préliminaire.

3.4. Une autre application du théorème 3.2 : construction de métriques de Q -courbure constante

Si le grand mérite de l'approche précédente consiste à s'affranchir du recours à la régularisation par le terme elliptique du quatrième ordre $\delta\Delta \text{scal}$, elle n'en conduit pas moins à un intéressant résultat d'existence de métriques de Q -courbure constante sous des hypothèses suffisamment souples pour être satisfaites par de nombreux exemples. La Q -courbure dont il s'agit ici est celle de la paire conforme (P, Q) associée à l'opérateur de Paneitz, (cf. (2.2) et (2.3) pour les définitions), un opérateur du quatrième ordre donc, dont le contenu géométrique a été discuté dans la partie 2.1. Rappelons que la caractéristique principale de la fonctionnelle I_2 – essentiellement la forme quadratique associée à l'opérateur formellement auto-adjoint P translatée par le potentiel Q (cf. (2.4)) – est d'admettre pour *points critiques* les métriques de Q -courbure constante (cf. (2.5)) et (2.6)).

Dans [CY], Chang S.-Y. A. et Yang P. établissent pour la fonctionnelle I_2 l'analogue suivant du théorème 2.1 ; la démonstration de ce résultat est parallèle à celle du Théorème 2.1 exposée précédemment et repose, comme celle-ci, sur l'inégalité de Moser-Trudinger-Adams (2.10).

THÉORÈME 3.3 ([CY], Theorem 1.2). — Sur une variété riemannienne compacte sans bord de dimension 4, (M, g) , le minimum de la fonctionnelle I_2 (cf. (2.4)) est atteint par une métrique de Q -courbure constante dès que l'opérateur de Paneitz P est positif ou nul et de noyau réduit aux fonctions constantes et que l'invariant conforme $\int_M \sigma_2(A) d\text{vol} = 2 \int_M Q d\text{vol}$ est strictement inférieur à celui de la sphère ronde (S^4, can) , égal à $16\pi^2$.

Dans [G2] M. Gursky démontre ensuite que la seconde hypothèse est satisfaite pour toute variété riemannienne compacte et sans bord de dimension 4 de courbure scalaire strictement positive qui n'est pas conforme à la sphère standard, et que la première l'est dès que la courbure scalaire et l'invariant conforme $\int_M \sigma_2(A), d\text{vol}$ sont positifs ou nuls, établissant ainsi, comme corollaire du théorème 3.3, l'existence de métriques de Q -courbure constante sous ces seules hypothèses.

Comme corollaire du cas particulier $\alpha = x_1 = 0$ du théorème 3.2, M. Gursky et J. Viaclovsky établissent l’extension substantielle suivante de ce résultat.

THÉORÈME 3.4 (cf. [GV], Theorem 1.4). — *Il existe une métrique de Q -courbure constante dans la classe conforme de toute variété riemannienne compacte sans bord de dimension 4 dont la courbure scalaire et l’invariant conforme $\int_M Q \, d\text{vol} + 1/6 \mu(g)^2$ sont strictement positifs.*

Comme l’indique la discussion précédente, ces métriques sont obtenues comme minima de la fonctionnelle I_2 , l’existence de ces derniers étant garantie par le théorème 3.3 dès que l’opérateur de Paneitz est positif ou nul et de noyau réduit aux constantes. Puisque $2 \int_M Q \, d\text{vol} = \int_M \sigma_2(A) \, d\text{vol}$, l’hypothèse correspond au cas particulier $\alpha = x_1 = 0$ du théorème 3.2 qui assure alors l’existence dans la classe conforme d’une métrique, h , pour laquelle la courbure scalaire et $\sigma_2(A_0)$ sont strictement positifs. Des identités $A_0 = \text{ric}$ (cf. la définition (3.2) de A_x) et $\sigma_2(A_0) = 1/2 (\text{tr } A_0)^2 - 1/2 |A_0|^2$, et de la stricte positivité de $\sigma_2(A_0)$ et de la courbure scalaire nous déduisons la majoration

$$(3.26) \quad \text{ric}(h) < \text{scal}(h) h .$$

(On peut aussi vérifier facilement la minoration $\text{ric}(h) > -1/2 \text{scal}(h) h$, intéressante dans l’absolu, mais sans conséquence pour notre argument.) Il reste à observer que la minoration spectrale de l’opérateur de Paneitz à laquelle nous avons réduit la preuve du théorème 3.4 est automatique dans ce cas : de l’expression (2.2) de l’opérateur de Paneitz nous déduisons immédiatement

$$(3.27) \quad \int_M f P f \, d\text{vol} = \int_M ((\Delta f)^2 - 2 \text{ric}(df^\sharp, df^\sharp) + \frac{2}{3} \text{scal} |df|^2) \, d\text{vol} .$$

En contractant contre la 1-forme df , et en intégrant sur M l’identité de Bochner appliquée à df , qui s’écrit

$$\Delta f = dd^*df = D^*Ddf + \text{ric}(df^\sharp, .),$$

nous dérivons l’identité

$$\int_M ((\Delta f)^2 - \text{ric}(df^\sharp, df^\sharp) - |\text{Hess}f|^2) \, d\text{vol} = 0 ,$$

que nous reportons dans (3.27) pour obtenir finalement

$$(3.28) \quad \int_M f P f \, d\text{vol} = \frac{1}{3} \int_M (4 |\text{Hess}_0 f|^2 + 2 (\text{scal} g - \text{ric})(df^\sharp, df^\sharp)) \, d\text{vol},$$

où nous avons noté $\text{Hess}_0 f := \text{Hess}f + 1/4 \Delta f g$ la composante de trace nulle du hessien.

L’intégrale $\int_M f P f \, d\text{vol}$ est donc positive ou nulle dès que $\text{ric}(h) \leq \text{scal}(h) h$. Si, de plus, $\int_M f P f \, d\text{vol} = 0$, nous déduisons de la même identité (3.28) que la composante de trace nulle du hessien, $\text{Hess}_0 f$, est identiquement nulle. Un résultat classique de M. Obata assure alors que f est constante, ou que (M, h) est de courbure sectionnelle constante ; dans ce dernier cas l’identité (3.28) devient $\int_M f P f \, d\text{vol} = 1/3 \int_M (4 |d\text{Hess}_0 f|^2 + 3/2 \text{scal} |df|^2) \, d\text{vol}$, ce qui établit que, cette fois encore, f est nécessairement constante, et conclut la démonstration de

cette *minoration spectrale* de l'opérateur de Paneitz sous l'hypothèse $\text{ric}(h) \leq \text{scal}(h)h$, et donc la preuve du Théorème 3.4.

Toute métrique de courbure de Ricci positive ou nulle satisfaisant évidemment l'hypothèse de pincement de la courbure de Ricci (3.26), la minoration spectrale de l'opérateur de Paneitz que nous venons d'établir permet d'invoquer le théorème 3.3 pour construire des métriques de Q -courbure constante dans la classe conforme de toute telle métrique. Dans [SY], Sha J.-P. et Yang D. en démontraient l'existence sur toute variété de dimension 4 (à homéomorphisme près) admettant une métrique de courbure scalaire strictement positive – c'est-à-dire, concrètement, sur les sommes connexes $a(S^2 \times S^2)$ et $(a+b)\mathbb{P}_{\mathbb{C}}^2 \# b\mathbb{P}_{\mathbb{C}}^2$, a, b entiers naturels.

En invoquant le théorème 3.4 nous allons élargir la classe des variétés dont on sait qu'elles admettent des métriques de Q -courbure constante, en nous affranchissant par exemple de l'hypothèse de simple-connexité requise dans les constructions de Sha J.P. et Yang D. .

COROLLAIRE 3.5 (cf. [GV], Theorem 7.1). — *Soient a et b deux entiers naturels, $(M_i, g_i)_{i=1,2}$ deux variétés riemanniennes compactes sans bord de dimension 4, et $\mu(g_i)$ leurs invariants de Yamabe (cf. (3.5)). Les variétés $a(S^1 \times S^3) \# b\mathbb{P}_{\mathbb{R}}^4$, $M_1 \# a(S^1 \times S^3)$, $M_1 \# b\mathbb{P}_{\mathbb{R}}^4$ et $M_1 \# M_2$ admettent des métriques de Q -courbure constante dès que, respectivement,*

1. $2a + b \leq 9$, pour la première famille ;
 2. $\mu(g_1) - 4\sqrt{3}a\pi > 0$, $a \in \{0, \dots, 7\}$ et $\int_{(M_1, g_1)} Q d\text{vol} \geq 0$, pour la seconde famille ;
 3. $\mu(g_1) - 8\sqrt{3}\pi > 0$, $b \in \{0, \dots, 8\}$ et $\int_{(M_1, g_1)} Q d\text{vol} \geq 0$, pour la troisième ;
- ou
4. $\mu(g_i) - 4\sqrt{3}\pi > 0$ et $\int_{(M_i, g_i)} Q d\text{vol} \geq 0$, $i \in \{1, 2\}$, dans le dernier cas.

Les sommes connexes $S^2 \times S^2 \# a(S^1 \times S^3)$, $a \leq 5$; $\mathbb{P}_{\mathbb{C}}^2 \# a(S^1 \times S^3)$, $a \leq 5$; $\mathbb{P}_{\mathbb{C}}^2 \# b\mathbb{P}_{\mathbb{R}}^4$, $b \leq 8$; et $\mathbb{P}_{\mathbb{C}}^2 \# c\bar{\mathbb{P}}_{\mathbb{C}}^2 \# (S^1 \times S^3)$, $c \in \{3, \dots, 7\}$, sont ainsi des exemples de variétés admettant des métriques de Q -courbure constante.

D'après le théorème 3.4, il suffit pour établir ce corollaire de construire sur les variétés de l'énoncé des métriques de courbure scalaire strictement positive satisfaisant de plus à l'hypothèse $\int_M Q d\text{vol} = \frac{1}{2} \int_M \sigma_2(A) d\text{vol} > 1/6 \mu(g)^2$.

Rappelons pour cela les constructions élémentaires suivantes

LEMME 3.6 (cf. [K], Lemma 3.2, ou [ABKS], Proposition 4.1). — *Pour tout $\epsilon > 0$, toute variété riemannienne compacte et sans bord, (M, g) , de dimension, n , supérieure ou égale à 3, et tout point P de M , il existe une métrique \tilde{g} telle que*

1. \tilde{g} est localement conformément plate (i.e. $W(\tilde{g}) \equiv 0$ si $n \geq 4$) sur un voisinage de P ;
2. $|\mu(\tilde{g}) - \mu(g)| < \epsilon$;

et

$$3. \left| \int_{(M,g)} |W|^2 d\text{vol} - \int_{(M,\tilde{g})} |W|^2 d\text{vol} \right| < \epsilon.$$

LEMME 3.7 (cf. [K] Theorem 2). — Soient $\epsilon > 0$ et $(M_i, g_i)_{i=1,2}$ deux variétés riemanniennes compactes sans bord de même dimension supérieure ou égale à 3, de courbure scalaire partout strictement positive, et localement conformément plates au voisinage d'un de leurs points $P_i \in M_i$, $i = 1, 2$. Il existe alors sur la somme connexe $M_1 \# M_2$ une métrique g_ϵ telle que

$$1. \quad \mu(M_1 \# M_2, g_\epsilon) \geq \min_{i \in \{1,2\}} \{\mu(g_i)\} - \epsilon, \text{ et,}$$

$$2. \quad \int_{(M_1 \# M_2, g_\epsilon)} |W|^2 d\text{vol} = \int_{(M_1, g_1)} |W|^2 d\text{vol} + \int_{(M_2, g_2)} |W|^2 d\text{vol}.$$

En particulier, si $\mu(M_i) := \sup_{g \in \mathcal{M}(M_i)} \mu(g)$ est positif ou nul, $\mu(M_1 \# M_2) \geq \min_{i \in \{1,2\}} \mu(M_i)$.

Appliqué aux cylindres $S^{n-1} \times \ell I$, $\ell \in \mathbb{R}_+^*$, recollés en deux points de la sphère standard, l'argument conduisant au lemme précédent démontre aussi

COROLLAIRE 3.8. — Pour tout entier naturel a ,

$$\mu(a(S^{n-1} \times S^1)) = \mu(S^{n-1} \times S^1) = \mu(S^n).$$

La démonstration de ces résultats élémentaires est sans surprise et n'a rien à voir avec les idées discutées dans ce texte ; [K] étant rédigé avec beaucoup de soin, nous y renvoyons le lecteur que ces énoncés laisseraient perplexes.

Reprenons la démonstration du corollaire 3.5 : de la formule de Gauss-Bonnet-Chern (1.3) et de l'hypothèse $\int_M \sigma_2(A) d\text{vol} \geq 0$, nous déduisons la majoration $\int_{(M_i, g_i)} |W|^2 d\text{vol} \leq 32 \pi^2 \chi(M)$. Pour tout $\epsilon, \epsilon > 0$, le lemme 3.6 permet de modifier la métrique g_i au voisinage d'un point arbitraire P_i de M_i pour l'y rendre conformément plate sans modifier de plus de ϵ ni l'invariant de Yamabe $\mu(g_i)$, ni $\int_{M_i} |W|^2 d\text{vol}$. Notons \tilde{g}_i les métriques ainsi obtenues.

Considérons pour commencer la seconde famille, $M \# a(S^1 \times S^3)$: d'après le corollaire 3.8, il existe sur $a(S^1 \times S^3)$ une métrique localement conformément plate, h_ϵ , telle que $\mu(h_\epsilon) \geq \mu(a(S^1 \times S^3)) - \epsilon = \mu(S^4) - \epsilon$; le lemme 3.7 entraîne alors l'existence d'une métrique \tilde{g} sur $M \# a(S^1 \times S^3)$ telle que, d'une part

$$1. \quad \mu(\tilde{g}) \geq \min(\mu(h_\epsilon), \mu(\tilde{g})) - \epsilon \geq \mu(\tilde{g}) - \epsilon \geq \mu(g) - 2\epsilon,$$

où nous utilisons la majoration $\mu(g) \leq \mu(S^n, \text{can})$ due à T. Aubin et valable pour n'importe quelle métrique sur n'importe quelle variété compacte sans bord de dimension n , le cas d'égalité étant caractéristique – c'est le cœur de la solution du problème de Yamabe – de la sphère conforme standard. Cette majoration intervient d'ailleurs de façon implicite dans l'énoncé du corollaire 3.8. Et, de l'autre

$$2. \quad \int_{(M \# a(S^1 \times S^3), \tilde{g})} |W|^2 d\text{vol} = \int_{(M, \tilde{g})} |W|^2 d\text{vol} \leq \int_{(M, g)} |W|^2 d\text{vol} + \epsilon.$$

En posant $N_a = M \# a(S^1 \times S^3)$, nous déduisons des deux inégalités précédentes et de la formule de Gauss-Bonnet-Chern appliquée successivement à N_a et à M , la minoration

$$\int_{(N_a, \tilde{g})} Q \, d\text{vol} + \frac{1}{6} \mu(\tilde{g})^2 = 4\pi^2 \mathcal{X}(N_a) - \frac{1}{8} \int_{(N_a, \tilde{g})} |W|^2 \, d\text{vol} + \frac{1}{6} \mu(\tilde{g})^2 \geq$$

$$4\pi^2 \mathcal{X}(N_a) - \frac{1}{8} \int_{(M, \tilde{g})} |W|^2 \, d\text{vol} + \frac{1}{6} (\mu(\tilde{g}) - \epsilon)^2 \geq 4\pi^2 (\mathcal{X}(N_a) - \mathcal{X}(M)) + \frac{1}{6} (\mu(g) - 2\epsilon)^2 - \frac{\epsilon}{8},$$

cette dernière expression étant strictement positive pour un choix convenable de ϵ dès que $\mu(g) > 4\sqrt{3}a\pi$. Ceci achève la discussion du second point du corollaire 3.5, la borne sur a provenant de la majoration de la constante de Yamabe $\mu(g) \leq \mu(S^n, \text{can}) = 8\sqrt{6}\pi$ rappelée ci-dessus : elle entraîne clairement $a < 8$.

Pour la troisième famille on procède identiquement : $\mathbb{P}_{\mathbb{R}}^4$ est, comme $S^1 \times S^3$, localement conformément plat. L'hypothèse $\mu(g) > 8\sqrt{3}\pi = \mu(\mathbb{P}_{\mathbb{R}}^4, \text{can})$ entraîne ici que le minimum $\min(\mu(g), \mu(\mathbb{P}_{\mathbb{R}}^4, \text{can}))$ est atteint par $\mathbb{P}_{\mathbb{R}}^4$; pour tout $\epsilon > 0$, le lemme 4 garantit alors l'existence d'une métrique \tilde{g} sur $N^a := M \# a\mathbb{P}_{\mathbb{R}}^4$ dont l'invariant de Yamabe est minoré par $8\sqrt{3}\pi - \epsilon$ et telle que

$$\int_{(N^a, \tilde{g})} |W|^2 \, d\text{vol} = \int_{(M, \tilde{g})} |W|^2 \, d\text{vol} \leq \int_{(M, g)} |W|^2 \, d\text{vol} + \epsilon.$$

On conclut alors comme précédemment.

Pour la première famille $N_{ab} := a(S^1 \times S^3) \# b\mathbb{P}_{\mathbb{R}}^4$ de l'énoncé, on reprend l'argument précédent. La première étape est même dans ce cas inutile, puisque la construction du lemme 3.7 permet de choisir la métrique g_ϵ localement conformément plate dès que g_1 et g_2 le sont. D'après le corollaire 3.8 et la majoration de $\mu(g)$ par $\mu(S^4)$,

$$\mu(a(S^{n-1} \times S^1)) = \mu(S^n) \geq \mu(b\mathbb{P}_{\mathbb{R}}^4);$$

le lemme 3.7 assure alors l'existence d'une métrique localement conformément plate, \tilde{g} , dont l'invariant de Yamabe $\mu(\tilde{g})$ est minoré par $\mu(b\mathbb{P}_{\mathbb{R}}^4) - \epsilon \geq \mu(\mathbb{P}_{\mathbb{R}}^4) - \epsilon$. Nous concluons dans ce cas par la minoration suivante

$$\int_{(N_{ab}, \tilde{g})} Q \, d\text{vol} + \frac{1}{6} \mu(\tilde{g})^2 = 4\pi^2 \mathcal{X}(N_{ab}) - \frac{1}{8} \int_{(N_{ab}, \tilde{g})} |W|^2 \, d\text{vol} + \frac{1}{6} \mu(\tilde{g})^2 \geq$$

$$4\pi^2 (\mathcal{X}(a(S^1 \times S^3)) + \mathcal{X}(b\mathbb{P}_{\mathbb{R}}^4) - 2) + \frac{1}{6} \mu(\mathbb{P}_{\mathbb{R}}^4) - \epsilon = 4\pi^2(10 - 2a - b) - \epsilon,$$

le dernier terme de cette identité étant strictement positif pour un choix adéquat de ϵ dès que les entiers naturels a, b vérifient la relation $2a + b \leq 9$.

Pour le cas restant, on applique le lemme 3.6 aux deux points arbitraires P_1 (P_2) de M_1 (resp. M_2) pour obtenir des métriques $\tilde{g}_i, i = 1, 2$, égales aux métriques g_i loin des points P_i , et localement conformément plates au voisinage des points $P_i, i = 1, 2$, les invariants conformes $\mu(\tilde{g}_i)$ et $\int_{(M_i, \tilde{g}_i)} |W|^2 \, d\text{vol}$ coïncidant – à ϵ près – avec ceux des métriques $g_i, i = 1, 2$. Du lemme 3.7 nous déduisons alors l'existence d'une métrique \tilde{g} sur $M_1 \# M_2$ telle que

1.
$$\int_{M_1 \# M_2} |W|^2 \, d\text{vol} = \sum_{i=1,2} \int_{(M_i, \tilde{g}_i)} |W|^2 \, d\text{vol} \leq \sum_{i \in \{1,2\}} \int_{(M_i, g_i)} |W|^2 \, d\text{vol} + 2\epsilon,$$

et

$$2. \quad \mu(M_1 \# M_2, \tilde{g}) \geq \min_{i \in \{1,2\}} \{\mu(M_i, g_i)\} - \epsilon .$$

Comme précédemment, on invoque la formule de Gauss-Bonnet-Chern pour conclure

$$\begin{aligned} \int_{(M_1 \# M_2, \tilde{g})} Q \, d\text{vol} + \frac{1}{6} \mu(\tilde{g}) &= 4\pi^2 \mathcal{X}(M_1 \# M_2) - \frac{1}{8} \int_{(M_1 \# M_2, \tilde{g})} |W|^2 \, d\text{vol} + \frac{1}{6} \mu(\tilde{g})^2 \geq \\ 4\pi^2 (\mathcal{X}(M_1) + \mathcal{X}(M_2) - 2) - \frac{1}{8} \int_{M_1} |W|^2 \, d\text{vol} - \frac{1}{8} \int_{M_2} |W|^2 \, d\text{vol} + \frac{1}{6} (\min_{i \in \{1,2\}} \mu(g_i))^2 - C\epsilon &\geq \\ -8\pi^2 + \frac{1}{6} (\min_{i \in \{1,2\}} \mu(g_i))^2 - C\epsilon . \end{aligned}$$

Sous l’hypothèse $\mu(g_i) > 4\sqrt{3}\pi$, cette dernière expression est strictement positive pour un choix pertinent de ϵ , ce qui conclut la preuve de l’énoncé principal du corollaire 3.5 .

Notons pour finir que les exemples proposés vérifient l’hypothèse de l’énoncé : c’est banal, sauf peut-être pour le dernier pour lequel nous rappelons que $\mathbb{P}^2 \#_c \bar{\mathbb{P}}^2$, $c \in \{3, \dots, 8\}$ admet une métrique de Kähler-Einstein g_c pour laquelle $\mu(g_c) = 4\pi\sqrt{18 - 2c}$, (cf. [G1]).

4. LE CAS LIMITE : $\int_M (\sigma_2(A) - |W|^2/4) \, d\text{vol} = 0$

On se ramène sans douleur au cas où $\int_M |W|^2 \, d\text{vol}$ est strictement positif en rappelant le résultat de M. Gursky (cf. [G1]) d’après lequel toute variété de dimension 4 conformément plate de courbure scalaire strictement positive et de caractéristique d’Euler nulle (si $\int_M \sigma_2(A) \, d\text{vol} = \frac{1}{4} \int_M |W|^2 = 0$, la caractéristique d’Euler de M , $\mathcal{X}(M)$, s’annule nécessairement d’après la formule de Chern (1.3)) est conforme à un quotient compact du produit riemannien $(S^3, \text{can}) \otimes \mathbb{R}$ par un sous-groupe d’isométries. Pour tout $\alpha < 1$, $\int_M (\sigma_2(A) - \alpha/4 |W|^2) \, d\text{vol} = (1 - \alpha)/4 \int_M |W|^2 \, d\text{vol}$ est alors strictement positif et il existe donc dans la classe conforme de g , une métrique g_α pour laquelle $\sigma_2(A) - \alpha/4 |W|^2$ est strictement positif. Malheureusement les estimations a priori dérivées précédemment dépendent lourdement du paramètre $\int_M (\sigma_2(A) - \alpha/4 |W|^2) \, d\text{vol}$, qui tend ici vers 0 lorsque α approche 1 ; nous devons donc les reprendre avant d’espérer conclure à l’existence d’une métrique pour laquelle le polynôme $4\sigma_2(A) - |W|^2$ serait partout positif ou nul (et donc nul sous notre hypothèse) en prenant la limite d’une suite adéquate de métriques g_{α_k} , $\alpha_k \rightarrow 1$.

Pour ce faire, Chang S.-Y. A., M. Gursky et Yang P. commencent par associer à chaque $\alpha < 1$ une solution g_α « canonique » du problème précédent, telle que le polynôme $\sigma_2(A) - \alpha/4 |W|^2$ est non seulement strictement positif, mais aussi *constant*, une sorte de *jauge*, associée à un problème de Yamabe pour l’invariant quadratique $\sigma_2(A) - \alpha/4 |W|^2$. On démontre pour cela le résultat plus général suivant

4.1. Le problème de Yamabe pour un polynôme quadratique en la courbure de Ricci :
l'équation $4\sigma_2(A) - \alpha|W|^2 = \varphi$

THÉORÈME 4.1. — Soit α un réel positif ou nul. Sur une variété riemannienne compacte sans bord de dimension 4, (M, g) , qui n'est pas conforme à la sphère standard et pour laquelle $\int_M (\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol}$ et la courbure scalaire sont strictement positifs, il existe, pour toute fonction infiniment différentiable et strictement positive φ , une métrique dans la classe conforme de g pour laquelle

$$(4.1) \quad \sigma_2(A) - \frac{\alpha}{4}|W|^2 = \varphi .$$

Les solutions $e^{2f_\alpha}g$ de (4.1) vérifient de plus l'estimation a priori uniforme

$$(4.2) \quad \sup(|df_\alpha| + e^{f_\alpha}) \leq C ,$$

où la constante C ne dépend que de (M, g) , et de la norme C^2 et du minimum de la fonction φ .

Le cas particulier $\alpha = 0$ de ce résultat est l'objet de l'article [CGY2] auquel le lecteur de [CGY0] est renvoyé en guise de preuve. L'énoncé de Chang S.-Y. A., M. Gursky et Yang P. est en fait plus fort que celui du théorème 6 en ce qu'il affirme une majoration indépendante du minimum de φ . Ceci présente l'inconvénient de faciliter la discussion du cas limite, mais aussi l'inconvénient de conduire à la

Contradiction : Le couple (*Theorem 1.1, Proposition 1.7*) de [CGY0], tout comme son cas particulier (*Main Theorem, Corollary B*) objet de [CGY2], est *contradictoire* : soient α un réel positif, (M, g) une variété riemannienne compacte sans bord de dimension 4 de courbure scalaire strictement positive qui n'est pas conformément équivalente à la sphère standard (S^4, can) et telle que $\int_M (\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} > 0$. Notons qu'il existe de telles variétés : $(\mathbb{P}_{\mathbb{C}}^2, F\text{-S})$, satisfait par exemple les hypothèses précédentes pour tout $\alpha, 0 \leq \alpha < 1$. D'après [CGY0] il existe une solution f à l'équation $(\sigma_2(A) - \frac{\alpha}{4}|W|^2)(e^{2f}g) \equiv 1$; l'homogénéité du polynôme $4\sigma_2(A) - \alpha|W|^2$ entraîne par ailleurs que les métriques $e^{2(f+a)}g$, $a \in \mathbb{R}$, satisfont alors les équations $4\sigma_2(A) - \alpha|W|^2 = e^{-4a}$. Pour $a \in \mathbb{R}_+$, les normes C^2 des fonctions (constantes) e^{-4a} sont majorées par 1, et d'après la *proposition 1.7* de [CGY0] il existerait une constante majorant $\sup_M (e^{f+a} + |df|)$ uniformément en $a \geq 0$.

Cette petite remarque démontre de même que la *proposition 1.7* de [CGY0] et le *Theorem A* de [CGY1] sont contradictoires.

On s'en tiendra donc à l'énoncé proposé ici. Sa démonstration repose sur l'énoncé du Théorème 1, d'après lequel il existe une métrique dont le polynôme de courbure $\sigma_2(A) - \frac{\alpha}{4}|W|^2$ est strictement positif dans la classe conforme de toute métrique vérifiant les hypothèses de positivité énoncées. On procède en deux étapes, de natures très différentes.

4.1.1. *Majoration C^2 a priori pour les solutions de l'équation $\sigma_2(A) - \frac{\alpha}{4}|W|^2 = \varphi$.* —

PROPOSITION 4.2. — Soient (M, g) une variété riemannienne compacte sans bord de dimension 4 de courbure scalaire strictement positive qui n'est pas conforme à la sphère standard (S^4, can) , φ une fonction infiniment différentiable et strictement positive sur M , et (a, b, c)

trois nombres réels positifs ou nuls. Toute solution f de $(\sigma_2(A) - \frac{\alpha}{4}|W|^2)(e^{2f}g) = \varphi$, telle que $|\varphi|_{C^2} \leq a$, $\inf \varphi \geq b$ et $0 \leq \alpha \leq c$ vérifie les majorations a priori $\sup(|df_\alpha| + e^{f_\alpha}) \leq C(a, b, c, g)$ et $|f|_{C^2} \leq C(a, b, c, \alpha, g)$.

Pour nous en convaincre commençons par l'estimation a priori locale d'après laquelle toute solution de courbure scalaire strictement positive de l'équation $\sigma_2(A) - \frac{\alpha}{4}|W|^2 = \varphi$ sur une boule de rayon ρ , $e^{2f}g$, vérifie la majoration a priori $|\text{Hess}f|_{B(P, \rho/2)} \leq C$, où la constante C ne dépend que du rayon ρ , des normes C^2 de la métrique et de φ , ainsi que des normes sup de $|f|$ et de $|df|$, toute ces normes étant prises sur la boule $B(P, \rho)$. Il s'agit là d'une version « localisée » d'estimations que nous avons discutées précédemment dans leur version globale (cf. la partie 2.4, en particulier les paragraphes 2.4.1 et 2.4.2). À partir de l'identité

$$(4.3) \quad (G, \text{Hess}(\text{scal})) = -3 \Delta \sigma_2(A) + 3(|D \text{ric}_0|^2 - \frac{1}{12}|d\text{scal}|^2) + 6 \text{tr ric}_0^3 + \text{scal} |\text{ric}_0|^2 - 6W(\text{ric}_0, \text{ric}_0) - 6(\text{ric}_0, B),$$

déjà implicitement utilisée pour dériver la majoration (2.16), et de son alter ego pour $(G, \text{Hess}|df|^2)$, on établit facilement l'inéquation différentielle

$$(G, \text{Hess}(\text{scal} + 12|df|^2)) \geq \frac{1}{48}(\text{scal} + 12|df|^2)^3 - C(\rho, |\varphi|_{C^2(B(P, \rho))}, |g|_{C^2(B(P, \rho))}, |f|_{C^1(B(P, \rho))}).$$

Par hypothèse $\sigma_2(A)$ est strictement positif et G , qui est minoré par $3\sigma_2(A)/\text{scal} g$, est donc défini positif; le principe du maximum appliqué à l'inéquation différentielle précédente conduit alors à une majoration de la courbure scalaire sur $B(P, \rho/2)$ en terme de la constante figurant dans l'inéquation, et donc des seuls paramètres dont celle-ci dépend. L'expression de la courbure scalaire dans une classe conforme (2.13) permet finalement de conclure que les dérivées secondes sont uniformément bornées :

$$(4.4) \quad \sup_M |\text{Hess} f| \leq C(\rho, |\varphi|_{C^2(B(P, \rho))}, |g|_{C^2(B(P, \rho))}, |f|_{C^1(B(P, \rho))})$$

4.1.1.1. *L'éclatement.* — Reprenons la preuve de la proposition 4.2 et démontrons, sous les hypothèses de l'énoncé, que l'expression $e^f + |df|^2$ est uniformément majorée : nous aurions sinon une suite de réels positifs ou nuls α_k , une suite φ_k de fonctions uniformément minorées, $\varphi_k \geq a > 0$, et de normes C^2 uniformément majorées, $|\varphi_k|_{C^2} \leq b$, une suite, f_k , de solutions du système $(\sigma_2(A) - \frac{\alpha_k}{4}|W|^2)(e^{2f_k}g) = \varphi_k$ et $\text{scal}(e^{2f_k}g) > 0$, et une suite de points, P_k , de M , tels que la suite

$$(|df_k| + e^{f_k})(P_k) = \sup_M (|df_k| + e^{f_k}) := \epsilon_k^{-1}$$

diverge lorsque k tend vers l'infini. Choisissons un réel positif, r_0 , inférieur au rayon d'injectivité de (M, g) et identifions les boules $B(P_k, r_0)$ avec « la » boule euclidienne $B(r_0)$ via l'exponentielle de la métrique g . En notant, pour tout ϵ strictement positif, h_ϵ l'homothétie de rapport ϵ , introduisons pour toute fonction, f , sur la boule $B(r_0)$ son ϵ -renormalisation, définie sur la boule $B(r_0/\epsilon)$ par l'identité $f_\epsilon = f \circ h_\epsilon + \log \epsilon$; pour tout entier positif k , la fonction f_{k, ϵ_k} est alors définie sur la boule $B(r_0/\epsilon_k)$ et vérifie sur cette boule la majoration

$$(4.5) \quad |df_{k, \epsilon_k}| + e^{f_{k, \epsilon_k}} \leq 1.$$

Par construction nous avons aussi l'identité

$$(4.6) \quad (|df_{k,\epsilon_k}| + e^{f_{k,\epsilon_k}})(0) = 1 .$$

Introduisons encore les métriques $\hat{g}_k = e^{2f_{k,\epsilon_k}} g \circ h_{\epsilon_k}$ et $\hat{g}_k = g \circ h_{\epsilon_k}$, définies sur la boule $B(r_0/\epsilon_k)$; la suite (\hat{g}_k) converge clairement vers la métrique euclidienne (en norme C^ℓ , ℓ entier arbitraire, et uniformément sur tout compact de \mathbb{R}^4), tandis que l'équation $\hat{g} = h_{\epsilon_k}^*(e^{2f_k} g)$ entraîne par « naturalité » et homogénéité du polynôme $\sigma_2(A) - \frac{\alpha}{4} |W|^2$ l'identité

$$(4.7) \quad \sigma_2(A(\hat{g}_k)) - \frac{\alpha_k}{4} |W(\hat{g}_k)|^2 = (\sigma_2(A) - \frac{\alpha_k}{4} |W|^2) (e^{2f_k} g) \circ h_{\epsilon_k} = \varphi_k \circ h_{\epsilon_k} .$$

Si $\lim_{k \rightarrow +\infty} e^{f_{k,\epsilon_k}(0)} = 0$, nous introduisons la suite de métriques \check{g}_k , définies sur $B(r_0/\epsilon_k)$ par $\check{g}_k := e^{2(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))} \hat{g}_k$. En invoquant à nouveau l'homogénéité du polynôme $\sigma_2(A) - \frac{\alpha_k}{4} |W|^2$, l'équation $\check{g}_k = e^{-2f_{k,\epsilon_k}(0)} \hat{g}_k$ permet d'écrire l'identité (4.7) sous la forme

$$(4.8) \quad \sigma_2(A(\check{g}_k)) - \frac{\alpha_k}{4} |W(\check{g}_k)|^2 = e^{4f_{k,\epsilon_k}(0)} \varphi_k \circ h_{\epsilon_k} .$$

Pour tout R positif, et tout k assez grand pour que $\epsilon_k < r_0/R$, nous avons encore par construction

$$(4.9) \quad |d(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))| \leq 1 ,$$

$$(4.10) \quad (f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))(0) = 0 ,$$

et donc la majoration $\sup_{B(R)} |f_{k,\epsilon_k} - f_{k,\epsilon_k}(0)| \leq R$: sur tout compact de \mathbb{R}^4 , la suite $(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))$ est uniformément bornée dans la topologie C^1 . Nous avons aussi, toujours par construction

$$(4.11) \quad \lim_{k \rightarrow +\infty} |d(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))|(0) = \lim_{k \rightarrow +\infty} |df_{k,\epsilon_k}|(0) = 1 .$$

Dans ce cas, l'estimation locale a priori (4.4) conduit à une borne uniforme sur les hessiens des fonctions $(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))$: en passant à la limite en k sur cette suite de fonctions nous obtenons une métrique $C^{1,1}$, $\check{g} = e^{2\check{f}} (\sum_{i=1}^4 dx_i^2)$ et, pour tout β , $\beta < 1$, une sous-suite des fonctions $(f_{k,\epsilon_k} - f_{k,\epsilon_k}(0))$ convergeant uniformément sur tout compact vers \check{f} dans la topologie $C^{1,\beta}$. De plus la métrique \check{g} vérifie les équations limites (cf. 4.8)

$$(4.12) \quad \sigma_2(A(\check{g})) = 0, \quad \text{scal}(\check{g}) \geq 0, \quad \text{et } |d\check{f}|(0) = \lim_{k \rightarrow +\infty} |df_{k,\epsilon_k}|(0) = 1 .$$

Sinon $\limsup_k e^{f_{k,\epsilon_k}(0)} = \ell$, $0 < \ell \leq 1$, et une sous-suite, que nous noterons encore – abusivement – f_{k,ϵ_k} , vérifie alors que $f_{k,\epsilon_k}(0)$ est minoré uniformément en k : $-c \leq f_{k,\epsilon_k}(0) \leq 0$, $c \in \mathbb{R}_+$. Puisque par construction $|df_{k,\epsilon_k}| \leq 1$, la suite f_{k,ϵ_k} est uniformément bornée en norme C^1 sur tout compact de \mathbb{R}^4 (nous ne considérons évidemment que les entiers $k \geq k_0$, où k_0 est suffisamment grand pour que $\epsilon_{k_0} < r_0/R$, R étant lui-même choisi tel que la boule $B(R)$ contient le compact considéré). Les fonctions φ_k étant uniformément minorées par hypothèse et α étant positif ou nul, les équations (4.7) satisfaites par les fonctions f_{k,ϵ_k} sont uniformément elliptiques : nous avons déjà discuté à plusieurs reprises la façon dont la concavité de l'opérateur $\sqrt{\sigma_2(A)}$ permet de passer des bornes C^2 provenant de ce qui précède et de l'estimation locale (4.4) à une borne en norme $C^{2,\beta}$, $\beta > 0$; nous savons aussi comment utiliser la

théorie elliptique ordinaire pour passer de la borne $C^{2,\beta}$ à une borne uniforme dans toutes les normes C^p , $p \in \mathbb{N}$.

Dans le cas où $\limsup e^{f_{k,\epsilon_k}(0)} = \ell > 0$, une sous-suite des fonctions f_{k,ϵ_k} converge donc, dans les normes C^m , $m \in \mathbb{N}$, et de façon uniforme sur tout compact, vers une fonction infiniment différentiable f satisfaisant de plus l'équation limite

$$(4.13) \quad \sigma_2(A(e^{2f} \sum_{i=1}^4 dx_i^2)) = b > 0, \quad (\text{et donc, } \text{scal}(e^{2f} \sum_{i=1}^4 dx_i^2) > 0);$$

on notera en effet que dans l'identité (4.7) les coefficients α_k sont uniformément majorés par c par hypothèse et que le terme $|W(\hat{g}_k)|^2$ tend vers 0 lorsque k tend vers l'infini, puisque la covariance conforme de la courbure de Weyl et la relation $\hat{g}_k = e^{2f_{k,\epsilon_k}} \hat{g}_k$ entraînent l'identité $|W(\hat{g}_k)|^2 = e^{-4f_{k,\epsilon_k}} |W(\hat{g}_k)|^2$, où la suite f_{k,ϵ_k} est uniformément bornée sur tout compact de \mathbb{R}^4 , et où les métriques (\hat{g}_k) convergent par construction vers la métrique euclidienne, plate, et donc conformément plate.

4.1.1.2. *Classification des éclatements.* — Commençons par démontrer que le système (4.12) satisfait par la métrique \check{g} de \mathbb{R}^4 obtenue par l'éclatement précédent dans le cas où $\lim_k e^{f_{k,\epsilon_k}(0)} = 0$ n'admet aucune solution : pour toute métrique conforme à la métrique euclidienne, $g = e^{2f} \sum_{i=1}^4 dx_i^2$, pour toute fonction de troncature $\eta := \eta_R$, de support la boule $B(2R)$, $R > 0$, égale à 1 sur $B(R)$, de gradient et de hessien majorés par C/R et C/R^2 respectivement, où C , $C \in \mathbb{R}_+^*$, est une constante universelle, on vérifie la majoration générale (et élémentaire) suivante

$$\begin{aligned} & \int_{B(2R)} \left(\frac{\text{scal}}{2} e^{2f} |df|^2 + |df|^4 \right) \eta^4 \, d\text{vol} \leq \int_{B(2R)} (f - \bar{f}^R) \sigma_2(A) e^{4f} \eta^4 \, d\text{vol} \\ & + C R^{-2} \int_{B(2R) \setminus B(R)} (f - \bar{f}^R) |df|^2 \eta^2 \, d\text{vol} + C R \int_{B(2R) \setminus B(R)} (f - \bar{f}^R) |df|^3 \eta^3 \, d\text{vol} \\ & \leq \int_{B(2R)} (f - \bar{f}^R) \sigma_2(A) e^{4f} \eta^4 \, d\text{vol} + C \left(\int_{B(2R) \setminus B(R)} \eta^4 |df|^4 \, d\text{vol} \right)^{1/2}, \end{aligned}$$

où par \bar{f}^R on représente la moyenne de f sur la boule $B(2R)$. La métrique \check{g} , satisfaisant les équations $\sigma_2(A(\check{g})) = 0$ et $\text{scal}(\check{g}) \geq 0$ (cf. (4.12)), la majoration précédente s'écrit dans ce cas particulier $\int_{B(2R)} \eta^4 |df|^4 \, d\text{vol} \leq C \sqrt{\int_{B(2R) \setminus B(R)} \eta^4 |df|^4 \, d\text{vol}}$, et implique clairement que la fonction \check{f} est nécessairement constante, ce qui contredit l'identité $|d\check{f}(0)| = 1$ (cf. 4.12)). De ceci nous déduisons que $\limsup e^{f_{k,\epsilon_k}(0)}$ est nécessairement strictement positive et que la métrique obtenue sur \mathbb{R}^4 par éclatement satisfait donc toujours l'identité de courbure (4.13).

Nous démontrons maintenant qu'une telle métrique est nécessairement de courbure sectionnelle constante, et donc l'image de la métrique d'une sphère ronde – ou d'une de ses images par le groupe conforme – par la projection stéréographique. Cette preuve peut être considérée comme une généralisation (pas tout à fait immédiate) de l'énoncé (facile, lui) de M. Obata d'après lequel toute métrique de courbure scalaire constante dans la classe conforme de la sphère ronde est en fait de courbure sectionnelle constante, le tenseur

$$L := \frac{1}{4} |\text{ric}_0|^2 g + \frac{1}{6} \text{scal} \, \text{ric}_0 - \text{ric}_0^2$$

jouant ici le rôle tenu par ric_0 dans l'argument de M. Obata. Notons pour commencer que le tenseur L est non seulement clairement de *trace nulle*, mais aussi de *divergence nulle*, d'après la seconde identité de Bianchi : ce sont là les deux propriétés clés du tenseur ric_0 dans l'argument de M. Obata. Une petite discussion algébrique démontre ensuite, sous les hypothèses $\sigma_2(A) > 0$ et $\text{scal} > 0$, que la trace (L, ric_0) est partout positive ou nulle – le tenseur ric_0 s'annulant de plus identiquement là où $(L, \text{ric}_0) = 0$ –, ainsi que la majoration $|L|^2 \leq 1/3 \text{scal} (L, \text{ric}_0)$.

En utilisant la fonction de troncature η_R introduite précédemment, ces propriétés conduisent facilement à la majoration

$$\int_{B(2R)} (L, \text{ric}_0) e^f \eta_R^2 d\text{vol} \leq C \sqrt{\int_{B(2R) \setminus B(R)} (L, \text{ric}_0) e^f \eta_R^2 d\text{vol}} \sqrt{\int_{B(2R)} \text{scal} |df|^2 e^f |d\eta_R|^2 d\text{vol}} ;$$

pour conclure à l'annulation de ric_0 – et donc à la platitude projective de $e^{2f} \sum_{i=1}^4 dx_i^2$ – il reste à majorer $\int_{B(2R)} \text{scal} |df|^2 e^f |d\eta_R|^2 d\text{vol}$ indépendamment de R , c'est-à-dire à majorer l'expression $R^{-2} \int_{B(2R) \setminus B(R)} \text{scal} e^f |df|_0^2 d\text{vol}_0$ uniformément en R (où l'indice 0 renvoie à la métrique euclidienne). Ceci s'avère moins banal qu'attendu, la clé en étant une minoration (asymptotique) du facteur conforme e^{2f} de toute métrique conforme à la métrique euclidienne et de courbure scalaire positivement minorée, qui découle de [KMPS] : dès que $\inf_M \text{scal} > 0$, et pour $|x|$ assez grand, $e^{2f(x)} \geq C |x|^{-4}$, où la constante $C = C(\inf_M \text{scal})$ tend vers zero avec le minimum de la courbure scalaire, $\inf_M \text{scal}$.

4.1.1.3. Conclusion. — Si la majoration uniforme de $|df|^2 + e^f$ n'a pas lieu, nous disposons d'après la discussion précédente d'une suite de fonctions f_{k, ϵ_k} convergeant vers une fonction infiniment différentiable f telle que $e^{2f} \sum_{i=1}^4 dx_i^2$ est une métrique de courbure constante sur \mathbb{R}^4 . Cette métrique $e^{2f} \sum_{i=1}^4 dx_i^2$ est la limite $C^k, k \geq 2$, uniforme sur tout compact, des métriques $e^{2f_k, \epsilon_k} \hat{g}_k = \hat{g}_k = h_{\epsilon_k}^*(e^{2f_k} g)$; pour tout $R > 0$ nous avons donc

$$\begin{aligned} \int_{(B(R), e^{2f} \sum_{i=1}^4 dx_i^2)} \sigma_2(A) d\text{vol} &= \lim_{k \rightarrow +\infty} \int_{B(R)} \sigma_2(A(e^{2f_k} g)) \circ h_{\epsilon_k} h_{\epsilon_k}^*(d\text{vol}(e^{2f_k} g)) \\ &= \lim_{k \rightarrow +\infty} \int_{(B(P_k, \epsilon_k R), e^{2f_k} g)} \sigma_2(A) d\text{vol} \leq \lim_{k \rightarrow +\infty} \int_{(M, e^{2f_k} g)} \sigma_2(A) d\text{vol} , \end{aligned}$$

où nous avons utilisé la « naturalité » de la courbure – i.e. l'identité tautologique $R(F^*g) = F^*R(g)$, valable pour tout difféomorphisme F – (pour la première identité), et l'hypothèse $\sigma_2(A(e^{2f_k} g)) = \varphi_k + \frac{\alpha}{4} |W(e^{2f_k} g)|^2 > 0$ (pour la dernière majoration). En passant alors à la limite sur $R, R \rightarrow +\infty$, et en rappelant l'invariance conforme de l'intégrale $\int_M \sigma_2(A) d\text{vol}$, nous en déduisons la minoration

$$16\pi^2 = \int_{(S^4, \text{can})} \sigma_2(A) d\text{vol} = \int_{(\mathbb{R}^4, e^{2f} \sum_{i=1}^4 dx_i^2)} \sigma_2(A) d\text{vol} \leq \int_{(M, g)} \sigma_2(A) d\text{vol} .$$

Mais on sait que l'inégalité opposée est vraie pour toute variété compacte sans bord de dimension 4 et de courbure scalaire strictement positive (il suffit en fait que l'invariant de Yamabe soit positif ou nul), et que le cas d'égalité caractérise la classe conforme de la

sphère standard (cf. [G2], Theorem B), ce qui conclut la preuve de la majoration uniforme de $|df|^2 + e^{2f}$ annoncée :

$$(4.14) \quad \sup_M (|df|^2 + e^{2f}) \leq C(|\varphi|_{C^2}, \inf_M \varphi, g).$$

En intégrant alors l'équation $\sigma_2(A) - \frac{\alpha}{4}|W|^2 = \varphi$ en une solution f , et en rappelant que par hypothèse α est positif ou nul, nous obtenons la minoration

$$\begin{aligned} 0 < \int_{(M,g)} (\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} &= \int_{(M, e^{2f}g)} (\sigma_2(A) - \frac{\alpha}{4}|W|^2) d\text{vol} = \int_{(M, e^{2f}g)} \varphi d\text{vol} \\ &= \int_{(M,g)} \varphi e^{4f} d\text{vol} \leq \sup_M \varphi \text{vol}(g) e^{4\sup_M f}. \end{aligned}$$

Ceci garantit l'existence d'un minorant du maximum, $\sup_M f$, ne dépendant que de g , d'un minorant strictement positif de $\int_M \sigma_2(A) - \frac{\alpha}{4}|W|^2 d\text{vol}$ et de $\sup_M |\varphi|$ (notons que $\int_{(M,g)} \varphi d\text{vol}$ suffirait par l'argument précédent, puisque par hypothèse $\varphi \geq 0$); d'après la majoration uniforme (4.14) de f et de $|df|$, $|f|$ – et donc $|\text{Hess} f|$ par l'estimation a priori locale (4.4) – sont alors uniformément majorés par une constante ne dépendant que de g , de α , de la norme C^2 de φ , et d'un minorant de φ , ce qui achève la démonstration de la proposition 4.2.

4.1.2. Invariance homotopique du degré de Leray-Schauder-Li. — Nous supposons dans ce paragraphe que (M^4, g) n'est pas conformétement équivalente à la sphère standard. Pour résoudre l'équation $4\sigma_2(A) - \alpha|W|^2 = \varphi$ nous invoquons la théorie du degré de Leray-Schauder : Li Y.Y. introduit dans [L] un tel degré pour tout opérateur elliptique du second ordre, F , sur une variété riemannienne compacte. La stratégie est classique (cf. [FP]) : on remarque, l'opérateur $(\Delta + 1d)$ étant inversible, que F , et sa composition avec $(\Delta + 1d)$, $(\Delta + 1d) \circ F$, ont les mêmes solutions, l'opérateur du quatrième degré $(\Delta + 1d) \circ F$ présentant l'avantage de pouvoir s'écrire $(a(x, f, df, D^2f) \otimes g, D^4f) + C(x, f, \dots, D^3f)$, où a représente le symbole principal de l'opérateur elliptique F , un 2-tenseur symétrique défini positif. En introduisant l'opérateur linéaire $L_f\varphi = (a_f \otimes g, D^4\varphi) + C_f$, nous pouvons encore écrire $(\Delta + 1d) \circ F(f)$ sous la forme $L_f(f)$. Il reste à vérifier que, pour tout entier k assez grand, l'opérateur $M_{f,k} := L_f - C_f + k(\Delta + 1d)$ est un isomorphisme et à observer que l'opérateur $C_f - k(\Delta + 1d)$ est compact (de $C^{4,\alpha}$ vers C^α), ce qui permet alors de considérer le degré – au sens de Leray et Schauder – de la perturbation compacte de l'identité $1d + M_{f,k}^{-1}(C_f - k(\Delta + 1d))$. En invoquant l'invariance homotopique du degré de Leray-Schauder on démontre que ce degré ne dépend pas de k ; par ailleurs, il « compte » bien les solutions de notre problème puisque, $(\Delta + 1d)$ et $M_{f,k}$ étant des isomorphismes, $F(f) = 0$, $(\Delta + 1d)F(f) = 0$, $L_f(f) = (M_{f,k} + C_f - k(\Delta + 1d))f = 0$, et $(1d + M_{f,k}^{-1}(C_f - k(\Delta + 1d)))f = 0$ sont des énoncés équivalents. Comme on l'aura deviné, Li Y.Y. établit l'*invariance homotopique* de ce degré – au moins pour une déformation *uniformément elliptique* et sous l'existence d'une borne a priori $C^{4,\beta}$ uniforme sur les solutions. Il vérifie de plus que le degré de l'opérateur elliptique du second ordre F coïncide, en une solution f où la linéarisation de F

est inversible, avec le degré en 0 du linéarisé, et que ce dernier est donné par la formule

$$(4.15) \quad \sum_{\lambda_i < 0} (-1)^{\beta_i} ,$$

où par $(\lambda_i)_{i \in \mathbb{N}}$ on représente le spectre du linéarisé en f et par β_i la multiplicité de la i -ième valeur propre λ_i . Bien qu'élémentaire et aujourd'hui classique, cette théorie n'en reste pas moins élégante et bien utile.

Sous l'hypothèse de positivité du théorème 4.1 nous avons établi (et de deux façons très différentes) l'existence d'une solution f_α^\diamond , $\alpha \in \mathbb{R}_+$, à l'équation $\sigma_2(A) - \frac{\alpha}{4} |W|^2 > 0$. Posons $\varphi_\alpha = (\sigma_2(A) - \frac{\alpha}{4} |W|^2) (e^{2f_\alpha^\diamond} g)$ et introduisons la déformation

$$(4.16) \quad \sigma_2 - \frac{\alpha}{4} |W|^2 = (1 - t) \varphi + t \varphi_\alpha .$$

La minoration uniforme

$$(4.17) \quad \inf_M ((1 - t) \varphi + t \varphi_\alpha) \geq \min(\inf \varphi, \inf \varphi_\alpha) > 0$$

permet d'invoquer la proposition 4.2 pour majorer uniformément les solutions de (4.16) en termes de g , des normes C^2 de φ et φ_α , et de $\min(\inf \varphi, \inf \varphi_\alpha)$ uniquement. La minoration (4.17) garantit par ailleurs que les symboles principaux sont uniformément minorés, puisqu'égaux au tenseur gravitationnel $G = -\text{ric} + 1/2 \text{ scal } g$, minoré par $3 \frac{\sigma_2(A)}{\text{scal}} g$, la borne C^2 donnée par la proposition 4.2 permettant par ailleurs de majorer uniformément la courbure scalaire.

Pour établir l'existence d'une solution de l'équation (4.16) _{$t=0$} , il suffit, d'après la théorie du degré de Li rappelée ci-dessus, de démontrer que le degré de l'équation (4.16) _{$t=1$} en sa solution f_α n'est pas nul. Pour cela il peut être utile d'écrire l'équation (4.16) _{$t=1$} sous la forme d'une *divergence*

$$(4.18) \quad d^*(M(f) df) - \varphi_\alpha (e^{4f} - 1) = 0 ,$$

où le 2-tenseur symétrique $M(f) := G(e^{2f} g) + G(g) + |df|_g^2 g$ est défini positif dès que $G(e^{2f} g)$ et $G(g)$ le sont, en particulier lorsque $\sigma_2(A(e^{2f} g))$, $\sigma_2(A(g))$, $\text{scal}(e^{2f} g)$ et $\text{scal}(g)$ sont strictement positifs.

Pour calculer le degré de (4.18) en sa solution $f = 0$, nous invoquons à nouveau l'invariance homotopique du degré que nous appliquons à la déformation suivante de l'équation (4.18)

$$(4.19) \quad d^*(M(f) df) + \varphi_\alpha = (1 - t) \varphi_\alpha e^{4f} + t \varphi_\alpha \int_{(M, e^{2f} g)} e^{4f} d\text{vol} ;$$

nous nous ramenons ainsi à calculer le degré de l'équation (4.19) _{$t=1$} en 0 (remarquez, en supposant sans restriction que $\text{vol}(g) = 1$, que $f = 0$ est solution de (4.19) pour tout t).

Les estimations a priori $C^{4,\alpha}$ requises par la théorie de Li Y.Y. sont établies, pour t assez loin de 1, en raisonnant comme pour la démonstration de la proposition 4.2 ; pour t proche de 1 (et loin de zéro, donc) les auteurs s'appuient directement sur une inégalité « classique »

de J. Moser et N. Trudinger d’après laquelle, pour une constante universelle $c_1, c_1 > 0$, l’intégrale

$$\int_M \exp\left(\frac{c_1 |\varphi - \bar{\varphi}|}{\left(\int_M |d\varphi|^4 d\text{vol}\right)^{1/4}}\right)^{4/3} d\text{vol}$$

est uniformément majorée sur $L^{1,4}((M^4, g))$.

La minoration

$$\begin{aligned} (\sigma_2(A) - \frac{\alpha}{4} |W|^2) (e^{2f} g) &= e^{-4f} d^*(M(f) df) = e^{-4f} t \varphi_\alpha \int_{(M, e^{2f} g)} e^{4f} d\text{vol} + (1 - t) \varphi_\alpha \geq \\ &\min(e^{-12 \sup_M |f|}, 1) \inf \varphi_\alpha \end{aligned}$$

permet alors de minorer uniformément le symbole principal, G , comme précédemment.

Il reste à évaluer le degré en $t = 1$. En intégrant sur M la spécialisation en $t = 1$ de l’équation (4.19) nous dérivons la condition nécessaire

$$(4.20) \quad \int_{(M, e^{2f} g)} e^{4f} d\text{vol} = 1,$$

en vertu de laquelle l’équation $(4.19)_{t=1}$ se simplifie pour donner l’identité $d^*(M(f) df) = 0$.

En intégrant cette dernière contre df , on établit – la forme bilinéaire symétrique $M(f)$ étant définie positive – que f est nécessairement constante, et donc nulle d’après l’identité (4.20) sous la normalisation, toujours possible : $\text{vol}(g) = 1$. Considérons alors la linéarisation de l’équation $(4.19)_{t=1}$ en la fonction identiquement nulle, son unique solution :

$$L_0(u) = -2(G(g), \text{Hess } u) - 4 \varphi_\alpha \int_{(M, g)} u d\text{vol}.$$

Puisque la fonction φ_α est strictement positive, en intégrant sur M cette équation on démontre que tout élément du noyau de L_0 est d’intégrale nulle. Le tenseur gravitationnel G étant défini positif, le principe du maximum (ou, de façon équivalente, le tenseur gravitationnel étant de divergence nulle, l’identité $-\int_M (G, \text{Hess } u) d\text{vol} = \int_M G(du^\sharp, du^\sharp) d\text{vol}$ entraîne alors que la fonction u est identiquement nulle. Puisque le noyau de L_0 est réduit à $\{0\}$, nous pouvons invoquer la formule (4.15) pour calculer le degré de L_0 : si λ est valeur propre de L_0 , l’équation $L_0(u_\lambda) = \lambda u_\lambda$ s’intègre en l’identité

$$(4.21) \quad (\lambda + 4 \int_M \varphi_\alpha d\text{vol}) \int_M u_\lambda d\text{vol} = 0;$$

$-4 \int_M \varphi_\alpha d\text{vol}$ est effectivement valeur propre, les constantes appartenant à l’espace propre associé. Pour les autres valeurs propres, nous déduisons de l’identité (4.21) la condition nécessaire $\int_M u_\lambda d\text{vol} = 0$, et donc, en rappelant que le tenseur gravitationnel est de divergence nulle, l’identité $\lambda u_\lambda^2 = (L_0 u_\lambda, u_\lambda) = 2 u_\lambda d^*(G(du_\lambda^\sharp))$: ces valeurs propres, qui satisfont alors l’identité $\lambda = \int_M G(du_\lambda^\sharp, du_\lambda^\sharp) d\text{vol} / \int_M |du_\lambda|^2 d\text{vol}$, sont toutes (strictement) positives. L’opérateur L_0 admet ainsi une unique valeur propre négative, de multiplicité nécessairement égale à un en tant que « plus petite valeur propre » de L_0 – un corollaire classique du principe du maximum. Le degré de L_0 à l’origine, calculé à l’aide de la formule (4.15), vaut donc -1 . Par les arguments qui précèdent, c’est encore le degré de l’équation $\sigma_2 - \alpha/4 |W|^2 = \varphi$,

dont nous démontrons ainsi qu'elle admet (au moins) une solution. Ceci conclut la preuve du théorème 4.1.

4.2. Démonstration du théorème 2

4.2.1. *Construction d'une solution $C^{1,1}$.* — Dans les premières lignes de ce dernier chapitre nous avons expliqué comment ramener le problème au cas où $\int_M |W|^2 d\text{vol}$ est strictement positif : nous supposons donc désormais que la variété (M, g) n'appartient pas à la classe conforme de la sphère standard. Pour tout α , $\alpha \in [0, 1)$, $\int_M (\sigma_2(A) - \alpha/4 |W|^2) d\text{vol} = (1 - \alpha)/4 \int_M |W|^2 d\text{vol}$ étant strictement positif, le théorème 4.1 appliqué à la fonction $\varphi \equiv 1$ assure l'existence d'une solution f_α de l'équation

$$(4.22) \quad (\sigma_2(A) - \frac{\alpha}{4} |W|^2)(e^{2f_\alpha} g) = 1 ,$$

ainsi qu'une majoration du $\sup_M (e^{f_\alpha} + |df_\alpha|)$ uniforme en α , $\alpha \in [0, 1)$. En intégrant l'équation (4.22) nous dérivons cependant la majoration banale suivante du minimum de f_α :

$$(4.23) \quad \begin{aligned} \frac{1 - \alpha}{4} \int_M |W|^2 d\text{vol} &= \int_M (\sigma_2(A) - \frac{\alpha}{4} |W|^2) d\text{vol} = \int_{(M, e^{2f_\alpha} g)} (\sigma_2(A) - \frac{\alpha}{4} |W|^2) d\text{vol} \\ &= \int_{(M, e^{2f_\alpha} g)} d\text{vol} = \int_{(M, g)} e^{4f_\alpha} d\text{vol} \geq e^{4 \inf_M f_\alpha} \int_{(M, g)} d\text{vol} , \end{aligned}$$

et concluons que nécessairement $\lim_{\alpha \rightarrow 1} \inf_M f_\alpha = -\infty$: les solutions f_α ne sauraient satisfaire les conclusions de la proposition 4.2 (cf. les commentaires faisant suite à l'énoncé du Théorème 4.1). Par contre, en définissant pour tout α , $\alpha \in [0, 1)$, a_α et \tilde{f}_α par les formules

$$(4.24) \quad e^{4a_\alpha} := \frac{1 - \alpha}{4} \int_M |W|^2 d\text{vol} , \quad \text{et} \quad \tilde{f}_\alpha := f_\alpha - a_\alpha ,$$

les identités précédentes démontrent que $\int_{(M, g)} e^{4\tilde{f}_\alpha} d\text{vol} = 1$; nous en déduisons une minoration de $\sup_M \tilde{f}_\alpha$, et une majoration de $\inf_M \tilde{f}_\alpha$, cette fois toutes les deux uniformes en α . Puisque les normes des différentielles $|d\tilde{f}_\alpha| = |df_\alpha|$ sont aussi uniformément majorées, nous concluons que les fonctions \tilde{f}_α sont uniformément bornées : il existe une constante C , indépendante de α , $\alpha \in [0, 1)$, telle que

$$(4.25) \quad \sup_M |\tilde{f}_\alpha| < C .$$

Les fonctions \tilde{f}_α satisfaisant les équations

$$(4.26) \quad (\sigma_2(A) - \frac{\alpha}{4} |W|^2)(e^{2\tilde{f}_\alpha} g) = e^{4a_\alpha} (\sigma_2(A) - \frac{\alpha}{4} |W|^2)(e^{2f} g) = e^{4a_\alpha} ,$$

et les normes C^2 des fonctions (constantes) e^{4a_α} , $\alpha \in [0, 1)$, étant uniformément bornées d'après l'identité 4.24, la proposition 4.2 (en particulier l'estimation a priori (4.4)), garantit alors que les fonctions $(\tilde{f}_\alpha)_{0 \leq \alpha < 1}$ sont uniformément bornées en norme C^2 . Il existe donc une métrique $C^{1,1}$, $e^{2\tilde{f}} g$, telle que, pour tout $\beta \in [0, 1)$, il existe une suite α_k , $\alpha_k \rightarrow 1$, pour laquelle les métriques $e^{2\tilde{f}_{\alpha_k}} g$ convergent dans la topologie $C^{1,\beta}$ vers $e^{2\tilde{f}} g$. De l'équation 4.26 nous déduisons de plus que la limite $e^{2\tilde{f}} g$ vérifie, au sens faible, l'équation $4\sigma_2(A) - |W|^2 = 0$. On ne peut cependant pas espérer utiliser cette équation pour améliorer les estimations C^2

précédentes, puisqu'elle cesse d'être elliptique là où $\sigma_2(A)$ s'annule – c'est-à-dire là où la métrique initiale g est conformément plate.

Par contre, au voisinage d'un point, P , où la courbure de Weyl de la métrique g ne s'annule pas, nous disposons de la minoration

$$\sigma_2(A(e^{2\tilde{f}_\alpha} g)) = \frac{\alpha}{4} |W(e^{2\tilde{f}_\alpha} g)|^2 + e^{4a_\alpha} \geq \frac{\alpha}{4} e^{-4\tilde{f}_\alpha} |W(g)|^2,$$

la borne uniforme (4.25) entraînant alors une minoration uniforme de $\sigma_2(A(e^{2\tilde{f}_\alpha} g))$ sur un voisinage de P . En écrivant l'équation sous la forme $\sqrt{\sigma_2(A)} - \sqrt{\frac{\alpha}{4} |W|^2 + e^{4a_\alpha}}$, nous pouvons invoquer une dernière fois la théorie d'Evans et Krylov pour les équations concaves strictement elliptiques qui établit alors l'existence de bornes $C^{2,\beta}$ uniformes, $\beta > 0$, sur un voisinage un peu plus petit ; pour tout entier naturel l , $l \in \mathbb{N}$, la théorie elliptique standard garantit ensuite la convergence dans la norme C^l d'une sous-suite $e^{2\tilde{f}_{\alpha_k}} g$ vers la métrique $e^{2\tilde{f}} g$, démontrant ainsi que la métrique $e^{2\tilde{f}} g$ est infiniment différentiable en dehors des zéros de la courbure conforme de g . Pour conclure la preuve du théorème 2, il me reste à vous convaincre que cet ensemble $|W|^{-1}(0)$, le lieu des zéros de la courbure de Weyl, est nécessairement *vide* dans les cas pertinents.

4.2.2. Rigidité des métriques plates au sens de Bach. — S'il existe sur M une métrique telle que $\int_M (\sigma_2(A) - |W|^2/4) d\text{vol} > 0$, c'est-à-dire telle que $\int_M |W|^2 d\text{vol} < 16\pi^2 \chi(M)$, le théorème 1 affirme que M est diffeomorphe à S^4 ou à $\mathbb{P}_{\mathbb{R}}^4$: pour établir le théorème 2 nous pouvons donc supposer sans restriction que la métrique pour laquelle $\int_M (\sigma_2(A) - |W|^2/4) d\text{vol} = 0$ minimise la fonctionnelle $\|W\|_{L^2}$ et qu'elle satisfait à ce titre l'équation d'Euler associée, à savoir l'annulation du tenseur de Bach, $B := -\text{tr}_{13} \text{tr}_{25} D^2W - 1/2 W(\text{ric})$, introduit précédemment pour sa propriété de covariance conforme (cf. (2.27)). Une première propriété intéressante des métriques plates au sens de Bach est l'identité intégrale suivante, inspirée par une formule équivalente de A. Derdzinski pour les métriques de courbure de Weyl harmonique.

LEMME 4.3. — *Toute métrique plate au sens de Bach sur une variété compacte sans bord de dimension 4 satisfait l'identité intégrale suivante :*

$$(4.27) \quad \int_M |DW|^2 d\text{vol} = \int_M (72 \det W^+ + 72 \det W^- - \frac{1}{2} \text{scal} |W|^2 + 2 W(\text{ric}_0, \text{ric}_0)) d\text{vol} .$$

Pour la première fois dans ce texte intervient la décomposition de la courbure en composantes auto-duale, $*W^+ = W^+$, et anti-autoduale, $*W^- = -W^-$, sous l'action de l'opérateur $*$ de Hodge. Si la preuve du lemme 4.3 n'est pas immédiate, elle n'en reste pas moins élémentaire, reposant essentiellement sur les deux identités de Bianchi.

Une autre propriété utile découle de l'identité (4.3), valable pour toute métrique en dimension 4, et qui, spécialisée à une métrique plate au sens de Bach, s'intègre en l'identité intégrale suivante

$$(4.28) \quad \int_M (3 |D\text{ric}_0|^2 - 1/4 |d\text{scal}|^2 + 6 \text{tr} \text{ric}_0^3 + \text{scal} |\text{ric}_0|^2 - 6 W(\text{ric}_0, \text{ric}_0)) d\text{vol} = 0 .$$

Pour les métriques $e^{2\tilde{f}_\alpha}g$ construites au paragraphe précédent, $\sigma_2(A) - \alpha/4|W|^2 = e^{4a_\alpha} \geq 0$; nous en déduisons facilement la minoration $3\alpha/2 |DW|^2 + 3|D\text{ric}_0|^2 - 1/4|d\text{scal}|^2 \geq 0$, qui donne après intégration, et en invoquant la combinaison linéaire adéquate des identités (4.27) et (4.28), la majoration fondamentale suivante, valable pour tout α , $\alpha \in [0, 1)$, et toute métrique $e^{2\tilde{f}_\alpha}g$ solution de l'équation $\sigma_2(A) - \alpha/4|W|^2 = e^{4a_\alpha}$, où le nombre réel a_α est défini par l'identité (4.24) :

$$(4.29) \quad \int_M (6 \text{tr ric}_0^3 + \text{scal} |\text{ric}_0|^2 - 3(\alpha + 2) W(\text{ric}_0, \text{ric}_0) - 108 \alpha (\det W^+ + \det W^-) + 3 \alpha/4 \text{scal} |W|^2) d\text{vol} \leq 0.$$

L'intégrand se réduit à $6 \text{tr ric}_0^3 + \text{scal} |\text{ric}_0|^2$ sur le lieu des zéros de $|W|$, $|W|^{-1}(0)$. Ce polynôme en la courbure est positif ou nul dès que $\sigma_2(A)$ l'est ; les métriques $e^{2\tilde{f}_\alpha}g$ satisfaisant l'équation $\sigma_2(A) - \frac{\alpha}{4}|W|^2 = e^{4a_\alpha} > 0$ (cf. (4.24)), la majoration (4.29) reste donc valable si l'on restreint l'intégrale à $M^4 \setminus |W|^{-1}(0)$.

La convergence C^l , $l \geq 2$, des métriques $e^{2\tilde{f}_\alpha}g$ sur $M \setminus |W|^{-1}(0)$ discutée au paragraphe précédent permet de passer à la limite $\alpha \rightarrow 1$ dans l'intégrale précédente sur $M \setminus |W|^{-1}(0)$ pour établir que la limite $e^{2\tilde{f}}g$ satisfait la majoration suivante

$$(4.30) \quad \int_{M \setminus |W|^{-1}(0)} (6 \text{tr ric}_0^3 + \text{scal} |\text{ric}_0|^2 - 9 W(\text{ric}_0, \text{ric}_0) - 108 (\det W^+ + \det W^-) + 3/4 \text{scal} |W|^2) d\text{vol} \leq 0.$$

Nous sommes ici au cœur de la preuve du théorème de rigidité : une discussion algébrique fine, s'appuyant sur des arguments proches de ceux développés dans [M], établit que l'intégrand qui figure dans (4.30) est un polynôme en la courbure *universellement positif ou nul* en dimension 4 et que son annulation en un point correspond à l'annulation, en ce même point, soit de la partie sans trace de la courbure de Ricci, ric_0 , soit de la courbure conforme, W .

Puisque nous avons supposé $\|W\|_{L^2} > 0$, l'image inverse de 0 par $|W|$ est un fermé strict de M . Considérons une composante connexe (non vide) de son complémentaire, un ouvert O de M , sur lequel la partie sans trace de la courbure de Ricci, $\text{ric}_0(e^{2\tilde{f}}g)$ est, d'après ce qui précède, identiquement nulle. La restriction de la métrique $e^{2\tilde{f}}g$ à l'ouvert O est donc, par définition, d'Einstein et la seconde identité de Bianchi démontre que sa courbure scalaire est en particulier constante. De l'équation $4\sigma_2(A) - |W|^2 = 0$ satisfaite par $e^{2\tilde{f}}g$ nous déduisons alors que la norme de la courbure conforme est, elle aussi, constante sur O : $|W(e^{2\tilde{f}}g)|^2|_O \equiv C$. M étant connexe par hypothèse, si O était un ouvert strict de M il existerait une suite de points P_i dans O convergeant vers un point P de $|W|^{-1}(0)$. Pour cette suite, nous aurions par construction :

$$0 < C = |W(e^{2\tilde{f}}g)|^2(P_i) = e^{-4\tilde{f}(P_i)}|W(g)|^2(P_i).$$

La métrique g étant infiniment différentiable, sa courbure conforme l'est aussi, qui satisferait en particulier la propriété de continuité : $\lim_{i \rightarrow +\infty} |W(g)|_g^2(P_i) = |W(g)|_g^2(P) = 0$; mais ceci contredirait la continuité de la limite \tilde{f} , $\tilde{f} \in C^{1,1}$, construite au paragraphe précédent.

Donc l'ouvert O coïncide avec la variété M , la courbure conforme W ne s'annule nulle part, et la métrique infiniment différentiable $e^{2\tilde{f}}g$ est une solution de l'équation $4\sigma_2(A) - |W|^2 \equiv 0$; l'argument précédent démontre de plus que sa courbure scalaire est (constante et) partout strictement positive. On déduit alors le théorème 2 (d'un cas particulier facile) du résultat de rigidité ([M], Theorem 2) d'après lequel toute variété riemannienne de dimension 4 compacte, sans bord 1/6-faiblement pincée est, sinon difféomorphe à S^4 ou $\mathbb{P}_{\mathbb{R}}^4$, un plan projectif standard ou un quotient d'un produit $(S^3, \lambda \text{ can}) \times \mathbb{R}$, $\lambda \in \mathbb{R}_+^*$, par un sous-groupe cocompact d'isométries.

RÉFÉRENCES

- [A] D. ADAMS – *A sharp inequality of J. Moser for higher order derivatives*, Annals of Math. 128 (1988) 385-398.
- [CY] CHANG S.-Y. A. and YANG P. – *Extremal metrics of zeta functional determinants on 4-manifolds*, Annals of Math. 142 (1995) 171-212.
- [CGY0] CHANG S.-Y. A., M. GURSKY and YANG P. – *A conformally invariant sphere theorem in four dimensions*, Publ. Math. Inst. Hautes Études Sci. n° 98 (2003) 105-143.
- [CGY1] CHANG S.-Y. A., M. GURSKY and YANG P. – *An equation of Monge-Ampère type in conformal geometry and 4-manifolds of positive Ricci curvature*, Annals of Math. 155 (2002) 711-789.
- [CGY2] CHANG S.-Y. A., M. GURSKY and YANG P. – *An a priori estimate for a fully nonlinear equation on 4-manifolds*, J. d'Analyse Mathématique 87 (2002)
- [CGY3] CHANG S.-Y. A., M. GURSKY and YANG P. – *Regularity of a fourth order nonlinear PDE with critical exponent*, Amer. J. Math. 121 (1999) 215-257.
- [CQ] CHANG S.-Y. A. and QING J. – *The Zeta functional determinants on manifolds with boundary II – Extremum metrics and compactness of isospectral set*, J. Funct. Anal. 147 (1997) n° 2 363-399.
- [G1] M. GURSKY – *The Weyl functional, de Rham cohomology and Kähler-Einstein metrics*, Annals of Math. 148 (1998) 315-337
- [G2] M. GURSKY – *The principal eigenvalue of a conformally invariant differential operator, with application to semi-linear elliptic PDE*, Comm. Math. Phys. 207 (1999) 131-143.
- [GV] M. GURSKY and J. VIACLOVSKY – *A fully nonlinear equation on 4-manifolds with positive scalar curvature*, J. Diff. Geom. 63 (2003) 131-154.

- [GW] GUAN P. and WANG G. – *Local estimates for a class of fully nonlinear equations arising from conformal geometry*, Int. Math. Res. Not. 26 (2003) 1413-1432.
- [K] O. KOBAYASHI – *Scalar curvature of a metric with unit volume*, Math. Ann. 279 (1987) 253-265.
- [KMPS] N. KOREVAAR, R. MAZZEO, F. PACARD and R. SCHOEN – *Refined asymptotics for constant scalar curvature metrics with isolated singularities*, Invent. Math. 135 (1999) 233-272.
- [L] LI Y.Y. – *Degree theory for second order nonlinear elliptic operators and its applications*, Commun. in P.D.E. 14 (11) (1989) 1541-1578.
- [LL] LI A. and LI Y.Y. – *On some conformally invariant fully nonlinear equations*, Comm. Pure Appl. Math. 56 (2003) n° 10 1416-1464.
- [M] C. MARGERIN – *A sharp characterization of the smooth 4-sphere in curvature terms* Comm. Anal. Geom. 6 (1998) 21-65.
- [S] R. SCHOEN – *Analytic aspects of the harmonic map problem* in *Seminar on nonlinear P.D.E.*, MSRI Vol. 2 Springer Verlag (1984) 321-368.
- [SU] R. SCHOEN and K. UHLENBECK – *A regularity theory for harmonic maps*, J. Diff. Geom. 17 (1982) 307-335
- [SY] SHA J.-P. and YANG D.-G. – *Positive Ricci curvature on compact 1-connected 4-manifolds*, Proc. Sympos. Pure Math. vol. 54 (1993) 529-538.
- [Y] YANG D. – *L^p pinching and compactness theorems for compact Riemannian manifolds*, Forum Math. 4 (1992) 323-333.

Références bibliographiques complémentaires

- [ABKS] K. AKUTAGAWA, B. BOTVINNIK, O. KOBAYASHI and H. SESHADRI – *The Weyl functional near the Yamabe Invariant*, J. Geom. Anal. 13 n° 1 (2003) 1-20.
- [Be] W. BECKNER – *Sharp Sobolev inequalities on the sphere and the Moser-Trudinger inequality*, Annals of Math. 138 (1993) 213-242.
- [B-Ø] T. BRANSON and B. ØRSTED – *Explicit functional determinants in four dimensions*, Proc. AMS 113 (1991), 669-682.
- [CY2] CHANG S.-Y. A. and YANG P. – *Non-linear partial differential equations in conformal geometry*, Plenary talk ; Proceedings of the International Congress of Mathematicians Vol.1 Beijing Higher Ed. Press 189-207
- [FG1] C. FEFFERMAN and C.R. GRAHAM – *Conformal invariants* in « *Élie Cartan et les mathématiques d'aujourd'hui* », Astérisque (1985) p.95-116.
- [FG2] C. FEFFERMAN and C.R. GRAHAM – *Q-curvature and Poincaré metrics*, Math. Res Lett. 9, 2-3 (2002) 139-152.

- [FH] C. FEFFERMAN and K. HIRACHI – *Ambient metric construction of Q -curvature in conformal and CR geometries*, Math. Res. Lett. 10 n° 5-6 (2003) 819-831.
- [FP] P. FITZPATRICK and J. PEJSACHOWICZ – *An extension of the Leray-Schauder degree for fully nonlinear elliptic problems*, Proc. Sympos. Pure Math. 45-1 (1986) 425-438.
- [GJMS] C.R. GRAHAM, R. JENNE, L. MASON and G. SPARLING – *Conformally invariant powers of the Laplacian I : existence*, J. London Math Soc. 46 (1992) n° 2, 557-565.
- [GP] A.R. GOVER and L. PETERSON – *Conformally invariant powers of the Laplacian, Q -curvature and tractor calculus*, Comm. Math. Phys. 235 n° 2 (2003) 339-378.
- [GV2] M. GURSKY and J. VIACLOVSKY – *Fully nonlinear equations on Riemannian manifolds with negative curvature*, Indiana Univ. Math. J. 52 (2003) n° 399-420.
- [GZ] C.R. GRAHAM and M. ZWORSKY – *Scattering matrix in conformal geometry*, Invent. Math. 152 (2003) n° 1 89-118.
- [OPS1] B. OSGOOD, R. PHILLIPS, and P. SARNAK – *Extremals of determinants of Laplacians*, J. Func. Anal. 80 (1988) 148-211.
- [OPS2] B. OSGOOD, R. PHILLIPS, and P. SARNAK – *Compact isospectral sets of surfaces*, J. Func. Anal. 80 (1988) 212-234.

Christophe MARGERIN

École Polytechnique

UMR 7640 du CNRS

Centre de Mathématiques Laurent Schwartz

F-91128 PALAISEAU Cédex

E-mail : margerin@math.polytechnique.fr

**PROBLÈMES DE RECOUVREMENT ET POINTS EXCEPTIONNELS
POUR LA MARCHE ALÉATOIRE
ET LE MOUVEMENT BROWNIEN**
[d'après Dembo, Peres, Rosen et Zeitouni]

par **Zhan SHI**

INTRODUCTION

La marche aléatoire (ou marche au hasard) sur Z^d est un objet fondamental de la théorie des probabilités. Elle représente le mouvement aléatoire d'une particule (dont le point de départ est, disons, l'origine) qui, à chaque unité de temps, se déplace de façon équiprobable vers l'un de ses $2d$ plus proches voisins. On suppose de plus que tous les déplacements sont indépendants.

En 1960, Erdős et Taylor [17] ont posé le problème suivant pour la marche aléatoire sur Z^2 : quel est le nombre maximal de visites que la marche aléatoire peut effectuer en un site pendant les n premières étapes ? Plus précisément, introduisons $T_n(x)$, le nombre de visites de la marche aléatoire au point x lors des n premières étapes, et T_n^* , le maximum des variables $T_n(x)$, c'est-à-dire $T_n^* := \max_{x \in Z^2} T_n(x)$. Erdős et Taylor [17] ont prouvé que, presque sûrement,

$$(1) \quad \frac{1}{4\pi} \leq \liminf_{n \rightarrow \infty} \frac{T_n^*}{(\log n)^2} \leq \limsup_{n \rightarrow \infty} \frac{T_n^*}{(\log n)^2} \leq \frac{1}{\pi}.$$

Ils ont, de plus, conjecturé le résultat suivant :

CONJECTURE 0.1 (Erdős et Taylor [17]). — *Presque sûrement,*

$$\lim_{n \rightarrow \infty} \frac{T_n^*}{(\log n)^2} = \frac{1}{\pi}.$$

Le but de cet exposé est de faire une présentation succincte et non technique des travaux de Dembo, Peres, Rosen et Zeitouni ([5]–[13]) qui démontrent, entre autres, cette conjecture. Ces travaux apportent également une réponse définitive à plusieurs autres problèmes ouverts concernant la marche aléatoire et le mouvement brownien, dont les énoncés sont d'une simplicité remarquable – à l'image de la conjecture 0.1. En particulier, ils décrivent, de façon précise, la nature multi-fractale d'une nouvelle classe d'ensembles de points exceptionnels liés au recouvrement d'une partie de R^d pour le mouvement brownien.

1. POINTS FAVORIS ET POINTS ÉPAIS

1.1. Marche aléatoire

Commençons par définir rigoureusement une marche aléatoire en dimension 2 : on considère une trajectoire aléatoire $S : N \rightarrow Z^2$, issue de $S_0 = 0$, telle que les accroissements $\xi_n := S_n - S_{n-1}$ soient des variables aléatoires indépendantes et identiquement distribuées, dont la loi est donnée par $P(\xi_n = e) = \frac{1}{4}$, pour tout e tel que $\|e\| = 1$, où « $\|\cdot\|$ » désigne la norme euclidienne sur Z^2 .

Posons, comme ci-dessus, $T_n(x) := \#\{i : 0 \leq i \leq n, S_i = x\}$ et $T_n^* := \max_{x \in Z^2} T_n(x)$. Voici une réponse affirmative à la conjecture 0.1:

THÉORÈME 1.1 ([10]). — *Presque sûrement,*

$$\lim_{n \rightarrow \infty} \frac{T_n^*}{(\log n)^2} = \frac{1}{\pi}.$$

Rappelons que, pour tout $x \in Z^2$ fixé, $T_n(x)$ est approximativement de l'ordre de $\log n$ lorsque $n \rightarrow \infty$ (voir Erdős et Taylor [17] pour un énoncé précis). Un point x tel que $T_n(x) \geq \alpha(\log n)^2$ (pour un $\alpha > 0$) est donc «beaucoup visité» par la marche aléatoire. Le théorème suivant décrit la taille de l'ensemble de ces points qui sont beaucoup visités.

THÉORÈME 1.2 ([10]). — (i) *Pour tout $\alpha \in]0, \frac{1}{\pi}]$,*

$$(2) \quad \lim_{n \rightarrow \infty} \frac{\log \#\{x \in Z^2 : T_n(x) \geq \alpha(\log n)^2\}}{\log n} = 1 - \alpha\pi, \quad \text{p.s.}$$

(ii) *Presque sûrement, pour toute suite aléatoire $(x_n)_{n \geq 0}$ telle que $T_n(x_n) = T_n^*$, on a*

$$\lim_{n \rightarrow \infty} \frac{\log \|x_n\|}{\log n} = \frac{1}{2}.$$

Remarquons que (2) donne la borne inférieure cruciale qui manquait dans la conjecture d'Erdős et Taylor [17].

Un point favori (à l'étape n) est un point $x \in Z^2$ tel que $T_n(x) = T_n^*$ (notion introduite par Erdős et Révész [15]). La seconde partie du théorème 1.2 nous dit donc que les points favoris sont, dans l'échelle logarithmique, plutôt près de la frontière de l'ensemble des points visités.

Il est à signaler que très peu de propriétés sont connues à ce jour pour les points favoris de la marche aléatoire. On peut consulter le livre de Révész [23] (pages 130–131) pour une liste de dix questions ouvertes, dont la plupart ont été reprises de l'article d'Erdős et Révész [15]. Signalons, par exemple, l'une des questions fondamentales, à savoir la possibilité d'avoir au moins trois points favoris au fil du temps. Cette question, comme tant d'autres, reste sans réponse même en dimension 1, malgré un joli résultat partiel de Tóth [28].

1.2. Mouvement brownien

Il existe un résultat analogue à celui du théorème 1.2 pour un processus aléatoire pour lequel le temps est continu. Pour introduire ce nouveau processus aléatoire, on considère une marche aléatoire $(S_n, n \geq 0)$ à valeurs dans Z^d (avec $d \geq 1$ quelconque, pour l'instant), et on effectue un changement d'échelle

$$W_t^{(N)} := \left(\frac{d}{N}\right)^{1/2} S_{[Nt]}, \quad t \geq 0.$$

Le théorème central limite assure que pour tous $0 \leq t_1 \leq \dots \leq t_k$, $(W_{t_1}^{(N)}, \dots, W_{t_k}^{(N)})$ converge en loi (lorsque N tend vers l'infini) vers $(W_{t_1}, \dots, W_{t_k})$, où $(W_t, t \geq 0)$ est un processus aléatoire à valeurs dans R^d tel que, pour tous $0 \leq s \leq t$, $W_t - W_s$ suit la loi de Gauss de moyenne nulle et de matrice de covariances $(t - s)\text{Id}$. On peut faire en sorte que $t \mapsto W_t$ soit une fonction continue sur R_+ . Le processus $(W_t, t \geq 0)$ porte le nom de «mouvement brownien».

Il est souvent plus aisé d'étudier le mouvement brownien que la marche aléatoire, grâce, par exemple, à la propriété d'auto-similarité (ou, en français, propriété de «scaling») du mouvement brownien. On reviendra, dans la section 5.3, sur une relation trajectorielle entre le mouvement brownien et la marche aléatoire, en plus de la convergence en loi décrite ci-dessus.

Dans cette section, on suppose $d = 2$. Soit $\bar{\theta} := \inf\{t \geq 0 : \|W_t\| = 1\}$. Considérons la mesure d'occupation

$$\mu_{\bar{\theta}}(A) := \int_0^{\bar{\theta}} \mathbf{1}_A(W_t) dt, \quad \forall A \subset R^2 \text{ borélien,}$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de A . L'analogie du théorème 1.2 pour le mouvement brownien s'énonce comme suit. On note $B(x, r)$ le disque ouvert (dans R^2) centré en x et de rayon r , et $\dim(A)$ la dimension de Hausdorff de A .

THÉORÈME 1.3 ([10]). — (i) *Pour tout $a \in]0, 2]$, on a, presque sûrement,*

$$(3) \quad \dim \left\{ x \in R^2 : \lim_{\varepsilon \rightarrow 0} \frac{\mu_{\bar{\theta}}(B(x, \varepsilon))}{\varepsilon^2 \log^2(1/\varepsilon)} = a \right\} = 2 - a.$$

(ii) *Presque sûrement,*

$$(4) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{x \in R^2} \frac{\mu_{\bar{\theta}}(B(x, \varepsilon))}{\varepsilon^2 \log^2(1/\varepsilon)} = 2.$$

À titre de comparaison, rappelons que, typiquement, pour un point x sur la trajectoire de $(W(t), t \in [0, \bar{\theta}])$, $\mu_{\bar{\theta}}(B(x, \varepsilon))$ se comporte à peu près comme $\varepsilon^2 \log(1/\varepsilon)$ lorsque ε est petit (voir Ray [22] pour un énoncé précis). Un point x dans l'ensemble considéré dans (3), qui est donc en quelque sorte «souvent visité», est appelé «point épais» (*thick point* en anglais).

La partie (ii) du théorème 1.3 fournit une réponse affirmative à une conjecture de Perkins et Taylor [21]. Perkins et Taylor ont obtenu la borne supérieure dans (4), et une borne inférieure non optimale (quatre fois plus petite que la limite conjecturée).

2. POINTS TARDIFS ET TEMPS DE RECOUVREMENT

2.1. Marche aléatoire sur un compact

Question : Combien de temps faut-il pour qu'un graphe fini soit recouvert par une marche aléatoire ?

Il s'agit d'un problème important en probabilités, en combinatoire, et en informatique. L'exemple du tore bidimensionnel $Z_n^2 := Z^2/nZ^2$ est le plus célèbre, et est précisément ce qui nous intéresse ici. Soit C_n le temps nécessaire pour que le tore Z_n^2 soit recouvert par une marche aléatoire sur Z_n^2 .

Il y a une quinzaine d'années, Aldous et Lawler ont obtenu des bornes asymptotiques pour C_n : lorsque $n \rightarrow \infty$,

$$(5) \quad \frac{2}{\pi} + o(1) \leq \frac{C_n}{(n \log n)^2} \leq \frac{4}{\pi} + o(1), \quad \text{en probabilité,}$$

la borne supérieure étant due à Aldous [1], et la borne inférieure à Lawler [19].

Aldous [1] a de plus conjecturé que la borne supérieure est optimale, ce qui a été récemment confirmé par le théorème suivant :

THÉORÈME 2.1 ([12]). — *Si C_n désigne le temps de recouvrement de Z_n^2 par une marche aléatoire sur Z_n^2 , alors*

$$\lim_{n \rightarrow \infty} \frac{C_n}{(n \log n)^2} = \frac{4}{\pi}, \quad \text{en probabilité.}$$

Le problème du temps de recouvrement est, en quelque sorte, le dual de la conjecture de Perkins et Taylor énoncée dans (4).

De même que pour le théorème 1.2 sur les points beaucoup visités par la marche aléatoire, il existe un résultat analogue à celui du théorème 2.1 pour la courbe du mouvement brownien (ou plutôt, pour son voisinage, la saucisse de Wiener).

La preuve du théorème 2.1 réserve une étude spéciale concernant les « points tardifs » (*late points* en anglais) qui sont les derniers à être atteints avant le complet recouvrement. La preuve décrit, en outre, le spectre multi-fractal de l'ensemble des points tardifs.

Pour plus de détails, voir [12].

2.2. Marche aléatoire sur Z^2

Passons maintenant aux problèmes de recouvrement pour la marche aléatoire sur Z^2 (et non plus sur le tore).

On peut se poser deux questions.

Question 1 : Quel est le rayon ϱ_n du plus grand disque, centré à l'origine, recouvert par les n premiers pas de la marche aléatoire sur Z^2 ?

Question 2 : Quel est le rayon R_n du plus grand disque recouvert par les n premiers pas de la marche aléatoire sur Z^2 ?

Commençons par la question 1. Les premiers résultats, dus à Révész [23], ont montré l'existence de constantes $0 < a < b < \infty$ telles que, pour tout $y > 0$,

$$(6) \quad e^{-by} \leq \liminf_{n \rightarrow \infty} P \left(\frac{\log^2 \varrho_n}{\log n} \geq y \right) \leq \limsup_{n \rightarrow \infty} P \left(\frac{\log^2 \varrho_n}{\log n} \geq y \right) \leq e^{-ay}.$$

Révész a conjecturé l'existence d'une constante λ , sans précision sur sa valeur, telle que $P(\frac{\log^2 \varrho_n}{\log n} \geq y)$ converge vers $e^{-\lambda y}$. Lawler [19] a prouvé (6) avec les constantes $a = 2$ et $b = 4$, et a mentionné une conjecture de Kesten qui consistait à dire que la limite vaudrait e^{-4y} . Ceci est confirmé par le théorème suivant :

THÉORÈME 2.2 ([12]). — *Si ϱ_n désigne le rayon du plus grand disque centré à l'origine totalement recouvert par la marche aléatoire sur Z^2 lors des n premières étapes, alors, pour tout $y > 0$,*

$$\lim_{n \rightarrow \infty} P \left(\frac{\log^2 \varrho_n}{\log n} \geq y \right) = e^{-4y}.$$

On peut donc dire que, grosso modo, ϱ_n se comporte comme $\exp(\sqrt{\log n})$.

Passons à la question 2.

Il apparaît que R_n est beaucoup plus grand que ϱ_n . En effet, on a, presque sûrement, pour tout n suffisamment grand,

$$n^{\theta_1} \leq R_n \leq n^{\theta_2}.$$

Ces deux bornes sont dues à Révész qui a démontré le résultat pour $\theta_1 := 0,02$ ([24]) et pour $\theta_2 := 0,42$ ([25]). Il était donc naturel de penser (Révész [24]) que $R_n = n^{\theta+o(1)}$ p.s., pour un certain exposant $\theta \in [\theta_1, \theta_2]$. Ceci est confirmé par le résultat suivant :

THÉORÈME 2.3 ([6]). — *Soit R_n le rayon du plus grand disque recouvert par la marche aléatoire sur Z^2 dans les n premières étapes. On a, presque sûrement,*

$$(7) \quad \lim_{n \rightarrow \infty} \frac{\log R_n}{\log n} = \frac{1}{4}.$$

Signalons que la preuve du Théorème 2.3 a permis de répondre à la question suivante : quel est le comportement presque sûr, en limite supérieure, du temps nécessaire à la marche aléatoire pour rencontrer un point qu'elle n'a encore jamais visité ? Il s'agissait là d'un problème ouvert d'Erdős et Révész [16].

3. DIMENSIONS SUPÉRIEURES

La dimension 2 est critique pour la marche aléatoire et pour le mouvement brownien, son analogue à temps continu. C'est la raison pour laquelle l'étude de ces processus aléatoires en dimension 2 est particulièrement compliquée. Les notes de cours de Le Gall [20] restent une référence de base pour l'étude du mouvement brownien en dimension 2. Sur ce sujet, de récents progrès, parmi les plus importants en théorie des probabilités, faisant intervenir les SLE, ont été effectués depuis (voir les notes de cours de Werner [29]).

Dans cette section, on s'intéresse au cas $d \geq 3$. Soit $W = (W_t, t \geq 0)$ un mouvement brownien à valeurs dans R^d , issu de $W_0 = 0$. Il est connu que W est transient, c'est-à-dire que, presque sûrement, $\|W_t\| \rightarrow \infty$, lorsque $t \rightarrow \infty$. On définit

$$\mu_\infty(A) := \int_0^\infty \mathbf{1}_A(W_t) dt, \quad A \subset R^d \text{ borélien.}$$

Soit q_d le plus petit réel positif tel que $J_{(d-2)/2}(q_d) = 0$, où $J_{(d-2)/2}(\cdot)$ désigne la fonction de Bessel d'indice $(d-2)/2$.

THÉORÈME 3.1 ([8]). — *Supposons $d \geq 3$. Pour tout $a \in]0, \frac{4}{q_d^2}]$, on a, presque sûrement,*

$$\dim \left\{ x \in R^d : \limsup_{\varepsilon \rightarrow 0} \frac{\mu_\infty(B(x, \varepsilon))}{\varepsilon^2 \log(1/\varepsilon)} = a \right\} = 2 - \frac{aq_d^2}{2}.$$

En plus, pour tout $r > 0$,

$$(8) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{\|x\| \leq r} \frac{\mu_\infty(B(x, \varepsilon))}{\varepsilon^2 \log(1/\varepsilon)} = \frac{4}{q_d^2}, \quad \text{p.s.}$$

Par conséquent,

$$\sup_{x \in R^d} \limsup_{\varepsilon \rightarrow 0} \frac{\mu_\infty(B(x, \varepsilon))}{\varepsilon^2 \log(1/\varepsilon)} = \frac{4}{q_d^2}, \quad \text{p.s.}$$

L'identité (8) donne une réponse affirmative à une conjecture de Taylor [27].

La présence de la constante q_d , dont la valeur dépend de la fonction de Bessel $J_{(d-2)/2}(\cdot)$, est due à une relation très curieuse, découverte par Ciesielski et Taylor [2], entre le temps d'occupation de la boule unité par le mouvement brownien de dimension $d \geq 3$, et le temps d'atteinte de la sphère unité par le mouvement brownien de dimension $d-2$.

Le mouvement brownien en dimension $d \geq 3$ possède une propriété de « localisation » : les passages de W dans $B(x, \varepsilon)$ se trouvent, pour l'essentiel, concentrés dans un très faible intervalle de temps (alors qu'en dimension 2, le mouvement brownien revient très régulièrement au voisinage de x). L'énoncé précis de cette propriété peut être trouvé dans [8]. Cette propriété facilite l'étude des points épais en dimension $d \geq 3$, et permet de se passer de la très délicate « méthode multi-échelle du second moment » évoquée dans la section 5.2.

4. POINTS FINS

Soit $(W_t, t \geq 0)$ un mouvement brownien à valeurs dans R^d (avec $d \geq 2$), issu de $B_0 = 0 \in R^d$. Fixons $T > 0$. Soit

$$\mu_T(A) := \int_0^T \mathbf{1}_A(W_t) dt, \quad \forall A \subset R^d \text{ borélien.}$$

On peut voir que pour tout point x de la trajectoire $\{W_t, t \in [0, T]\}$, $\mu_T(B(x, \varepsilon))$ est au moins de l'ordre de $\varepsilon^2 / \log(1/\varepsilon)$. Le résultat suivant nous confirme que cette borne est effectivement atteinte par certains points exceptionnels, et donne la dimension de Hausdorff de l'ensemble de ces points.

THÉORÈME 4.1 ([9]). — *Soit $d \geq 2$. Fixons $T > 0$. Pour tout $a \geq 1$, on a, presque sûrement,*

$$(9) \quad \dim \left\{ x \in R^d : \liminf_{\varepsilon \rightarrow 0} \frac{\mu_T(B(x, \varepsilon))}{\varepsilon^2 / \log(1/\varepsilon)} = a \right\} = 2 - \frac{2}{a}.$$

De plus,

$$\inf_{t \in]0, T[} \liminf_{\varepsilon \rightarrow 0} \frac{\mu_T(B(W_t, \varepsilon))}{\varepsilon^2 / \log(1/\varepsilon)} = 1, \quad \text{p.s.}$$

Un point x dans l'ensemble considéré dans (9), qui est donc très peu visité, est appelé «point fin» (*thin point* en anglais).

5. QUELQUES IDÉES DE PREUVE

Je tâcherai ici de faire un résumé, le moins technique possible, sur les idées de base des travaux de Dembo, Peres, Rosen et Zeitouni. Pour plus de détails, on peut consulter les notes de cours de Dembo [4].

5.1. Difficultés essentielles

On remarque immédiatement que ces travaux répondent à des conjectures, souvent de longue date, dont les bornes supérieures et inférieures différaient par un facteur de 4 ou 2 (conjecture d'Erdős et Taylor dans (1), conjecture de Perkins et Taylor dans (4), conjecture d'Aldous dans (5), conjecture de Kesten–Révész dans (6), conjecture de Révész dans (7), conjecture de Taylor dans (8)). En réalité, certains éléments fondamentaux manquaient à la compréhension de ces problèmes.

Prenons par exemple le théorème 1.3. On choisit un réseau de points (x_j) dans $B(0, 1)$, et on considère

$$Z := \sum_j \mathbf{1}_{\left\{ \frac{\mu_T(B(x_j, \varepsilon))}{\varepsilon^2 \log^2(1/\varepsilon)} \geq a \right\}}.$$

Le réseau (x_j) est choisi de sorte que, pour tout x , il existe x_j suffisamment proche de x tel que le temps d'occupation autour de x est approximativement celui autour de x_j . Donc, l'événement $\{Z \geq 1\}$ correspond à peu près à la situation

$$\sup_{x \in B(0,1)} \frac{\mu_{\bar{\theta}}(B(x, \varepsilon))}{\varepsilon^2 \log^2(1/\varepsilon)} \geq a.$$

D'autre part, il est possible d'estimer $E(Z)$. En utilisant la méthode du premier moment (à savoir $P(Z \geq 1) \leq E(Z)$ par l'inégalité de Markov), Perkins et Taylor [21] ont pu obtenir une borne supérieure pour $\sup_{x \in B(0,1)} \frac{\mu_{\bar{\theta}}(B(x, \varepsilon))}{\varepsilon^2 \log^2(1/\varepsilon)}$. La borne supérieure ainsi obtenue s'est avérée être optimale.

En revanche, à cause des fortes corrélations des temps d'occupation autour de chacun des points du réseau, $E(Z^2)$ est tellement grande que la méthode du second moment n'aboutit pas.

Une approche naturelle est de tronquer certaines excursions (opération qui revient à ignorer les visites trop éloignées dans le temps), ce qui réduit effectivement le moment d'ordre 2, mais conduit hélas à une réduction simultanée de la moyenne. On obtient ainsi le résultat de Perkins et Taylor [21] qui laisse un facteur 4 entre les bornes supérieure et inférieure.

Cette remarque s'applique en effet à tous les problèmes cités dans cet exposé.

La principale nouveauté, dans les travaux de Dembo, Peres, Rosen et Zeitouni, a été la mise au point d'une nouvelle méthode qui assouplit la méthode du second moment, bien au-delà de ce qu'avait pu faire auparavant la méthode de troncature.

Nous allons expliquer, dans le paragraphe suivant, les grandes lignes de cette méthode, dans son application au mouvement brownien en dimension 2. Le passage du mouvement brownien à la marche aléatoire sera ensuite expliqué dans la section 5.3.

5.2. Méthode multi-échelle du second moment

On procède en deux étapes.

Étape 1. Commençons par étudier la marche aléatoire d'une particule sur un arbre régulier de hauteur N , où chaque sommet, qui n'est pas une feuille, admet b descendants immédiats.

À chaque pas, la particule se déplace de façon équiprobable vers l'un de ses $b + 1$ voisins, et revient sur ses pas lorsqu'elle se trouve sur une des b^N feuilles de l'arbre. La marche aléatoire part de la feuille la plus à gauche. On considère sa trajectoire jusqu'au premier instant où elle touche la racine de l'arbre.

Une feuille x est dite «très visitée» par la marche aléatoire si le nombre de passages en x est au moins de l'ordre de N^2 (lorsque N est grand). On cherche désormais à minorer le cardinal M_N de l'ensemble (aléatoire) de ces feuilles très visitées.

On peut assez facilement estimer $E(M_N)$, mais il se trouve que la variance de M_N est si grande, que la méthode du second moment, une nouvelle fois, n'aboutit pas.

L'idée va être d'introduire un sous-ensemble de l'ensemble des feuilles très visitées dont l'espérance du cardinal sera comparable à $E(M_N)$ mais dont la variance sera considérablement plus faible que celle de M_N !

On qualifiera les éléments de ce sous-ensemble, des feuilles « n -parfaites». Définissons-les.

Commençons par poser $h_k := \lfloor k \log k \rfloor$, pour $k = 2, 3, \dots, \lfloor N/\log N \rfloor =: n$. Une feuille x est alors dite « n -parfaite» si, pour tout $k \in \{2, 3, \dots, n\}$, le nombre des excursions effectuées par la marche aléatoire, entre l'ancêtre de x à la génération h_{k-1} et l'ancêtre de x à la génération h_k , est approximativement $k^2 \log k$.

Il se trouve que l'espérance du cardinal de l'ensemble des feuilles « n -parfaites» est proche de $E(M_N)$, et que, dans le même temps, la variance de ce cardinal est bien plus faible que celle de M_N . Pour évaluer cette variance, il s'agit d'estimer la probabilité que deux feuilles soient toutes deux « n -parfaites». La méthode du second moment usuelle nous permet alors de dire que le nombre total des points « n -parfaits» est relativement proche de son espérance, ce qui nous amène finalement à une borne inférieure satisfaisante pour M_N .

Étape 2. Soit W un mouvement brownien en dimension 2. Introduisons le réseau des points (x_j) comme dans la section 5.1, et considérons, pour chaque point x_j du réseau, des disques concentriques $(B(x_j, \varepsilon_k))_{k \geq 1}$, centrés au point x_j , où (ε_k) est une suite qui décroît assez rapidement vers 0. On choisit en fait $\varepsilon_k := \exp(-k \log k)$, de sorte que $\log(1/\varepsilon_k)$ correspond au h_k de l'étape précédente. Soit $N_k^{x_j}$ le nombre des excursions effectuées par W entre les cercles $\partial B(x_j, \varepsilon_{k-1})$ et $\partial B(x_j, \varepsilon_k)$.

En utilisant un résultat de grandes déviations, concernant la concentration de la somme de variables aléatoires i.i.d. autour de sa moyenne, on peut estimer le temps d'occupation d'une couronne à l'aide du nombre d'excursions la traversant. Il s'ensuit que, presque sûrement, pour qu'un point x_j du réseau soit épais, il suffit que $N_k^{x_j}$ soit de l'ordre de $k^2 \log k$ pour tout k suffisamment grand.

On conclut en reliant le problème du comptage des points épais pour W avec le problème du comptage des feuilles « n -parfaites». L'estimation du second moment du cardinal des points épais du réseau se ramène à l'examen de la probabilité que deux points du réseau soient simultanément épais. Cette question est de même nature que celle de la probabilité que deux feuilles soient simultanément « n -parfaites», considérée à l'étape 1.

5.3. Passage du mouvement brownien à la marche aléatoire

Nous avons commencé cet exposé en introduisant la marche aléatoire, qui est un objet facile à définir. Nous avons tout de même mentionné que, du point de vue technique, il est plus facile d'étudier le mouvement brownien.

À partir d'un résultat pour le mouvement brownien, on peut souvent déduire un résultat analogue pour la marche aléatoire, à l'aide d'un théorème célèbre dû à Komlós,

Major et Tusnády [18], qui consiste à dire que, éventuellement sur un espace de probabilité élargi, on peut construire un mouvement brownien et une marche aléatoire tels que la différence de ces deux processus soit «très petite». Ce théorème de couplage, qui porte le nom de KMT, permet souvent de faire le passage entre un résultat pour le mouvement brownien et son analogue pour la marche aléatoire.

Le théorème de KMT est valable en dimension 1. Grâce à un argument géométrique simple, on peut l'étendre à la dimension 2. C'est donc ce théorème classique qui est à l'origine de la justification en dimension 2. (Attention : malgré tout, il faut encore pas mal de travail, car théoriquement, deux processus relativement proches l'un de l'autre peuvent avoir des mesures d'occupation très différentes.)

En dimension $d \geq 3$, on utilise une version multi-dimensionnelle du théorème de KMT, due à Einmahl [14], pour justifier le passage entre le mouvement brownien et la marche aléatoire.

Signalons, pour terminer, qu'il est possible de reprendre la méthode présentée dans cet exposé pour démontrer la conjecture d'Erdős–Taylor, uniquement par des considérations sur la marche aléatoire elle-même et sans utiliser le mouvement brownien comme outil (voir [26]).

6. QUELQUES REMARQUES

La plupart des résultats pour la marche aléatoire présentés dans cet exposé s'appliquent en fait à une grande classe de marches aléatoires au sens général. De telles marches ne sont pas contraintes à effectuer des pas de longueur 1. Les résultats restent alors valables sous certaines conditions sur la taille des pas (voir [10]).

On peut également étendre quelques-uns des résultats sur le mouvement brownien à certains processus aléatoires dont les trajectoires ne sont pas continues. Par exemple, les résultats sur les points épais du mouvement brownien en dimension $d \geq 3$ ont été étendus pour les processus stables symétriques transients ([7]).

La méthode développée par Dembo, Peres, Rosen et Zeitouni permet de traiter d'autres problèmes de recouvrement, par exemple, celui de la dimension de Hausdorff des points épais liés à l'intersection de plusieurs mouvements browniens indépendants ([11]), ainsi que celui de la couverture des disques par plusieurs marches aléatoires indépendantes ([6]), ou encore le problème du temps de recouvrement d'une variété riemannienne compacte par le mouvement brownien ([5], [12]). Signalons aussi que la méthode s'applique à l'étude de certaines interfaces aléatoires en mécanique statistique ([3]). Par manque de temps, ces aspects n'ont pas été abordés dans cet exposé.

RÉFÉRENCES

- [1] D. ALDOUS – *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York, 1989.
- [2] Z. CIESIELSKI et S.J. TAYLOR – First passage times and sojourn times for Brownian motion in space and the exact Hausdorff measure of the sample path. *Trans. Amer. Math. Soc.* **103** (1962), 434–450.
- [3] O. DAVIAUD – Extremes of the discrete two-dimensional Gaussian free field. Prépublication disponible sur ArXiv math.PR/0406609.
- [4] A. DEMBO – Lectures on favorite points, cover times and fractals. À paraître dans *École d’été de probabilités de Saint-Flour 2003*, Springer.
- [5] A. DEMBO, Y. PERES et J. ROSEN – Brownian motion on compact manifolds: cover time and late points. *Electron. J. Probab.* **8** (2003), no. 15.
- [6] A. DEMBO, Y. PERES et J. ROSEN – How large a disc is covered by a random walk in n steps? (Prépublication.)
- [7] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Thick points for transient symmetric stable processes. *Electron. J. Probab.* **4** (1999), no. 10.
- [8] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Thick points for spatial Brownian motion: multifractal analysis of occupation measure. *Ann. Probab.* **28** (2000), 1–35.
- [9] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Thin points for Brownian motion. *Ann. Inst. H. Poincaré Probab. Statist.* **36** (2000), 749–774.
- [10] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Thick points for planar Brownian motion and the Erdős–Taylor conjecture on random walk. *Acta Math.* **186** (2001), 239–270.
- [11] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Thick points for intersections of planar sample paths. *Trans. Amer. Math. Soc.* **354** (2002), 4969–5003.
- [12] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Cover times for Brownian motion and random walks in two dimensions. *Ann. Math.* **160** (2004), 433–464.
- [13] A. DEMBO, Y. PERES, J. ROSEN et O. ZEITOUNI – Late points for random walks in two dimensions. À paraître dans *Ann. Probab.*
- [14] U. EINMAHL – Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivar. Anal.* **28** (1989), 20–68.
- [15] P. ERDŐS et P. RÉVÉSZ – On the favourite points of a random walk. *Mathematical Structures – Computational Mathematics – Mathematical Modelling* **2** pp. 152–157. Publ. House Bulgar. Acad. Sci., Sofia, 1984.
- [16] P. ERDŐS et P. RÉVÉSZ – Three problems on the random walk in Z^d . *Studia Sci. Math. Hungar.* **26** (1991), 309–320.
- [17] P. ERDŐS et S.J. TAYLOR – Some problems on the structure of random walk paths. *Acta Math. Sci. Hungar.* **11** (1960), 137–162.

- [18] J. KOMLÓS, P. MAJOR et G. TUSNÁDY – An approximation of partial sums of independent RV’s and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** (1975), 111–131.
- [19] G.F. LAWLER - On the covering time of a disc by simple random walk in two dimensions. *Seminar on Stochastic Processes 1992*, 189–207, Progr. Probab., **33**. Birkhäuser, Boston, 1993.
- [20] J.-F. LE GALL - Some properties of planar Brownian motion. *École d’été de probabilités de Saint-Flour 1990*. Lecture Notes in Math. **1527**, pp. 111–235. Springer, Berlin, 1992.
- [21] E.A. PERKINS et S.J. TAYLOR – Uniform measure results for the image of subsets under Brownian motion. *Probab. Theory Related Fiends* **76** (1987), 257–289.
- [22] D. RAY – Sojourn times and the exact Hausdorff measure of the sample path for planar Brownian motion. *Trans. Amer. Math. Soc.* **106** (1963), 436–444.
- [23] P. RÉVÉSZ – *Random Walk in Random and Non-Random Environments*. World Scientific, Singapore, 1990.
- [24] P. RÉVÉSZ – Clusters of a random walk on the plane. *Ann. Probab.* **21** (1993), 318–328.
- [25] P. RÉVÉSZ – Covering problems. *Theory Probab. Appl.* **38** (1993), 367–379.
- [26] J. ROSEN – A random walk proof of the Erdős–Taylor conjecture. *Period. Math. Hung.* **50** (2005), 223–245.
- [27] S.J. TAYLOR – Regularity of irregularities on a Brownian path. *Ann. Inst. Fourier* **24** (1974), 195–203.
- [28] B. TÓTH – No more than three favorite sites for simple random walk. *Ann. Probab.* **29** (2001), 484–503.
- [29] W. WERNER – Random planar curves and Schramm–Loewner evolutions. *École d’été de probabilités de Saint-Flour 2002*. Lecture Notes in Math. **1840**, pp. 107–195. Springer, Berlin, 2004.

Zhan SHI

Université Paris VI
 Laboratoire de Probabilités et
 Modèles Aléatoires
 CNRS UMR 7599
 4 place Jussieu
 F-75252 PARIS Cedex 05
E-mail : zhan@proba.jussieu.fr

**COMPACTIFICATION DE L'ESPACE DES MODULES DES
VARIÉTÉS ABÉLIENNES PRINCIPALEMENT POLARISÉES**

[d'après V. Alexeev]

par Michel BRION

INTRODUCTION

Classiquement, les variétés abéliennes complexes de dimension g munies d'une polarisation principale sont paramétrées par le quotient A_g du demi-espace de Siegel \mathcal{H}_g sous l'action du groupe symplectique entier $\mathrm{Sp}_{2g}(\mathbb{Z})$. L'espace des modules A_g est un espace analytique complexe de dimension $g(g+1)/2$ n'ayant que des singularités quotient par des groupes finis. En fait, A_g est un ouvert de Zariski d'une variété projective \overline{A}_g^{\min} (construite par Satake, Baily et Borel dans le cadre plus général des espaces localement symétriques) : la compactification minimale, dont le bord est de codimension g . La variété \overline{A}_g^{\min} est en général bien plus singulière que A_g , mais on en connaît des désingularisations partielles : les compactifications toroïdales (construites pour les espaces localement symétriques par Ash, Mumford, Rapoport et Tai) dont le bord est un diviseur, et qui n'ont que des singularités quotient par des groupes finis.

Toutes ces compactifications de A_g admettent des modèles sur les entiers, les compactifications arithmétiques de Faltings et Chai. Cependant, elles sont construites par des procédés ad hoc qui n'en donnent pas d'interprétation modulaire, à savoir, comme espaces de paramètres d'objets géométriques (ce sens de l'adjectif «modulaire» est sans rapport avec les formes modulaires, qui ont des liens étroits avec la compactification minimale).

Des exemples importants de variétés abéliennes principalement polarisées sont les jacobiniennes des courbes algébriques irréductibles, lisses et complètes de genre $g \geq 2$. Ces courbes admettent un espace des modules M_g dont on connaît cette fois une compactification modulaire \overline{M}_g , paramétrant les courbes stables de genre arithmétique g . En associant à chaque courbe sa jacobienne, on obtient un morphisme $\mathfrak{t} : M_g \rightarrow A_g$ qui est injectif d'après le théorème de Torelli ; de plus, M_g , \overline{M}_g et \mathfrak{t} admettent des modèles entiers. La question se pose alors de construire une compactification modulaire et canonique de A_g , définie sur les entiers et qui permette de compactifier le morphisme de Torelli \mathfrak{t} .

Les travaux [2, 3, 4] d'Alexeev apportent une réponse complète à cette question. Sa compactification $\overline{A}_g^{\mathrm{mod}}$ est un espace des modules de «couples quasi-abéliens stables» ; il s'agit des couples (X, D) où X est une variété projective (connexe, mais non

nécessairement irréductible) dans laquelle une variété semi-abélienne G opère avec un nombre fini d'orbites, et D est un diviseur effectif et ample sur X qui ne contient aucune de ces orbites. On suppose de plus que X est équidimensionnelle de dimension g et semi-normale (c'est une petite restriction sur ses singularités) et que les stabilisateurs de l'action de G sont des tores.

L'exemple le plus simple d'un tel couple est formé d'une variété abélienne opérant dans elle-même par translations, et d'un diviseur thêta. Un exemple plus singulier est celui où X est une cubique plane nodale munie de l'action du groupe multiplicatif G et du diviseur D formé d'un point distinct du point double ; c'est une dégénérescence des cubiques planes lisses munies d'un point, c'est-à-dire des courbes elliptiques.

À tout couple quasi-abélien stable on peut associer un complexe de polytopes convexes entiers appelé son type. Les couples dont le type est un «pavage périodique par des polytopes convexes entiers» d'un espace vectoriel \mathbb{R}^r , $r \leq g$, sont paramétrés par la compactification modulaire $\overline{A}_g^{\text{mod}}$. Parmi ces couples, on trouve ceux associés comme précédemment aux variétés abéliennes principalement polarisées (c'est le cas où $r = 0$), et aussi les jacobiniennes compactifiées des courbes stables de genre arithmétique g . Ceci permet d'obtenir un morphisme de Torelli compactifié $\bar{\mathfrak{t}} : \overline{M}_g \rightarrow \overline{A}_g^{\text{mod}}$; son image est contenue dans l'adhérence de A_g , une composante irréductible de $\overline{A}_g^{\text{mod}}$ dont la normalisation est une compactification toroïdale particulière, notée $\overline{A}_g^{\text{Vor}}$.

En général, $\overline{A}_g^{\text{mod}}$ contient d'autres composantes irréductibles [2] ; autrement dit, ce n'est pas une compactification de A_g au sens usuel. Une description modulaire de la «composante principale» $\overline{A}_g^{\text{Vor}}$ est proposée par Olsson [21] en termes de géométrie logarithmique. Mais une question ouverte est de construire des compactifications modulaires des espaces plus généraux $A_{g,d}$ qui paramètrent les variétés abéliennes de dimension g munies d'une polarisation de degré $d \geq 2$; de même pour les espaces $A_{g,d,n}$ où on se donne aussi une structure de niveau n .

La définition des couples quasi-abéliens stables semble assez arbitraire : pourquoi faudrait-il s'intéresser à des objets aussi singuliers ? En fait, la construction d'espaces des modules de variétés projectives et lisses fait apparaître des objets très analogues : afin de pouvoir considérer de telles variétés X dont la classe canonique K_X n'est pas ample (par exemple, les variétés abéliennes pour lesquelles K_X est triviale), on est amené à introduire des couples (X, D) où D est un diviseur effectif sur X tel que $K_X + D$ est ample. Et pour obtenir des espaces des modules complets, il faut autoriser des dégénérescences singulières de ces couples en des «couples stables».

Les courbes stables pointées forment le premier exemple de tels couples ; d'autres exemples importants sont les surfaces stables de [11]. Dans le manuscrit [1], Alexeev formule une définition générale des couples stables, et montre que l'existence d'un espace des modules complet pour ceux-ci se déduit d'un ensemble de conjectures dans la classification des variétés algébriques : le programme de Mori logarithmique. Ces conjectures sont toujours ouvertes en grande dimension, et la construction de $\overline{A}_g^{\text{mod}}$ s'obtient par des méthodes spécifiques liées aux symétries des variétés considérées.

Le but de ce texte est d'exposer une partie des résultats des articles [2, 3, 4, 5] avec des prérequis modestes de géométrie algébrique (par exemple, le contenu du manuel [9]), dans l'espoir de rendre plus accessible un sujet où foisonnent les notations, les définitions et les concepts. C'est pourquoi on rassemble dans la première partie des résultats classiques sur les variétés abéliennes et leurs espaces des modules, tirés des ouvrages [13, 15, 8]. La seconde partie est consacrée à une construction de dégénérescences «maximales» de variétés abéliennes, qui fait apparaître beaucoup d'ingrédients de la compactification modulaire. Celle-ci fait l'objet de la troisième partie ; on y décrit la structure des couples stables qu'elle classe et on énonce les résultats principaux la concernant, en général sans démonstration détaillée.

Je remercie R. Bacher, O. Debarre, S. Druel et tout particulièrement V. Alexeev et G. Rémond pour des discussions très utiles et pour leurs commentaires sur les versions successives de ce texte ; il va de soi que je suis seul responsable des erreurs et imprécisions qui pourraient y subsister.

1. VARIÉTÉS ABÉLIENNES PRINCIPALEMENT POLARISÉES ET LEURS ESPACES DES MODULES

Dans tout ce texte, on appelle *variété* un schéma réduit, connexe, séparé et de type fini sur un corps algébriquement clos k ; avec cette convention, les variétés ne sont pas nécessairement intègres. On appelle *courbe* une variété de dimension pure 1. Enfin, on identifie chaque faisceau inversible au fibré en droites dont il est le faisceau des sections locales.

1.1. Variétés abéliennes

Une variété complète est dite *abélienne* si elle est munie d'une structure de groupe algébrique. Une telle variété A est intègre, projective et lisse, et sa loi de groupe est commutative ; on la note additivement. De plus, la structure de groupe sur la variété A est uniquement déterminée par la donnée de l'élément neutre 0 . Pour tout $a \in A$, on note

$$\tau_a : A \rightarrow A, \quad x \mapsto x + a$$

la translation par a .

Le sous-groupe du groupe de Picard de A formé des classes d'isomorphie des fibrés algébriquement équivalents au fibré trivial est noté $\text{Pic}^0(A)$ ou A^\vee ; c'est aussi une variété abélienne, la *duale* de A . Tout homomorphisme de variétés abéliennes $f : A \rightarrow B$ définit un homomorphisme dual $f^\vee : B^\vee \rightarrow A^\vee$, la restriction de $f^* : \text{Pic}(B) \rightarrow \text{Pic}(A)$.

Soit L un fibré en droites sur A . Pour tout $a \in A$, le fibré en droites $L^{-1} \otimes \tau_a^*(L)$ est algébriquement trivial ; on obtient ainsi un morphisme

$$\lambda_L : A \rightarrow A^\vee, \quad a \mapsto [L^{-1} \otimes \tau_a^*(L)]$$

qui est en fait un homomorphisme de groupes d'après le *théorème du carré*

$$L \otimes \tau_{a+b}^*(L) \simeq \tau_a^*(L) \otimes \tau_b^*(L) \quad \text{pour tous } a, b \in A.$$

Pour que λ_L soit trivial (c'est-à-dire $\tau_a^*(L) \simeq L$ pour tout $a \in A$), il faut et il suffit que $[L] \in \text{Pic}^0(A)$.

Lorsque le fibré en droites L est ample, λ_L est une *isogénie* (à savoir, un homomorphisme de groupes algébriques, surjectif et de noyau fini) et son degré est le carré de $h^0(L) := \dim H^0(A, L)$; de plus, $H^i(X, L) = 0$ pour tout $i \geq 1$. En particulier, A et A^\vee ont la même dimension, notée g , et le degré du fibré en droites ample L est $g! h^0(L)$.

Une *polarisation* de A est une isogénie

$$\lambda : A \rightarrow A^\vee$$

qui s'écrit sous la forme λ_L pour un fibré en droites ample L ; alors les fibrés en droites M tels que $\lambda = \lambda_M$ ne sont autres que les translatés $\tau_a^*(L)$, $a \in A$. Les classes de ces fibrés dans $\text{Pic}(A)$ forment un translaté de A^\vee noté $\text{Pic}^\lambda(A)$; l'entier positif $h^0(L) = h^0(\tau_a^*(L))$ est appelé le *degré* de la polarisation λ .

Une polarisation $\lambda = \lambda_L$ est dite *principale* si c'est un isomorphisme, c'est-à-dire si $h^0(L) = 1$; autrement dit, $L = \mathcal{O}_A(\Theta)$ pour un diviseur Θ effectif et ample, uniquement déterminé par L , et déterminé à translation près par λ . On dit alors que le couple (A, λ) est une *variété abélienne principalement polarisée*, qu'on abrège en v.a.p.p.

Les variétés abéliennes de dimension 1 ne sont autres que les courbes de genre 1 munies d'un point, qui définit une polarisation principale. En dimension au moins 2, certaines variétés abéliennes n'admettent aucune polarisation principale ; mais toute variété abélienne est isogène à une v.a.p.p.

Étant données deux variétés abéliennes polarisées (A, λ) et (B, μ) , un *morphisme* $f : (A, \lambda) \rightarrow (B, \mu)$ est un homomorphisme $f : A \rightarrow B$ tel que $f^\vee \circ \mu \circ f = \lambda$. Il en résulte que f est fini, et que c'est un isomorphisme lorsque λ et μ ont le même degré. De plus, le groupe des automorphismes $\text{Aut}(A, \lambda)$ est fini et non trivial ; en fait, il contient toujours l'involution $[-1] : a \mapsto -a$.

La classification des v.a.p.p est intimement liée à celles des courbes :

Exemple 1.1. — Soit C une courbe complète et lisse de genre $g := h^1(\mathcal{O}_C) \geq 1$. Soit $J = J(C) := \text{Pic}^0(C)$ sa jacobienne (formée des classes d'équivalence linéaire des diviseurs de degré 0) ; c'est une variété abélienne de dimension g . Le choix d'un point P de C définit un morphisme

$$f : C^{g-1} \rightarrow J, \quad (P_1, \dots, P_{g-1}) \mapsto P_1 + \dots + P_{g-1} - (g-1)P$$

dont l'image est un diviseur irréductible Θ de J ; un autre choix de P fournit un translaté de Θ , et ces diviseurs définissent une polarisation principale θ de J . D'après le théorème de Torelli, la classe d'isomorphie de la courbe C est uniquement déterminée par celle de la v.a.p.p (J, θ) .

1.2. L'espace des modules des variétés abéliennes principalement polarisées

Pour définir précisément cet espace qui paramètre les classes d'isomorphie des v.a.p.p de dimension donnée, on a besoin de quelques notions de nature schématique.

Tous les schémas considérés sont supposés localement noethériens. Un *schéma en groupes* sur un schéma de base S est un S -schéma $\pi : \mathcal{G} \rightarrow S$ muni de S -morphisms $\mu : \mathcal{G} \times_S \mathcal{G} \rightarrow \mathcal{G}$ (la multiplication), $\varepsilon : S \rightarrow \mathcal{G}$ (l'élément neutre) et $\iota : \mathcal{G} \rightarrow \mathcal{G}$ (l'inverse) qui vérifient les axiomes des groupes.

Un *schéma abélien* sur S est un schéma en groupes $\pi : \mathcal{A} \rightarrow S$ propre, lisse et à fibres géométriques connexes (c'est-à-dire $\mathcal{A}_{\bar{s}} := \mathcal{A} \times_S \text{Spec } \kappa(\bar{s})$ est connexe pour tout point s de S , où $\kappa(\bar{s})$ désigne une clôture algébrique du corps résiduel $\kappa(s)$). Chaque fibre géométrique $\mathcal{A}_{\bar{s}}$ est une variété abélienne sur $\kappa(\bar{s})$; on peut voir \mathcal{A} comme une famille de variétés abéliennes paramétrée par la base S . La loi de groupe μ est commutative et uniquement déterminée par la section nulle ε . Lorsque S est le spectre d'un corps K , on dit aussi que \mathcal{A} est une variété abélienne sur K .

Tout schéma abélien $\pi : \mathcal{A} \rightarrow S$ admet un *dual* $\mathcal{A}^\vee = \mathbf{Pic}^0(\mathcal{A}/S)$; c'est un schéma abélien sur S , dont chaque fibre géométrique est la duale de la fibre géométrique correspondante de \mathcal{A} [8, Sec.I.1.9]. Une *polarisation* est un morphisme $\lambda : \mathcal{A} \rightarrow \mathcal{A}^\vee$ de schémas en groupes sur S , qui induit une polarisation $\lambda_{\bar{s}} : \mathcal{A}_{\bar{s}} \rightarrow \mathcal{A}_{\bar{s}}^\vee$ pour tout point géométrique \bar{s} . Le degré de $\lambda_{\bar{s}}$ est constant sur toute composante connexe de S . La polarisation est *principale* si ce degré est 1, c'est-à-dire si λ est un isomorphisme. On dit alors que le couple (\mathcal{A}, λ) est un *schéma abélien principalement polarisé*, abrégé en s.a.p.p.

Parmi les s.a.p.p de dimension relative $g \geq 1$ fixée, il n'existe aucun schéma *universel*, dont tout autre s.a.p.p s'obtient par un unique changement de base ; en effet, comme on l'a vu, les v.a.p.p admettent des automorphismes non triviaux. On dit que les s.a.p.p n'ont pas d'*espace des modules fin.* Cependant, il existe un schéma A_g qui est la meilleure approximation schématique de la base d'un objet universel, et dont les points sur k ne sont autres que les classes d'isomorphie des v.a.p.p.

Plus précisément, considérons le foncteur contravariant \mathcal{A}_g de la catégorie des schémas vers celle des ensembles, qui à tout schéma S associe l'ensemble des classes d'isomorphie (dans un sens évident) des s.a.p.p de dimension relative g sur S . D'après [15, Thm.7.10], ce foncteur admet un *espace des modules grossier*, c'est-à-dire la donnée d'un schéma A_g et d'un morphisme de foncteurs $\varphi : \mathcal{A}_g \rightarrow h_{A_g} := \text{Mor}(-, A_g)$ tels que :

(i) L'application induite $\mathcal{A}_g(\text{Spec } k) \rightarrow h_{A_g}(\text{Spec } k) = A_g(k)$ est bijective pour tout corps algébriquement clos k .

(ii) Pour tout schéma M et tout morphisme de foncteurs $\psi : \mathcal{A}_g \rightarrow h_M$, il existe un unique morphisme de schémas $f : A_g \rightarrow M$ tel que $\psi = f \circ \varphi$.

(On note encore $f : h_{A_g} \rightarrow h_M$ la composition par f .)

La propriété universelle (ii) détermine le schéma A_g à un unique isomorphisme près. D’après [15, Thm.7.10] et [8, Thm.V.2.3], ce schéma est normal, plat et quasi-projectif sur $\text{Spec } \mathbb{Z}$, de dimension relative $g(g + 1)/2$.

De même, étant donné un entier $g \geq 2$, on considère le foncteur contravariant \mathcal{M}_g qui à tout schéma S associe l’ensemble des classes d’isomorphie des courbes de genre g sur S , c’est-à-dire des morphismes $\pi : \mathcal{C} \rightarrow S$ propres, lisses, et dont toute fibre géométrique est une courbe de genre g . Ce foncteur admet aussi un espace des modules grossier qu’on note M_g ; c’est un schéma normal, plat et quasi-projectif sur $\text{Spec } \mathbb{Z}$, de dimension relative $3g - 3$.

À toute courbe $\pi : \mathcal{C} \rightarrow S$ de genre g on associe sa jacobienne relative $\mathcal{J} = \mathcal{J}(\mathcal{C}) := \text{Pic}^0(\mathcal{C}/S)$. C’est un schéma abélien sur S dont chaque fibre géométrique $\mathcal{J}_{\bar{s}}$ est la jacobienne $J(\mathcal{C}_{\bar{s}})$; de plus, \mathcal{J} est projectif sur S , et les isomorphismes $\theta_{\bar{s}} : J(\mathcal{C}_{\bar{s}}) \rightarrow J(\mathcal{C}_{\bar{s}})^\vee$ se globalisent en un isomorphisme $\theta : \mathcal{J} \rightarrow \mathcal{J}^\vee$. Grâce à la propriété universelle de M_g , on obtient donc le morphisme de Torelli $\mathfrak{t} : M_g \rightarrow A_g$ qui est injectif sur les points géométriques (pour tout cela, voir [15]).

1.3. Variétés semi-abéliennes et théorème de réduction semi-stable

Une variété semi-abélienne est un groupe algébrique G , extension d’une variété abélienne A par un tore T (isomorphe à un produit fini \mathbb{G}_m^r de groupes multiplicatifs). Un tel groupe G est connexe et commutatif, et T est son unique sous-tore maximal ; la dimension r de T est appelée le rang de G , et $A = G/T$ est sa partie abélienne.

Rappelons la classification des variétés semi-abéliennes. Soient $\Lambda := \text{Hom}(T, \mathbb{G}_m) \simeq \mathbb{Z}^r$ le groupe des caractères de T , et $\pi : G \rightarrow A$ le quotient par T . On a une décomposition en espaces propres de T

$$(1) \quad \pi_*(\mathcal{O}_G) = \bigoplus_{\lambda \in \Lambda} L_\lambda$$

où chaque L_λ est un fibré en droites sur A . De plus, L_0 est le fibré trivial, et la multiplication de $\pi_*(\mathcal{O}_G)$ définit des isomorphismes $L_\lambda \otimes L_\mu \simeq L_{\lambda+\mu}$ pour tous $\lambda, \mu \in \Lambda$. Enfin, chaque L_λ est algébriquement trivial, car l’action de G par multiplication préserve la décomposition (1). On obtient donc un homomorphisme

$$c : \Lambda \rightarrow A^\vee, \quad \lambda \mapsto [L_\lambda]$$

qui classifie l’extension $1 \rightarrow T \rightarrow G \rightarrow A \rightarrow 0$.

Exemple 1.2. — Soit C une courbe complète et nodale (c’est-à-dire dont les seules singularités sont des points doubles ordinaires). Le groupe $\text{Pic}^0(C)$ (formé des classes d’isomorphie des fibrés en droites sur C dont la restriction à toute composante irréductible est de degré 0) est une variété semi-abélienne de dimension $g := h^1(\mathcal{O}_C)$, le genre arithmétique de C , et de rang $d - n + 1$ où d désigne le nombre des points doubles de C , et n le nombre de ses composantes irréductibles. La partie abélienne de $\text{Pic}^0(C)$ est la jacobienne de la normalisée \tilde{C} .

En effet, soit $f : \tilde{C} \rightarrow C$ la normalisation et soient C_1, \dots, C_n les composantes irréductibles de C ; alors \tilde{C} est la réunion disjointe des normalisées $\tilde{C}_1, \dots, \tilde{C}_n$. On a une suite exacte de faisceaux sur C

$$0 \rightarrow \mathcal{O}_C \rightarrow f_*(\mathcal{O}_{\tilde{C}}) \rightarrow \mathcal{F} \rightarrow 0$$

où \mathcal{F} est un faisceau gratte-ciel de fibre k en chaque point double, et 0 ailleurs. Puisque le morphisme f est fini, on en déduit une suite exacte longue

$$0 \rightarrow H^0(\mathcal{O}_C) = k \rightarrow H^0(\mathcal{O}_{\tilde{C}}) = k^n \rightarrow H^0(\mathcal{F}) = k^d \rightarrow H^1(\mathcal{O}_C) \rightarrow H^1(\mathcal{O}_{\tilde{C}}) \rightarrow 0$$

d'où $g = d - n + 1 + \sum_{i=1}^n g(\tilde{C}_i)$. De même, la suite exacte $0 \rightarrow \mathcal{O}_C^* \rightarrow f_*(\mathcal{O}_{\tilde{C}}^*) \rightarrow \mathcal{F}^* \rightarrow 0$ conduit à une suite exacte longue

$$(2) \quad 1 \rightarrow k^* \rightarrow (k^*)^n \rightarrow (k^*)^d \rightarrow H^1(\mathcal{O}_C^*) = \text{Pic}(C) \rightarrow H^1(\mathcal{O}_{\tilde{C}}^*) = \text{Pic}(\tilde{C}) \rightarrow 0.$$

D'où une suite exacte

$$1 \rightarrow \mathbb{G}_m^{d-n+1} \rightarrow \text{Pic}^0(C) \rightarrow \prod_{i=1}^n J(\tilde{C}_i) \rightarrow 0$$

ce qui démontre nos assertions.

Le tore maximal de $\text{Pic}^0(C)$ se lit aussi sur le *graphe dual* de C . Il s'agit du graphe non orienté, noté Γ ou $\Gamma(C)$, dont les sommets sont les composantes irréductibles de C ; deux sommets distincts C_i, C_j sont joints par autant d'arêtes que le nombre de leurs points (doubles) communs, et un sommet C_i porte autant de boucles que le nombre de points doubles de C_i . Le tore maximal de $\text{Pic}^0(C)$ est canoniquement isomorphe au groupe de cohomologie $H^1(\Gamma, k^*)$; en effet, dans la suite exacte (2), l'application $(k^*)^n \rightarrow (k^*)^d$ s'identifie au cobord $C^0(\Gamma, k^*) \rightarrow C^1(\Gamma, k^*)$. Par suite, le groupe des caractères de ce tore est le groupe d'homologie $H_1(\Gamma, \mathbb{Z})$, et le rang de $\text{Pic}^0(C)$ est le nombre de cycles libres du graphe Γ .

Les variétés semi-abéliennes apparaissent aussi dans les dégénérescences à un paramètre des variétés abéliennes.

Introduisons quelques notations : soit R un anneau de valuation discrète, complet, de corps des fractions K , d'idéal maximal \mathfrak{m} et de corps résiduel k . Le schéma $S := \text{Spec}(R)$ est appelé un *trait* de point générique $\eta := \text{Spec}(K)$ et de point fermé $s := \text{Spec}(k)$. Tout schéma \mathcal{X} sur S définit ainsi une fibre générique $\mathcal{X}_\eta := \mathcal{X} \times_S \eta$ et une fibre spéciale $\mathcal{X}_s := \mathcal{X} \times_S s$. On choisit un générateur z de \mathfrak{m} , qu'on peut voir comme une coordonnée locale sur le germe de courbe (S, s) . Pour tout entier n strictement positif, le quotient $R' := R[z']/(z'^n - z)$ est encore un anneau de valuation discrète complet, d'idéal maximal $\mathfrak{m}' = z'R'$ et de corps résiduel k ; c'est la clôture intégrale de R dans le corps $K' := K[z']/(z'^n - z)$. On dit que $S' := \text{Spec} R'$ est obtenu à partir de S par extension finie ramifiée.

On peut maintenant énoncer le théorème de réduction semi-stable [8, Sec.I.2].

THÉORÈME 1.3. — Soit \mathcal{A}_η un schéma abélien sur η (autrement dit, une variété abélienne sur K). Alors, quitte à faire un changement de base fini ramifié $S' \rightarrow S$, on peut étendre \mathcal{A}_η en un schéma en groupes lisse $\pi : \mathcal{A} \rightarrow S$ dont la fibre spéciale \mathcal{A}_s est une variété semi-abélienne sur k ; une telle extension est unique à isomorphisme près.

L'extension $\pi : \mathcal{A} \rightarrow S$ est un schéma semi-abélien, c'est-à-dire un schéma en groupes séparé, lisse, et dont toutes les fibres géométriques sont des variétés semi-abéliennes (dont le rang varie en général : un schéma abélien n'est pas nécessairement extension d'une variété abélienne par un tore). D'après [8, p. 35], la donnée d'une polarisation λ_η de \mathcal{A}_η induit une polarisation λ_s de la partie abélienne A_s de la fibre spéciale, et si λ_η est principale, alors λ_s l'est aussi.

Le théorème de réduction semi-stable admet un analogue pour les courbes de genre g : cette fois, la fibre spéciale est une *courbe stable de genre g* , c'est-à-dire une courbe complète, nodale, de genre arithmétique g , et dont le groupe des automorphismes est fini.

Plus généralement, une *courbe stable de genre g* sur un schéma S est un morphisme propre et plat $\pi : \mathcal{C} \rightarrow S$ dont toutes les fibres géométriques sont des courbes stables de genre g ; le théorème de réduction semi-stable reste valable [23]. Les courbes stables de genre g admettent un espace des modules grossier \overline{M}_g sur $\text{Spec } \mathbb{Z}$, qui vérifie les critères valuatifs de séparation et de propreté d'après ce même théorème. En fait, \overline{M}_g est projectif sur $\text{Spec } \mathbb{Z}$ (voir [15]) et les courbes stables de genre g sont les objets d'un *champ de Deligne–Mumford propre* qui compactifie le champ des courbes de genre g (pour ces notions, voir [12]).

1.4. Compactifications des espaces des modules des v.a.p.p.

D'après [8, Thm.V.2.3], A_g admet une compactification \overline{A}_g^{\min} dite *minimale* ; c'est un schéma normal, plat et projectif sur $\text{Spec } \mathbb{Z}$, qui contient A_g comme ouvert dense et dont le bord $\overline{A}_g^{\min} \setminus A_g$ est réunion disjointe de g sous-schémas localement fermés, isomorphes respectivement à A_{g-1}, \dots, A_1, A_0 . De plus, l'adhérence de chaque A_a dans \overline{A}_g^{\min} est isomorphe à la compactification minimale \overline{A}_a^{\min} .

Soit $\pi : \mathcal{A} \rightarrow S$ une dégénérescence à un paramètre comme dans le théorème 1.3. D'après la définition de A_g et le critère valuatif de propreté, il existe un unique morphisme $\varphi : S \rightarrow \overline{A}_g^{\min}$ tel que $\varphi(\eta)$ est la classe d'isomorphie de (A_η, λ_η) (vue comme un point de A_g). Et d'après [8, Thm.V.2.3], $\varphi(s)$ est la classe de (A_s, λ_s) , un point géométrique de A_a où a est la dimension de la partie abélienne A_s de la fibre spéciale. Ainsi, la compactification minimale ne rend compte que très partiellement des dégénérescences.

On connaît une famille d'autres compactifications de A_g , les *compactifications toroïdales* [6, 8] qui dépendent de données combinatoires qu'on va préciser. Soit V un espace vectoriel réel de dimension g muni d'un réseau Λ . L'espace vectoriel réel \mathcal{Q} des formes quadratiques sur V est alors muni du réseau Γ des formes *entières*, c'est-à-dire dont la forme bilinéaire symétrique associée est à valeurs entières sur $\Lambda \times \Lambda$. (On

convient que la forme bilinéaire B associée à une forme quadratique Q est donnée par $B(x, y) = Q(x + y) - Q(x) - Q(y)$. On note \mathcal{Q}^+ le cône convexe de \mathcal{Q} engendré par les formes quadratiques entières et positives ; les points de \mathcal{Q}^+ ne sont autres que les formes quadratiques positives sur V à noyau rationnel. Le groupe des automorphismes $\text{Aut}(\Lambda) \simeq \text{GL}_g(\mathbb{Z})$ opère dans \mathcal{Q} par changement de variables ; cette opération préserve Γ et \mathcal{Q}^+ .

Une *subdivision admissible* de \mathcal{Q}^+ est une famille Σ de parties de \mathcal{Q}^+ telles que :

- (i) Chaque $\sigma \in \Sigma$ est un cône convexe polyédral rationnel (pour le réseau Γ), et ces cônes recouvrent \mathcal{Q}^+ .
- (ii) Toute face d'un cône de Σ appartient à Σ .
- (iii) L'intersection de deux cônes de Σ est une face commune de ces cônes.
- (iv) Σ est invariant par l'action de $\text{Aut}(\Lambda)$ et ne contient qu'un nombre fini d'orbites pour cette action.

Une subdivision admissible Σ est dite *lisse* lorsque tous ses cônes maximaux sont engendrés par des bases de Γ ; elle est dite *projective* s'il existe une fonction continue et convexe $h : \mathcal{Q}^+ \rightarrow \mathbb{R}$ telle que :

- (a) $h(Q) > 0$ pour toute forme quadratique Q définie positive.
- (b) La restriction de h à chaque cône maximal de Σ s'étend en une (unique) forme linéaire sur \mathcal{Q} , à valeurs entières sur Γ .
- (c) Les formes linéaires associées à deux cônes maximaux distincts sont distinctes.

Il existe des subdivisions admissibles, et chacune d'elles peut être raffinée en une subdivision admissible, projective et lisse. De plus, deux subdivisions admissibles ont toujours un raffinement commun.

À toute subdivision admissible Σ on associe une compactification $\overline{A}_{g, \mathbb{C}}^\Sigma$ de l'espace analytique complexe $A_{g, \mathbb{C}}$. En général, $\overline{A}_{g, \mathbb{C}}^\Sigma$ n'est pas un schéma sur $\text{Spec } \mathbb{C}$, mais un espace analytique complexe compact qui admet une stratification indexée par les orbites de $\text{Aut}(\Lambda)$ dans Σ . La strate associée au cône nul n'est autre que $A_{g, \mathbb{C}}$, et l'identité de $A_{g, \mathbb{C}}$ s'étend en un morphisme $\overline{A}_{g, \mathbb{C}}^\Sigma \rightarrow \overline{A}_{g, \mathbb{C}}^{\min}$. Pour tout raffinement Σ' de Σ , on a aussi un morphisme $\overline{A}_{g, \mathbb{C}}^{\Sigma'} \rightarrow \overline{A}_{g, \mathbb{C}}^\Sigma$ qui étend l'identité de $A_{g, \mathbb{C}}$.

Lorsque Σ est projective, $\overline{A}_{g, \mathbb{C}}^\Sigma$ s'obtient à partir de $\overline{A}_{g, \mathbb{C}}^{\min}$ en éclatant un certain faisceau d'idéaux («critère de Tai», voir [6, Sec.IV.2]) ; en particulier, $\overline{A}_{g, \mathbb{C}}^\Sigma$ est une variété projective. Le critère de Tai reste valable sur les entiers grâce aux résultats de [8] ; on obtient ainsi un schéma \overline{A}_g^Σ , projectif et plat sur $\text{Spec } \mathbb{Z}$, qui est un modèle entier de $\overline{A}_{g, \mathbb{C}}^\Sigma$. Un tel modèle entier existe aussi lorsque Σ est lisse [8, Thm.IV.5.7] ; c'est un *espace algébrique* propre et plat sur $\text{Spec } \mathbb{Z}$. En fait, \overline{A}_g^Σ est l'espace des modules grossiers d'un champ de Deligne–Mumford propre qui compactifie le champ des s.a.p.p.

Parmi les subdivisions admissibles, on distingue la *deuxième subdivision de Voronoi* dont l'intérieur relatif de chaque cône est l'ensemble des formes quadratiques positives

qui donnent une décomposition de Delaunay fixée de V (la définition de la décomposition de Delaunay est rappelée en 2.1). Cette subdivision, notée Vor , est toujours projective [3, Cor.5.12.8], mais elle n'est lisse qu'en dimension $g \leq 4$ (voir [5, Sec.1.14] et ses références). De plus, le morphisme de Torelli $\mathbf{t} : M_g \rightarrow A_g$ se prolonge en un morphisme

$$\bar{\mathbf{t}}_{\mathbb{C}} : \bar{M}_{g,\mathbb{C}} \rightarrow \bar{A}_{g,\mathbb{C}}^{\text{Vor}}$$

qui est un isomorphisme lorsque $g = 2$, mais qui n'est pas injectif dès que $g \geq 3$ (voir [19] et ses références).

2. CONSTRUCTION DE DÉGÉNÉRESCENCES MAXIMALES DES VARIÉTÉS ABÉLIENNES PRINCIPALEMENT POLARISÉES

Les dégénérescences des variétés abéliennes $\pi : \mathcal{A} \rightarrow S$ obtenues grâce au théorème de réduction semi-stable (1.3) ont l'inconvénient de ne pas être propres en général : il existe de telles dégénérescences, dites *maximales*, dont la fibre spéciale est un tore. Mais on peut obtenir des dégénérescences propres $\pi : \mathcal{X} \rightarrow S$ de la fibre générique \mathcal{A}_η grâce à une construction de Mumford [14], reprise et développée par Faltings et Chai [8], puis par Alexeev et Nakamura [5]. Dans cette partie, on expose l'approche de [5] dans le cas particulier des dégénérescences maximales des v.a.p.p. (le cas général n'en est pas très éloigné, mais nécessite beaucoup plus de notations).

L'idée, due à Tate, est de voir \mathcal{A}_η comme le quotient d'un tore $(K^*)^g$ par un sous-groupe $\Lambda \simeq \mathbb{Z}^g$ de périodes. On construit des compactifications partielles $\tilde{\pi} : \tilde{\mathcal{X}} \rightarrow S$ de $(K^*)^g = \tilde{\mathcal{X}}_\eta$ qui sont munies d'une action propre de Λ prolongeant son action dans $(K^*)^g$ par multiplication, et dont la fibre spéciale $\tilde{\mathcal{X}}_s$ est la réunion des translatés par Λ d'une variété torique sous le tore $(k^*)^g$. Le quotient $\mathcal{X} = \tilde{\mathcal{X}}/\Lambda$ est alors une dégénérescence de \mathcal{A}_η , propre sur S et dont la fibre spéciale admet une action de $(k^*)^g$ avec un nombre fini d'orbites.

2.1. Décompositions de Delaunay et de Voronoi

On commence par présenter des objets de géométrie des nombres qui apparaîtront comme données combinatoires des dégénérescences.

Soit Λ un réseau d'un espace vectoriel réel V de dimension g . Soit Q une forme quadratique définie positive sur V . Étant donné $v \in V$, un point $\lambda \in \Lambda$ est dit à *distance minimale de v* si $Q(v - \lambda) = \min_{\mu \in \Lambda} Q(v - \mu)$. Il existe de tels points, et ils sont en nombre fini ; leur enveloppe convexe dans V est la *cellule de Delaunay* de v , notée $D(v)$. On a $D(v + \lambda) = D(v) + \lambda$ pour tous $v \in V$ et $\lambda \in \Lambda$.

Chaque cellule est un polytope convexe dont l'intersection avec Λ est formée de ses sommets ; les cellules minimales ne sont autres que les points de Λ . De plus, chaque face d'une cellule est une cellule, et l'intersection de deux cellules est une face de chacune d'elles. Enfin, les cellules recouvrent V et ne forment qu'un nombre fini d'orbites pour l'action de Λ par translation. On dit que les cellules de Delaunay forment un *pavage*

périodique de V par des polytopes convexes entiers, la *décomposition de Delaunay* Del_Q . Pour une forme quadratique générale, la décomposition est une triangulation, c'est-à-dire les cellules sont des simplexes.

Pour toute cellule de Delaunay σ , l'ensemble des $v \in V$ tels que $D(v) = \sigma$ est l'intérieur relatif d'un polytope convexe noté σ^\vee ou encore $V(\sigma)$, et appelé la *cellule de Voronoi duale* de σ . Lorsque σ est maximale, σ^\vee est formée d'un point unique : le centre de la sphère circonscrite aux sommets de σ , noté $v(\sigma)$ et appelé le *centre* de σ . En notant B la forme bilinéaire symétrique associée à Q (de sorte que $Q(v) = \frac{1}{2}B(v, v)$), le centre de σ est l'unique solution du système d'équations linéaires

$$(3) \quad B(v(\sigma) - \lambda, \mu - \lambda) = Q(\mu - \lambda)$$

où λ est un sommet fixé de σ , et μ décrit les autres sommets.

Les cellules de Voronoi forment aussi un pavage périodique de V par des polytopes convexes (en général non entiers) ; c'est la *décomposition de Voronoi* Vor_Q . Les cellules de Voronoi sont en bijection décroissante avec celles de Delaunay via $\sigma \mapsto \sigma^\vee$; de plus, $\dim(\sigma) = \text{codim}(\sigma^\vee)$.

Exemple 2.1. — Lorsque $g = 1$, la décomposition de Delaunay de $V \simeq \mathbb{R}$ est formée des intervalles entiers $[n, n + 1]$, et celle de Voronoi, des intervalles $[n - \frac{1}{2}, n + \frac{1}{2}]$.

Lorsque $g = 2$, on obtient deux décompositions de Delaunay de $V \simeq \mathbb{R}^2$, associées aux réseaux carré et hexagonal. Dans le premier cas, les cellules de Delaunay maximales sont le carré unité et ses translatés par $\Lambda \simeq \mathbb{Z}^2$; les cellules de Voronoi maximales sont les translatés de ces carrés par le vecteur $(\frac{1}{2}, \frac{1}{2})$.

Dans le deuxième cas, les cellules de Delaunay maximales sont des triangles équilatéraux ; elles forment deux orbites sous Λ . Les cellules de Voronoi maximales sont des hexagones réguliers.

Considérons maintenant l'application quadratique

$$(4) \quad F : V \rightarrow \mathbb{R}, \quad v \mapsto Q(v) + L(v)$$

où L est une forme linéaire sur V . Soit P l'enveloppe convexe dans $V \times \mathbb{R}$ des points $(\lambda, F(\lambda))$ où $\lambda \in \Lambda$. On obtient alors facilement :

LEMME 2.2. — *Pour tout $\lambda \in \Lambda$, le cône tangent à P en le point $(\lambda, F(\lambda))$ (c'est-à-dire le cône convexe de $V \times \mathbb{R}$ engendré par $-(\lambda, F(\lambda)) + P$) est l'ensemble des $(v, t) \in V \times \mathbb{R}$ tels que*

$$t \geq B(v(\sigma), v) + L(v) = dF_{v(\sigma)}(v)$$

où σ parcourt les cellules de Delaunay maximales contenant λ . De plus, l'intersection de P avec l'hyperplan d'équation $t - F(\lambda) = dF_{v(\sigma)}(v - \lambda)$ est l'enveloppe convexe des $(\mu, F(\mu))$ où μ décrit les sommets de σ .

Il en résulte que chaque $(\lambda, F(\lambda))$ est un sommet de P ; plus généralement, le bord de P admet un pavage par des polytopes convexes $P(\sigma)$ dont les projections sur V ne sont autres que les cellules de Delaunay σ . De plus, l'éventail normal de P (formé des cônes duaux aux cônes tangents aux sommets, ainsi que des faces de ces cônes duaux) est formé de l'origine et des cônes engendrés par les images des cellules de Voronoi via l'application affine injective

$$(-dF, 1) : V \rightarrow V \times \mathbb{R}, \quad v \mapsto (-dF_v, 1).$$

Pour tous $\lambda \in \Lambda$ et $v \in V$, on pose

$$h(\lambda; v) := \max_{\sigma \ni \lambda} dF_{v(\sigma)}(v)$$

où σ décrit les cellules de Delaunay maximales qui contiennent λ . La fonction $h(\lambda; -)$ est linéaire sur chaque cône engendré par $-\lambda + \sigma$, ou encore par les vecteurs $\mu - \lambda$ où μ décrit les sommets de σ . Ces vecteurs sont appelés les *vecteurs de Delaunay* de la cellule maximale σ en son sommet λ .

Comme les vecteurs de Delaunay appartiennent à Λ et engendrent l'espace vectoriel V , il existe un entier positif n tel que le réseau qu'ils engendrent contient $n\Lambda$. Le plus petit tel entier n est appelé l'*indice de nilpotence* de σ en λ . Quand σ et λ varient, les indices de nilpotence sont en nombre fini, et leur ppcm est appelé l'indice de nilpotence de la décomposition de Delaunay. Cet indice vaut toujours 1 en dimension $g \leq 4$, car les vecteurs de Delaunay engendrent alors Λ . Mais ceci ne s'étend pas aux dimensions $g \geq 5$; voir [5, Sec.1.14,1.15].

2.2. Construction de compactifications partielles

On conserve les notations de 2.1 et on suppose que F est à valeurs entières sur Λ , si bien que Q et L sont rationnelles. Les polytopes $P(\sigma)$ sont alors entiers, et les centres $v(\sigma)$ sont rationnels d'après (3) ; par suite, l'éventail normal à P est rationnel. On va associer à cet éventail une compactification partielle du tore $\text{Hom}(\Lambda, K^*) \simeq (K^*)^g$, avec les notations de 1.3.

Soit $R[\Lambda]$ l'algèbre du groupe Λ sur R , c'est-à-dire le R -module libre sur les e^λ , $\lambda \in \Lambda$, muni du produit défini par $e^\lambda e^\mu = e^{\lambda+\mu}$. Dans la R -algèbre graduée

$$R[\Lambda][\theta] = \bigoplus_{\lambda \in \Lambda, n \in \mathbb{N}} R e^\lambda \theta^n$$

où θ est une indéterminée de degré 1, on considère la sous- R -algèbre \mathcal{S} engendrée par les monômes

$$\zeta_\lambda := z^{F(\lambda)} e^\lambda \theta \quad (\lambda \in \Lambda)$$

ainsi que le sous- R -module \mathcal{R} engendré par les monômes $z^m e^\lambda \theta^n$ tels que $(\lambda, m) \in nP$. Alors \mathcal{R} est une sous- R -algèbre graduée de $R[\Lambda][\theta]$, entière sur sa sous-algèbre graduée \mathcal{S} . De plus, \mathcal{S} est engendrée par ses éléments de degré 1, et $\mathcal{R}_0 = \mathcal{S}_0 = R$. On obtient donc un schéma

$$\tilde{\mathcal{X}} := \text{Proj}(\mathcal{R})$$

sur S , muni d'un fibré en droites

$$\tilde{\mathcal{L}} := \mathcal{O}_{\tilde{\mathcal{X}}}(1).$$

Soit enfin

$$T := \text{Spec } R[\Lambda]$$

le tore scindé sur S dont le groupe des caractères est Λ ; ce tore opère dans le schéma $\tilde{\mathcal{X}}$, et le fibré en $\tilde{\mathcal{L}}$ est T -linéarisé (c'est-à-dire l'action de T dans $\tilde{\mathcal{X}}$ se relève en une action dans l'espace total de $\tilde{\mathcal{L}}$, linéaire dans les fibres).

Des propriétés des décompositions de Delaunay et de Voronoi énoncées ci-dessus, on déduit facilement :

PROPOSITION 2.3. — (i) *Le schéma $\tilde{\mathcal{X}}$ est recouvert par les ouverts affines $\tilde{\mathcal{X}}_\lambda := \text{Spec } \mathcal{R}_\lambda$, où $\lambda \in \Lambda$ et $\mathcal{R}_\lambda := \tilde{\mathcal{R}}[\frac{1}{\zeta_\lambda}]_0 = \bigcup_{n \in \mathbb{N}} \tilde{\mathcal{R}}_n \zeta_\lambda^{-n}$ est la sous- R -algèbre de $R[\Lambda]$ engendrée par les monômes $z^m e^\mu$ tels que $m \geq h(\lambda; \mu)$. Les monômes*

$$\zeta_{\lambda; \mu} = z^{\lceil h(\lambda; \mu) \rceil} e^\mu$$

(où $\lceil x \rceil$ désigne le plus petit entier $\leq x$) forment une base du R -module \mathcal{R}_λ .

(ii) *Toutes les R -algèbres \mathcal{R}_λ , $\lambda \in \Lambda$, sont isomorphes et de type fini sur S . En particulier, $\tilde{\mathcal{X}}$ est plat et localement de type fini sur S .*

(iii) *Le fibré en droites $\tilde{\mathcal{L}}$ est ample sur $\tilde{\mathcal{X}}$ (c'est-à-dire les ouverts associés aux sections des puissances positives $\tilde{\mathcal{L}}^n$ forment une base de la topologie de $\tilde{\mathcal{X}}$).*

(iv) *La fibre générique $\tilde{\mathcal{X}}_\eta$ n'est autre que le tore $\text{Hom}(\Lambda, K^*) = T_\eta$.*

(v) *La fibre spéciale $\tilde{\mathcal{X}}_s$ est un schéma localement de type fini sur k , muni d'une action du tore $T_s = \text{Hom}(\Lambda, k^*)$ dont les orbites sont en bijection avec les cellules de Delaunay. Cette bijection $\sigma \mapsto \mathcal{O}_\sigma$ vérifie $\dim \mathcal{O}_\sigma = \dim \sigma$ et $\overline{\mathcal{O}_\sigma} \cap \overline{\mathcal{O}_\tau} = \overline{\mathcal{O}_{\sigma \cap \tau}}$. En particulier, $\tilde{\mathcal{X}}_s$ est connexe.*

(vi) *Les composantes irréductibles de $\tilde{\mathcal{X}}_s$ sont les adhérences $\overline{\mathcal{O}_\sigma}$ où σ décrit les cellules de Delaunay maximales. La multiplicité du schéma $\tilde{\mathcal{X}}_s$ le long de $\overline{\mathcal{O}_\sigma}$ est le dénominateur de $dF_{v(\sigma)}$ (vu comme un point de l'espace vectoriel dual V^* , rationnel par rapport au réseau dual Λ^*).*

En particulier, $\tilde{\mathcal{X}}_s$ est génériquement réduite lorsque chaque $dF_{v(\sigma)}$ est entier, autrement dit, lorsque chaque $B(v(\sigma), -)$ est à valeurs entières sur Λ . D'après [5, Lem.3.12], cette condition est vérifiée si F est à valeurs dans $n\mathbb{Z}$ où n désigne l'indice de nilpotence de la décomposition de Delaunay. Le changement de base $S' \rightarrow S$, $z = z'^n$ permet de remplacer F par nF ; quitte à effectuer ce changement de base, on suppose désormais que tous les $dF_{v(\sigma)}$ sont entiers. Alors chaque fonction $h(\lambda; -)$ est à valeurs entières sur Λ . De plus, $\tilde{\mathcal{X}}_s$ est recouvert par les ouverts affines $\tilde{\mathcal{X}}_{\lambda, s} := \text{Spec } \bar{\mathcal{R}}_\lambda$ où le k -espace vectoriel $\bar{\mathcal{R}}_\lambda := \mathcal{R}_\lambda / \mathfrak{m} \mathcal{R}_\lambda$ a une base formée des monômes $\bar{\zeta}_{\lambda; \mu}$ (les images des $\zeta_{\lambda; \mu} = z^{h(\lambda; \mu)} e^\mu$). La multiplication dans $\bar{\mathcal{R}}_\lambda$ est donnée par

$$\bar{\zeta}_{\lambda; \mu_1} \cdots \bar{\zeta}_{\lambda; \mu_n} = \bar{\zeta}_{\lambda; \mu_1 + \cdots + \mu_n}$$

si μ_1, \dots, μ_n appartiennent au cône tangent en 0 à une même cellule de Delaunay $\sigma \ni 0$; sinon, le produit est nul.

Il en résulte aussitôt que $\bar{\mathcal{R}}_\lambda$, et donc $\tilde{\mathcal{X}}_s$, est réduite. En fait, on a un résultat un peu plus précis [3, Sec.2.3] :

PROPOSITION 2.4. — *La fibre spéciale $\tilde{\mathcal{X}}_s$ est semi-normale.*

Un schéma réduit X est dit *semi-normal* si, pour tout schéma réduit Y , tout morphisme fini et bijectif $\pi : Y \rightarrow X$ qui induit des isomorphismes sur les corps résiduels $\kappa(\pi(y)) \subseteq \kappa(y)$ est un isomorphisme. Par exemple, les courbes nodales sont semi-normales, mais non la courbe plane cuspidale d'équation homogène $y^2z = x^3$.

2.3. Action du réseau et passage au quotient

On va définir une action de Λ dans le schéma $\tilde{\mathcal{X}}$, compatible à l'action de T . Pour cela, on se donne un homomorphisme de groupes $\Lambda \rightarrow T_\eta = \text{Hom}(\Lambda, K^*)$, ou encore une application

$$b : \Lambda \times \Lambda \rightarrow K^*$$

telle que $b(\lambda + \mu, \nu) = b(\lambda, \nu) b(\mu, \nu)$ et $b(\lambda, \mu + \nu) = b(\lambda, \mu) b(\lambda, \nu)$ pour tous $\lambda, \mu, \nu \in \Lambda$; on dit que b est *bi-multiplicative*. Ceci définit une action de Λ dans la K -algèbre $K[\Lambda]$ via

$$\lambda \cdot e^\mu = b(\lambda, \mu) e^\mu.$$

Pour l'étendre en une action de Λ dans la K -algèbre graduée $K[\Lambda][[\theta]]$, posons

$$\lambda \cdot \theta := f(\lambda) e^\lambda \theta$$

pour tout $\lambda \in \Lambda$, où $f(\lambda) \in K^*$. Ceci définit bien une action pourvu que l'application $f : \Lambda \rightarrow K^*$ soit *quadratique multiplicative* et que b soit l'application bi-multiplicative associée, c'est-à-dire

$$f(\lambda + \mu) = f(\lambda) f(\mu) b(\lambda, \mu) \quad \text{pour tous } \lambda, \mu \in \Lambda.$$

En particulier, b est alors symétrique. Pour que cette action laisse stables les sous-algèbres \mathcal{S} et \mathcal{R} , il faut et il suffit que

$$f(\lambda) = u(\lambda) z^{F(\lambda)} \quad \text{où } u(\lambda) \in R^*.$$

Alors $b(\lambda, \mu) = u(\lambda + \mu) u(\lambda)^{-1} u(\mu)^{-1} z^{B(\lambda, \mu)}$. Le fait que la forme quadratique Q est définie positive équivaut donc à

$$(5) \quad b(\lambda, \lambda) \in \mathfrak{m} \quad \text{pour tout } \lambda \in \Lambda \setminus \{0\}.$$

On obtient ainsi une action de Λ dans $\tilde{\mathcal{X}}$ qui commute à l'action de T et qui se relève en une action dans $\tilde{\mathcal{L}}$. Pour cette dernière action, on a la relation de commutation $t \circ \lambda = \lambda(t) \lambda \circ t$. En d'autres termes, l'action de $T \times \Lambda$ sur $\tilde{\mathcal{X}}$ se relève en une action du groupe de Heisenberg $(\mathbb{G}_{m, S} \times_S T) \times \Lambda$ sur $\tilde{\mathcal{L}}$. Chaque $\lambda \in \Lambda$ envoie \mathcal{R}_μ sur $\mathcal{R}_{\lambda + \mu}$, et donc l'ouvert $\tilde{\mathcal{X}}_\mu$ sur $\tilde{\mathcal{X}}_{\mu - \lambda}$. De plus, l'action induite de λ dans $\tilde{\mathcal{X}}_s$ envoie chaque orbite \mathcal{O}_σ sur $\mathcal{O}_{-\lambda + \sigma}$. On en déduit :

PROPOSITION 2.5. — (i) Toute orbite de Λ dans $\tilde{\mathcal{X}}_s$ rencontre la réunion des \mathcal{O}_σ où σ décrit les cellules de Delaunay maximales qui contiennent 0.

(ii) Le groupe Λ opère proprement dans $\tilde{\mathcal{X}}_s$.

(iii) Le quotient $\mathcal{X}_s := \tilde{\mathcal{X}}_s/\Lambda$ est un schéma projectif sur k , et le fibré en droites $\tilde{\mathcal{L}}_s$ descend en un fibré en droites ample \mathcal{L}_s sur \mathcal{X}_s .

En fait, d’après les propositions 2.3 et 2.4, \mathcal{X}_s est une variété semi-normale dans laquelle T_s opère avec un nombre fini d’orbites, indexées par les orbites de Λ dans les cellules de Delaunay. En particulier, \mathcal{X}_s contient un unique point fixe de T_s .

Plus généralement, on considère pour tout entier $n \geq 1$ le point épais

$$S_n := \text{Spec } R/\mathfrak{m}^n$$

et la fibre spéciale épaissie

$$\tilde{\mathcal{X}}_n := \tilde{\mathcal{X}} \times_S S_n$$

munie du fibré en droites

$$\tilde{\mathcal{L}}_n := \tilde{\mathcal{L}} \times_S S_n.$$

Les assertions (ii) et (iii) ci-dessus s’étendent à ces fibres [14, Thm.3.10] ; on obtient ainsi un système inductif $(\mathcal{X}_n, \mathcal{L}_n)$ qui définit un schéma formel \mathfrak{X} sur le spectre formel de R , muni d’un fibré en droites ample \mathfrak{L} . Le couple $(\mathfrak{X}, \mathfrak{L})$ est donc algébrisable en un unique schéma \mathcal{X} , propre et plat sur S , muni d’un fibré en droites ample \mathcal{L} . De plus, \mathcal{X}_η est une variété abélienne d’après [14, Cor.4.9].

On va construire une section globale non nulle de \mathcal{L} . Pour cela, on va algébriser la série formelle

$$\tilde{\theta} := \sum_{\lambda \in \Lambda} \lambda \cdot \theta$$

vue comme une section globale de \mathfrak{L} . Posons

$$\xi_\lambda := \lambda \cdot \theta = f(\lambda) e^\lambda \theta \quad \text{et} \quad \xi_{\lambda;\mu} := \frac{\xi_{\lambda+\mu}}{\xi_\lambda} \quad (\lambda, \mu \in \Lambda)$$

si bien que $\xi_\lambda \in R^* \zeta_\lambda$ et $\xi_{\lambda;\mu} \in \mathcal{R}_\lambda$ pour tout $\mu \in \Lambda$. On vérifie alors :

LEMME 2.6. — (i) Pour tout entier $n \geq 1$, les $\mu \in \Lambda$ tels que $\xi_{\lambda;\mu} \notin \mathfrak{m}^n \mathcal{R}_\lambda$ sont en nombre fini.

(ii) $\xi_{\lambda;\mu} \notin \mathfrak{m} \mathcal{R}_\lambda$ si et seulement si μ est un vecteur de Delaunay en 0 (autrement dit, μ appartient à une cellule de Delaunay contenant 0).

Le quotient

$$\theta_\lambda := \frac{\tilde{\theta}}{\xi_\lambda} = \sum_{\mu \in \Lambda} \xi_{\lambda;\mu}$$

donne donc une somme finie dans chaque $\mathcal{R}_\lambda \otimes_R R/\mathfrak{m}^n$, et la famille des θ_λ définit une section Λ -invariante de $\tilde{\mathcal{L}}_n$ qui s’algébrise comme précédemment en la section cherchée, notée θ . De plus, θ_s ne s’annule identiquement sur aucune orbite de T_s , car l’image

de θ_λ dans $\bar{\mathcal{R}}_\lambda = \mathcal{R}_\lambda/\mathfrak{m}\mathcal{R}_\lambda$ ne s'annule pas en l'unique point fixe de T_s dans $\tilde{\mathcal{X}}_{\lambda,s}$. Enfin, θ engendre $\pi_*(\mathcal{L})$; il en résulte que \mathcal{L}_η définit une polarisation principale de \mathcal{X}_η .

En conclusion, on a obtenu :

THÉORÈME 2.7. — *À la donnée d'une application quadratique multiplicative $f : \Lambda \rightarrow K^*$ telle que l'application bi-multiplicative associée vérifie (5), on associe un schéma \mathcal{X} propre et plat sur S , de dimension relative g , et un diviseur de Cartier effectif et ample \mathcal{D} sur \mathcal{X} tels que :*

(i) *La fibre générique \mathcal{X}_η est une variété abélienne dont \mathcal{D}_η définit une polarisation principale.*

(ii) *La fibre spéciale \mathcal{X}_s est une variété projective et semi-normale dans laquelle le tore $T_s = \text{Hom}(\Lambda, k^*)$ opère avec un nombre fini d'orbites et un unique point fixe. Le diviseur \mathcal{D}_s ne contient aucune de ces orbites.*

Mais on notera que \mathcal{X} n'est pas toujours une compactification de la dégénérescence semi-stable de \mathcal{X}_η donnée par le théorème 1.3. En effet, l'adhérence dans \mathcal{X} de la section nulle de \mathcal{X}_η rencontre la fibre spéciale en un point dont l'orbite n'est pas nécessairement de dimension maximale [5, Sec.3.25]. Il faut plutôt voir \mathcal{X} comme une compactification d'un *torseur* (ou espace principal homogène) sous une variété abélienne ; ce point de vue sera développé dans la troisième partie.

Exemple 2.8. — Comme dans l'exemple 2.1, considérons les cas où $g \leq 2$.

Lorsque $\Lambda = \mathbb{Z}$, la fibre spéciale $\tilde{\mathcal{X}}_s$ est une chaîne infinie de courbes C_n , $n \in \mathbb{Z}$, où chaque C_n est isomorphe à la droite projective \mathbb{P}^1 , et le point ∞ de C_n est identifié au point 0 de C_{n+1} . Ainsi, C_n rencontre transversalement C_{n-1} et C_{n+1} , et ne rencontre aucun autre C_m . Le groupe $\Lambda \simeq \mathbb{Z}$ opère dans $\tilde{\mathcal{X}}_s$ par translations : son générateur 1 envoie isomorphiquement chaque $(C_n, 0, \infty)$ sur $(C_{n+1}, 0, \infty)$. Le quotient $\mathcal{X}_s = \tilde{\mathcal{X}}_s/\Lambda$ est une courbe nodale obtenue à partir de \mathbb{P}^1 en identifiant les points 0 et ∞ , et le diviseur \mathcal{D}_s est un point distinct du point double. Ainsi, $(\pi : \mathcal{X} \rightarrow S, \mathcal{D})$ réalise la dégénérescence d'une courbe de genre 1 munie d'un point, en une courbe rationnelle nodale munie d'un point lisse.

Lorsque $\Lambda \simeq \mathbb{Z}^2$ est le réseau carré, \mathcal{X}_s s'obtient à partir de $\mathbb{P}^1 \times \mathbb{P}^1$ en identifiant chaque point $(x, 0)$ avec (tx, ∞) , et chaque point $(0, y)$ avec (∞, ty) où t est un paramètre non nul. Le diviseur \mathcal{D}_s est l'image d'une section du fibré $\mathcal{O}(1, 1)$; c'est une courbe rationnelle avec deux points doubles, ou la réunion de deux courbes rationnelles avec un point double chacune, qui se coupent transversalement en un point.

Enfin, lorsque $\Lambda \simeq \mathbb{Z}^2$ est le réseau hexagonal, \mathcal{X}_s s'obtient à partir de la réunion disjointe de deux plans projectifs (associés aux deux types de triangles) en identifiant deux à deux les droites de coordonnées. Le diviseur \mathcal{D}_s est l'image de la réunion disjointe de deux droites en position générale ; il est formé de deux courbes rationnelles lisses qui se coupent transversalement en trois points.

3. COMPACTIFICATION MODULAIRE DE A_g ET DU MORPHISME DE TORELLI

Dans cette partie, on présente une partie des résultats de [2, 3, 4]. On commence par définir les couples quasi-abéliens stables, dont on donne plusieurs classes d'exemples. On décrit la structure de ces couples et on leur associe un invariant de nature combinatoire : le type, sous une hypothèse supplémentaire de linéarisation d'un fibré en droites. Puis on explique comment lever cette hypothèse et définir le type en toute généralité. Enfin, on énonce les résultats principaux concernant la compactification modulaire $\overline{A}_g^{\text{mod}}$ qui paramètre les couples quasi-abéliens stables de type dit périodique de degré 1, et le morphisme de Torelli compactifié qui à chaque courbe stable associe un couple quasi-abélien stable de ce type.

3.1. Variétés et couples quasi-abéliens stables

DÉFINITION 3.1. — *Une variété quasi-abélienne stable est une variété X munie d'une action d'une variété semi-abélienne G telle que :*

- (i) X ne contient qu'un nombre fini d'orbites de G .
- (ii) Pour tout point $x \in X$, le stabilisateur G_x est un tore (en particulier, le schéma G_x est réduit).
- (iii) X est équidimensionnelle de dimension $g = \dim G$.
- (iv) X est semi-normale.

Lorsque X est normale (et donc intègre), on dit que c'est une variété quasi-abélienne.

On utilisera aussi les abréviations v.q.-a.s et v.q.-a. (Bien avant que cette définition ne soit formulée dans [3], Nakamura [16] et Namikawa [18] avaient introduit une notion voisine, mais différente, de variété quasi-abélienne stable ; pour eux, il s'agit des fibres, éventuellement non réduites, d'une certaine famille sur $\overline{A}_{g,\mathbb{C}}^{\text{Vor}}$.)

Les conditions (ii) et (iii) entraînent que chaque composante irréductible d'une v.q.-a.s est une compactification équivariante de la variété semi-abélienne associée. En particulier, les v.q.-a.s sous une variété abélienne A ne sont autres que les toseurs sous A . Et les v.q.-a sous un tore T ne sont autres que les *variétés toriques*, c'est-à-dire les variétés normales dans lesquelles T opère avec une orbite ouverte dont le stabilisateur est trivial (les conditions (i) et (ii) sont alors bien connues).

Les v.q.-a.s sous un tore sont appelées *variétés toriques stables*. Des exemples de telles variétés sont les fibres spéciales des dégénérescences maximales construites dans la deuxième partie.

DÉFINITION 3.2. — *Un couple quasi-abélien stable est formé d'une v.q.-a.s projective X (sous une variété semi-abélienne G) et d'un diviseur de Cartier D sur X , effectif, ample et ne contenant aucune orbite de G .*

La variété X est dite polarisée par le fibré en droites $L := \mathcal{O}_X(D)$. On note s la section canonique de L dont le diviseur des zéros est D , si bien que la donnée du couple (X, D) est équivalente à celle du triplet (X, L, s) . Le degré de ce couple est $h^0(L)$.

Exemple 3.3. — En particulier, on a la notion de *couple abélien* (X, D) où X est un torseur sous une variété abélienne A , et D est un diviseur effectif et ample sur X .

Les couples abéliens de degré 1 sont en correspondance biunivoque avec les variétés abéliennes principalement polarisées. En effet, à toute v.a.p.p (A, λ) on associe la variété $X := \text{Pic}^\lambda(A)$ (vue comme un ensemble de diviseurs de X) et sa sous-variété D formée des diviseurs Θ qui contiennent 0. Alors X est un torseur sous A via l’isomorphisme $\lambda : A \rightarrow A^\vee$ et l’action de A^\vee par translations, et D s’identifie à un diviseur Θ . Réciproquement, à tout couple abélien (X, D) de degré 1, on associe la variété abélienne A sous-jacente à X , munie de la polarisation définie par les translatés de D .

Exemple 3.4. — Toute courbe complète et lisse C de genre g définit un couple abélien de degré 1, formé de la variété $\text{Pic}^{g-1}(C)$ des classes de diviseurs de degré $g - 1$ (un torseur sous $\text{Pic}^0(C)$) et de la sous-variété des classes des diviseurs effectifs.

Plus généralement, à toute courbe complète et nodale C de genre arithmétique g , on associe d’abord la variété semi-abélienne $\text{Pic}^0(C)$ (décrite dans l’exemple 1.2), puis la *jacobiennne compactifiée* $\text{Jac}^{g-1}(C)$; c’est l’espace des modules des *faisceaux semi-stables de degré $g - 1$ sur C* , c’est-à-dire des faisceaux cohérents \mathcal{F} sur C qui sont de rang 1 en chaque point générique et qui vérifient $h^0(\mathcal{F}) = h^1(\mathcal{F})$ et $h^0(\mathcal{G}) \leq h^1(\mathcal{G})$ pour tout sous-faisceau \mathcal{G} (en particulier, un faisceau semi-stable ne contient aucun faisceau gratte-ciel). Le groupe $\text{Pic}^0(C)$ opère dans $\text{Jac}^{g-1}(C)$ par produit tensoriel. Enfin, on note Θ le sous-schéma réduit de $\text{Jac}^{g-1}(C)$ formé des classes des faisceaux \mathcal{F} tels que $h^0(\mathcal{F}) \neq 0$. D’après [20] et [4], le couple $(\text{Jac}^{g-1}(C), \Theta)$ est quasi-abélien stable sous $\text{Pic}^0(C)$; lorsque C est une courbe stable, le degré de ce couple est 1.

On renvoie à [20] pour l’étude détaillée d’une famille de jacobiennes compactifiées $\text{Jac}^\phi(C)$ qui contient $\text{Jac}^{g-1}(C)$; chaque $\text{Jac}^\phi(C)$ est munie d’une action naturelle de $\text{Pic}^0(C)$ qui en fait une v.q.-a.s [4, Thm.5.1].

DÉFINITION 3.5. — La v.q.-a.s polarisée (X, L) sous G est dite *linéarisée* si L est munie d’une *linéarisation* pour le tore maximal T de G .

On rappelle qu’une linéarisation de L est la donnée d’une action de T dans l’espace total de ce fibré en droites, qui relève l’action de T dans X , et qui est linéaire dans les fibres. Puisque X est une variété complète, les linéarisations du fibré trivial s’identifient au groupe des caractères de T ; ce groupe opère donc simplement transitivement dans l’ensemble des linéarisations de L . Rappelons aussi que tout fibré en droites sur une variété normale est linéarisable ; en particulier, toute v.q.-a polarisée est linéarisable.

Exemple 3.6. — Les couples $(\mathcal{X}_s, \mathcal{D}_s)$ du théorème 2.7 sont des couples quasi-abéliens stables ; aucun d’eux n’est linéarisable.

En effet, lorsque X est une variété projective munie d'une action d'un tore T et d'un fibré en droites L ample et T -linéarisé, les sections d'une grande puissance L^n donnent une immersion T -équivariante $i : X \rightarrow \mathbb{P}(V)$ où V est un T -module (rationnel, de dimension finie), et $\mathbb{P}(V)$ désigne son projectivisé. Il en résulte que X est recouvert par des ouverts affines et invariants par T ; en particulier, toute courbe de X qui est invariante par T contient au moins deux points fixes. Mais on a vu que \mathcal{X}_s contient un seul point fixe de T_s .

Exemple 3.7. — Rappelons la classification des variétés toriques linéarisées (X, L) sous un tore T : en notant Λ le groupe des caractères de T , et V l'espace vectoriel réel associé à Λ , on a une correspondance biunivoque entre ces variétés et les polytopes convexes P dans V , entiers par rapport au réseau Λ , et d'intérieur non vide.

La classification des couples toriques s'en déduit aisément : pour tout entier $n \geq 0$, le T -module $H^0(X, L^n)$ est somme directe de droites propres dont les poids ne sont autres que les $\lambda \in \Lambda \cap nP$. En particulier, tout $s \in H^0(X, L)$ se décompose en somme de vecteurs propres s_λ . Pour que le diviseur des zéros D de s définisse un couple torique, il faut et il suffit que $s_\lambda \neq 0$ pour chaque sommet λ de P ; en effet, cette condition signifie que s ne s'annule en aucun point fixe de T dans X .

Enfin, les variétés toriques polarisées correspondent aux classes des polytopes convexes entiers modulo les translations entières.

DÉFINITION 3.8. — *Un morphisme du couple quasi-abélien stable (X, D) sous G , vers le couple quasi-abélien stable (Y, E) sous H , est la donnée d'un homomorphisme de groupes $f : G \rightarrow H$ et d'un morphisme $\varphi : X \rightarrow Y$ tels que φ est G -équivariant (pour l'action de G dans Y via f), et $\varphi^*(E) = D$.*

Pour un tel couple (f, φ) , le morphisme φ est fini, car D et E sont amples ; lorsque X et Y ont la même dimension, f est une isogénie. On montre que le groupe des automorphismes de tout couple quasi-abélien stable est fini.

Voici des versions schématiques des définitions précédentes :

DÉFINITION 3.9. — *Un schéma quasi-abélien stable sur un schéma S est la donnée d'un couple $(\mathcal{G}, \mathcal{X})$ où*

- (i) \mathcal{G} est un schéma en groupes semi-abélien sur S .
- (ii) \mathcal{X} est un schéma plat, séparé et de type fini sur S , muni d'une action de \mathcal{G} .
- (iii) Chaque fibre géométrique $\mathcal{X}_{\bar{s}}$ est une v.q.-a.s sous $\mathcal{G}_{\bar{s}}$.

Lorsque le morphisme $\pi : \mathcal{X} \rightarrow S$ est propre et que \mathcal{X} est muni d'un diviseur de Cartier effectif et relativement ample \mathcal{D} tel que chaque $(\mathcal{X}_{\bar{s}}, \mathcal{D}_{\bar{s}})$ est un couple quasi-abélien stable, on dit que $(\mathcal{X}, \mathcal{D})$ est un couple quasi-abélien stable sur S . Un tel couple est polarisé par le fibré en droites $\mathcal{L} := \mathcal{O}_{\mathcal{X}}(\mathcal{D})$; il est dit de degré d si $h^0(\mathcal{L}_{\bar{s}}) = d$ pour tout point s .

Exemple 3.10. — Les couples $(\mathcal{X}, \mathcal{D}) \rightarrow S$ du théorème 2.7 sont quasi-abéliens stables de degré 1.

D'autres exemples sont issus des courbes stables $\pi : \mathcal{C} \rightarrow S$ de genre g : d'après les résultats de [7, Chap.VIII], il existe en effet un *schéma de Picard relatif* $\mathbf{Pic}^0(\mathcal{C}/S)$ qui est un schéma en groupes semi-abélien ayant pour fibres géométriques les $\mathbf{Pic}^0(\mathcal{C}_{\bar{s}})$. Les jacobiniennes compactifiées des fibres géométriques se globalisent aussi en une *jacobienne compactifiée relative* $\mathbf{Jac}^{g-1}(\mathcal{C}/S)$; elle est munie d'une action du schéma en groupes $\mathbf{Pic}^0(\mathcal{C}/S)$ et d'un diviseur Θ qui définissent un couple quasi-abélien stable de degré 1 [4, Sec.5].

Pour tout couple quasi-abélien stable $(\mathcal{X}, \mathcal{D})$ sur S , le diviseur $\mathcal{D}_{\bar{s}}$ ne contient aucune composante irréductible de $\mathcal{X}_{\bar{s}}$. Il en résulte que \mathcal{D} est plat sur S ; autrement dit, \mathcal{D} est un *diviseur de Cartier relatif*.

On a une notion évidente de morphisme entre couples quasi-abéliens stables ; de plus, tout morphisme entre couples abéliens de même degré sur un même schéma est un isomorphisme. L'exemple 3.3 admet une version schématique [3, Cor.3.0.7] :

THÉORÈME 3.11. — *Les couples abéliens de degré 1 forment un champ isomorphe au champ des schémas abéliens principalement polarisés.*

3.2. Variétés quasi-abéliennes stables linéarisées

On se donne une variété semi-abélienne G de tore maximal T et de partie abélienne A . On commence par présenter une construction de v.q.-a.s sous G qui s'avéreront être des modèles locaux des v.q.-a.s linéarisées.

Soit Y une variété torique stable sous T . Ce dernier opère dans $G \times Y$ via $t \cdot (g, y) = (gt^{-1}, t \cdot y)$ et cette action commute à celle de G par multiplication sur le premier facteur. Notons

$$X = G \times^T Y$$

le quotient ; la projection $G \times Y \rightarrow G$ passe au quotient en un morphisme $p : X \rightarrow A$ qui est une fibration localement triviale de fibre Y . Il en résulte que X est une v.q.-a.s sous G , qu'on appelle l'*induite* de Y .

PROPOSITION 3.12. — *Soit (X, L) une v.q.-a.s linéarisée sous G .*

- (i) *X est recouvert par des ouverts invariants par G et induits de variétés toriques stables affines.*
- (ii) *L'adhérence de toute orbite de G dans X est normale.*
- (iii) *Toute composante irréductible de X est induite d'une variété torique.*

PREUVE. Soit Z une orbite fermée de G dans X . Comme X est projective, Z est formée de points fixes de T , d'où $Z \simeq G/T \simeq A$. Comme X ne contient qu'un nombre fini d'orbites, la réunion de ces orbites qui contiennent Z dans leur adhérence est un ouvert invariant par G . En remplaçant X par cet ouvert, on peut donc supposer que Z est l'unique orbite fermée ; c'est l'ensemble des points fixes de T . Pour tout $x \in X$,

on a $Z \subseteq \overline{G \cdot x} = G \cdot \overline{T \cdot x}$ (car G/T est complet), c'est-à-dire $\overline{T \cdot x}$ rencontre Z en ses points fixes de T . Il en résulte qu'il existe un sous-groupe à un paramètre $\lambda : \mathbb{G}_m \rightarrow T$ (dépendant de x) tel que $\lambda(0) \cdot x \in Z$, c'est-à-dire le morphisme $\mathbb{G}_m \rightarrow X$, $t \mapsto \lambda(t) \cdot x$ s'étend en un morphisme $\mathbb{A}^1 \rightarrow X$ qui envoie 0 sur un point de Z .

Puisque X est munie d'un fibré en droites ample et T -linéarisé, elle est isomorphe à une sous-variété localement fermée et invariante par T du projectivisé d'un T -module, $\mathbb{P}(V)$; on peut supposer que Z est contenue dans $\mathbb{P}(V_0)$ où V_0 désigne le sous-espace des points fixes de T . La projection T -invariante $V \rightarrow V_0$ donne une application rationnelle $p : \mathbb{P}(V) \dashrightarrow \mathbb{P}(V_0)$ qui est définie en chaque $x \in X$ et l'envoie sur $\lambda(0) \cdot x$. On vérifie que le morphisme $p : X \rightarrow Z$ obtenu ainsi est affine et G -équivariant, ce qui démontre l'assertion (i).

Comme l'assertion (ii) est de nature locale, on peut d'après (i) supposer que $G = T$, d'où X est torique stable et affine. Soit Y l'adhérence d'une orbite de T dans X , et $\nu_Y : \tilde{Y} \rightarrow Y$ la normalisation. Alors \tilde{Y} est une variété torique affine sous un quotient de T , et ν_Y induit une bijection entre les ensembles de T -orbites dans \tilde{Y} et dans Y . Grâce à la connexité des stabilisateurs T_y , $y \in Y$, il en résulte que ν_Y induit des isomorphismes entre les corps résiduels. De plus, les ν_Y se recollent en $\nu : \lim \tilde{Y} \rightarrow X$ où \lim désigne la limite inductive sur l'ensemble partiellement ordonné des adhérences des orbites. Comme X est semi-normale, ν est un isomorphisme, ce qui entraîne (ii).

Soit X' une composante irréductible de X . En identifiant l'orbite ouverte de X' à G et en notant Y' l'adhérence de T dans X' , on obtient un morphisme birationnel propre et G -équivariant $\varphi : G \times^T Y' \rightarrow X'$. Montrons que φ est bijectif. Dans le cas contraire, comme la restriction de φ à chaque fibre de $p : G \times^T Y' \rightarrow G/T$ est injective, il existe $g \in G \setminus T$ tel que $g \cdot Y'$ rencontre Y' . L'intersection $Y' \cap g \cdot Y'$ contient alors un point fixe x de T . Soient $x = x_1, \dots, x_n$ les points fixes de T dans Y' . Puisque L est ample et T -linéarisé, les poids de l'action linéaire de T dans les fibres L_{x_1}, \dots, L_{x_n} sont deux à deux distincts. Puisque G est connexe, ces poids sont constants sur chacune des orbites $G \cdot x_1, \dots, G \cdot x_n$. On a donc $x_1 = g \cdot x_1$ d'où $g \in T$, ce qui est absurde. Puisque X' est normale, φ est un isomorphisme, ce qui démontre (iii).

Exemple 3.13. — Il existe des v.q.-a.s non induites : soit $G := E \times \mathbb{G}_m$ où E est une courbe elliptique, et soit X_1 la variété obtenue à partir de $E \times \mathbb{P}^1$ en identifiant $E \times \{0\}$ à $E \times \{\infty\}$ via $(x, 0) = (-x, \infty)$. On vérifie que X_1 est une variété quasi-abélienne stable sous G , qui n'admet aucun morphisme équivariant vers E .

On construit de même des exemples de v.q.-a.s linéarisées qui ne sont pas induites : soit X_2 la variété obtenue à partir de la réunion disjointe de deux exemplaires de $E \times \mathbb{P}^1$ via les identifications $(x_1, 0) = (x_2, 0)$ et $(x_1, \infty) = (x_2 + e, \infty)$ où e est un point de 2-torsion de E . Alors X_2 est une v.q.-a.s sous G , qui n'admet aucun morphisme équivariant vers E . Mais X_2 est linéarisée par le fibré en droites L obtenu en identifiant deux exemplaires de $M \otimes \mathcal{O}(1)$ où M est un fibré en droites de degré 2 sur E (si bien que $\tau_e^*(M) \simeq M$).

On va déduire de la proposition 3.12 un résultat de structure des v.q.-a linéarisées. Fixons les notations : soient $\pi : G \rightarrow G/T = A$ la projection, Λ le groupe des caractères de T , $V = \Lambda \otimes \mathbb{R}$, et L_λ , $\lambda \in \Lambda$ les fibrés en droites sur A tels que $\pi_*(\mathcal{O}_G) = \bigoplus_{\lambda \in \Lambda} L_\lambda$.

PROPOSITION 3.14. — *Soit (X, L) une v.q.-a linéarisée.*

(i) *Il existe un unique polytope convexe entier P d'intérieur non vide dans V , et un unique fibré en droites ample M sur A tels que*

$$(6) \quad X \simeq \text{Proj}_A \bigoplus_{n=0}^{\infty} \bigoplus_{\lambda \in \Lambda \cap nP} M^n \otimes L_\lambda \quad \text{et} \quad L \simeq \mathcal{O}(1).$$

Toute adhérence d'une orbite de G dans X s'obtient en remplaçant P par une face F dans (6), et ceci définit une bijection croissante entre adhérences d'orbites et faces.

(ii) *On a $H^i(X, L^n) = 0$ pour tous $i \geq 1$ et $n \geq 1$, et*

$$H^0(X, L^n) \simeq \bigoplus_{\lambda \in \Lambda \cap nP} H^0(A, M^n \otimes L_\lambda)$$

pour tout $n \geq 0$.

(iii) *Soit D le diviseur des zéros de $s \in H^0(X, L)$. Écrivons $s = \sum_{\lambda \in \Lambda \cap P} s_\lambda$ où $s_\lambda \in H^0(A, M \otimes L_\lambda)$. Pour que D ne contienne aucune orbite de G dans X , il faut et il suffit que $s_\lambda \neq 0$ pour tout sommet λ de P .*

PREUVE. D'après la proposition 3.12, $X \simeq G \times^T Y$ où Y est une variété torique projective sous T ; on note $p : X \rightarrow A$ la projection. Et d'après les résultats rappelés dans l'exemple 3.7, il existe un unique polytope convexe entier P dans V , d'intérieur non vide, tel que

$$H^0(Y, L^n) \simeq \bigoplus_{\lambda \in \Lambda \cap nP} k_\lambda$$

comme T -modules, où k_λ désigne le T -module de dimension 1 et de poids λ . De plus, $H^i(Y, L^n) = 0$ pour tous $i \geq 1$ et $n \geq 0$. D'après le théorème de cohomologie et changement de base, ceci entraîne la décomposition en espaces propres de T

$$p_*(L^n) \simeq \bigoplus_{\lambda \in \Lambda \cap nP} L_{n,\lambda}$$

où chaque $L_{n,\lambda}$ est un fibré en droites sur A , ainsi que l'annulation des $R^i p_*(L^n)$ lorsque $i \geq 1$ et $n \geq 0$. La multiplication de l'algèbre $\bigoplus_{n=0}^{\infty} p_*(L^n)$ se restreint en des isomorphismes

$$L_{n,\lambda} \otimes L_{p,\mu} \simeq L_{n+p,\lambda+\mu}.$$

Il existe donc des fibrés en droites $L_{0,\lambda}$ ($\lambda \in \Lambda$) et $L_{1,0}$ sur A , tels que $L_{n,\lambda} \simeq L_{1,0}^n \otimes L_{0,\lambda}$ chaque fois que $\lambda \in nP$. En considérant les restrictions à l'orbite ouverte $G \simeq G \times^T T$, on obtient des isomorphismes $L_{0,\lambda} \simeq L_\lambda$. D'où, en posant $M := L_{1,0}$:

$$H^i(X, L^n) \simeq \bigoplus_{\lambda \in \Lambda \cap nP} H^i(A, M^n \otimes L_\lambda)$$

pour tout $i \geq 0$. Comme L est ample et chaque L_λ est algébriquement trivial, on en déduit que M est gros (c'est-à-dire $h^0(M^n)$ croît comme n^g) ; et comme A est une variété abélienne, il en résulte que M est ample. Ceci entraîne les autres assertions.

Ainsi, les v.q.-a linéarisées (X, L) (sous une variété semi-abélienne G non spécifiée) sont classifiées par les quadruplets

$$(A, c, [M], P)$$

où A est une variété abélienne, $c : \Lambda \rightarrow A^\vee$ est l'homomorphisme associé à l'extension $1 \rightarrow T \rightarrow G \rightarrow A \rightarrow 0$, $[M] \in \text{Pic}(A)$ est une classe ample, et P est un polytope convexe entier dans V , d'intérieur non vide. Les adhérences des orbites de G dans X sont en bijection croissante avec les faces de P .

Plus généralement, à toute v.q.-a.s linéarisée (X, L) on associe la famille $\Delta(X, L)$ des polytopes associés aux adhérences de ses orbites sous G . Grâce aux propositions 3.12 et 3.14, $\Delta(X, L)$ vérifie les conditions de la

DÉFINITION 3.15. — *Un complexe de polytopes convexes entiers au-dessus de V est un espace topologique $|\Delta|$ muni d'un recouvrement Δ par des fermés, et d'une application de référence $\rho : |\Delta| \rightarrow V$ telles que :*

- (i) *La restriction de ρ à chaque $P \in \Delta$ est un homéomorphisme de P sur un polytope convexe entier dans V .*
- (ii) *Toute face d'un $P \in \Delta$ appartient à Δ .*
- (iii) *Pour tous $P, Q \in \Delta$, l'intersection $P \cap Q$ (considérée dans $|\Delta|$) est une réunion de faces de P et de Q .*

DÉFINITION 3.16. — *Le complexe (fini) $\Delta(X, L)$ est le type de la v.q.-a.s linéarisée (X, L) .*

L'application de référence n'est pas nécessairement injective : pour la v.q.-a.s linéarisée (X_2, L) de l'exemple 3.8, le type est formé de deux intervalles $[0, 1]$ recollés en leurs extrémités.

3.3. Linéarisation de l'action d'un tore

Dans la deuxième partie, on a construit des exemples de variétés complètes X munies d'une action d'un tore T et d'un fibré en droites ample L , telles que L n'est pas T -linéarisable mais le devient après un revêtement infini étale dont le groupe est celui des caractères de T . En fait, ce phénomène est bien plus général, comme le montrent les résultats de [3, Sec.4] qu'on va présenter brièvement.

Soit X une variété complète. Le foncteur contravariant qui à tout schéma S associe le groupe $\text{Pic}(X \times S)/p_2^* \text{Pic}(S)$ est représentable par un schéma en groupes $\mathbf{Pic}(X)$, localement de type fini sur $\text{Spec } k$. La composante neutre $\mathbf{Pic}^0(X)$ est un schéma en groupes de type fini ; son sous-schéma réduit $\mathbf{Pic}^0(X)_{\text{red}}$ est un groupe algébrique commutatif qu'on peut identifier au groupe $\text{Pic}^0(X)$ des classes d'isomorphie des fibrés en droites algébriquement triviaux.

PROPOSITION 3.17. — *Lorsque X est semi-normale, tout morphisme $\mathbb{A}^1 \rightarrow \mathbf{Pic}(X)$ est constant. De plus, $\mathbf{Pic}^0(X)$ est une variété semi-abélienne.*

PREUVE. La première assertion se déduit de l’isomorphisme $\mathbf{Pic}(X \times \mathbb{A}^1) \simeq p_1^* \mathbf{Pic}(X)$ vérifié pour toute variété semi-normale X (voir [24, Thm.3.6] pour le cas affine, et [3, Lem.4.1.10] pour le cas général).

D’après un théorème de Chevalley, le groupe algébrique $\mathbf{Pic}^0(X)$ est extension d’une variété abélienne par un groupe algébrique linéaire G , qui est ici commutatif et connexe. Mais on a vu que G ne contient aucun sous-groupe fermé isomorphe au groupe additif \mathbb{G}_a , donc G est un tore.

(Dans cet énoncé, l’hypothèse de semi-normalité est essentielle comme le montre l’exemple de la cubique plane cuspidale munie de l’action du groupe additif.)

Soit G un groupe algébrique opérant dans la variété complète X . Pour tout fibré en droites L sur X , on dispose d’un *morphisme de polarisation*

$$\lambda_L : G \rightarrow \mathbf{Pic}(X), \quad g \mapsto [L^{-1} \otimes g^*(L)].$$

C’est un morphisme de schémas, qui est constant si L admet une G -linéarisation. Lorsque G est connexe et réduit, l’image de λ_L est contenue dans $\mathbf{Pic}^0(X)$.

PROPOSITION 3.18. — (i) *Lorsque X est munie d’une action d’une variété semi-abélienne G , le morphisme de polarisation λ_L est un homomorphisme de groupes pour tout fibré en droites L sur X . Autrement dit, le théorème du carré est vérifié pour l’action naturelle de G dans $\mathbf{Pic}(X)$.*

(ii) *Étant donné un tore T de groupe des caractères Λ , on a une correspondance biunivoque entre les homomorphismes $T \rightarrow \mathbf{Pic}(X)$ et les classes d’isomorphie des Λ -torseurs sur X , c’est-à-dire des schémas \tilde{X} munis d’une action propre de Λ telle que $\tilde{X}/\Lambda \simeq X$.*

PREUVE. (i) Pour tout entier $m \geq 1$, soit ${}_m\mathbf{Pic}^0(X)$ le noyau (ensembliste) de l’endomorphisme $[L] \mapsto [L^m]$ du groupe $\mathbf{Pic}^0(X)$. Les ${}_m\mathbf{Pic}^0(X)$ forment une suite croissante de sous-groupes finis dont la réunion est dense dans $\mathbf{Pic}^0(X)$. L’action naturelle de G dans $\mathbf{Pic}(X)$ préserve chaque ${}_m\mathbf{Pic}^0(X)$; puisque G est connexe, il opère trivialement dans chaque ${}_m\mathbf{Pic}^0(X)$ et donc dans $\mathbf{Pic}^0(X)$. Ceci entraîne l’isomorphisme

$$g_1^*(L^{-1} \otimes g_2^*(L)) \simeq L^{-1} \otimes g_2^*(L)$$

pour tous $g_1, g_2 \in G$, et donc le théorème du carré.

(ii) s’obtient en adaptant celle de l’énoncé analogue pour les homomorphismes d’un schéma en groupes finis vers $\mathbf{Pic}(X)$ [23, Prop.6.2.1].

Dans [3, Sec.4.1,4.2], la proposition 3.18 est généralisée aux schémas $\pi : \mathcal{X} \rightarrow S$ propres, plats et à fibres géométriques connexes et semi-normales. Cette généralisation permet d’établir un résultat clé [3, Thm.4.3] :

THÉORÈME 3.19. — *Étant donné un tore scindé T sur un schéma S , de groupe des caractères Λ , on a une équivalence de catégories entre :*

(a) *Les couples $(\mathcal{X}, \mathcal{L})$, où*

(i) *\mathcal{X} est un schéma propre sur S , à fibres géométriquement connexes, réduites et semi-normales, muni d'une action de T .*

(ii) *\mathcal{L} est un fibré en droites relativement ample sur \mathcal{X} .*

(b) *Les couples $(\tilde{\mathcal{X}}, \tilde{\mathcal{L}})$, où*

(i) *$\tilde{\mathcal{X}}$ est un schéma localement de type fini sur S , à fibres géométriquement réduites et semi-normales, muni d'une action de $T \times \Lambda$.*

(ii) *$\tilde{\mathcal{L}}$ est un fibré en droites relativement ample sur $\tilde{\mathcal{X}}$, muni d'actions de T et de Λ qui relèvent leurs actions dans $\tilde{\mathcal{X}}$.*

(iii) *Λ opère proprement dans $\tilde{\mathcal{X}}$ et le quotient $\tilde{\mathcal{X}}/\Lambda$ est propre sur S , à fibres géométriquement connexes.*

(iv) *Pour les actions de T et Λ dans $\tilde{\mathcal{L}}$, on a la relation de commutation $t \circ \lambda = \lambda(t) \lambda \circ t$ pour tout point fonctoriel t de T et tout $\lambda \in \Lambda$.*

Lorsque S est connexe, les composantes connexes de $\tilde{\mathcal{X}}$ sont paramétrées par le groupe des caractères du noyau de l'homomorphisme de polarisation $\lambda_L : T \rightarrow \mathbf{Pic}(\mathcal{X}/S)$.

En particulier, à toute v.q.-a.s polarisée (X, L) sous la variété semi-abélienne G de tore maximal T , on associe un revêtement étale infini $\pi : \tilde{X} \rightarrow X$ de groupe Λ , où \tilde{X} est un schéma localement de type fini muni d'une action de G qui relève l'action dans X et commute à celle de Λ . Par suite, \tilde{X} est semi-normale et toute réunion finie de ses composantes irréductibles est une v.q.-a.s linéarisée relativement au fibré en droites linéarisé $\tilde{L} := \pi^*(L)$. À l'aide des propositions 3.12 et 3.14, on peut donc définir le *type*

$$\tilde{\Delta} := \Delta(\tilde{X}, \tilde{L}).$$

C'est un complexe *localement fini* de polytopes convexes entiers, muni d'une action de Λ telle que l'application de référence $\tilde{\rho} : |\tilde{\Delta}| \rightarrow V$ est équivariante pour l'action de Λ par translations sur V . Le quotient $\Delta := \tilde{\Delta}/\Lambda$ peut être vu comme un complexe de polytopes convexes entiers muni d'une application de référence à valeurs dans le tore réel V/Λ .

Passons en revue quelques classes d'exemples, en commençant par les plus simples.

Exemple 3.20. — Les couples abéliens sont ceux dont le type est un point.

Exemple 3.21. — Lorsque le fibré en droites L est linéarisé, le schéma \tilde{X} est réunion disjointe d'exemplaires de X notés X_λ , $\lambda \in \Lambda$. L'action de chaque $\lambda \in \Lambda$ dans \tilde{X} est donnée par les identifications $X_\mu \rightarrow X_{\lambda+\mu}$. La restriction à X_λ du fibré en droites \tilde{L} est le fibré L muni de sa linéarisation tordue par le caractère λ .

Exemple 3.22. — Pour la dégénérescence maximale $(\mathcal{X}_s, \mathcal{L}_s)$ du théorème 2.7, l'espace topologique $|\tilde{\Delta}|$ est identifié à V par l'application de référence, qui identifie $\tilde{\Delta}$ à la décomposition de Delaunay Del_Q .

Exemple 3.23. — On va décrire le type du couple $(\text{Jac}^{g-1}(C), \Theta)$ associé dans l'exemple 3.4 à une courbe complète et nodale C de genre arithmétique g . Avec les notations de cet exemple et de l'exemple 1.2, on obtient $\Lambda = H_1(\Gamma, \mathbb{Z}) \subseteq C_1(\Gamma, \mathbb{Z}) \simeq \mathbb{Z}^d$, d'où $V = H_1(\Gamma, \mathbb{R}) \subseteq \mathbb{R}^d$. D'après [4, Sec.2], on a encore $|\tilde{\Delta}| \simeq V$ via l'application de référence, et $\tilde{\Delta}$ n'est autre que la décomposition de Delaunay de V relative à la restriction de la forme quadratique standard sur \mathbb{R}^d . Ainsi, le type est la trace sur V du pavage entier de \mathbb{R}^d en «cubes» découpés par les hyperplans de coordonnées et leurs translatés entiers.

Ces deux derniers exemples motivent la

DÉFINITION 3.24. — *Le type $\tilde{\Delta}$ d'une v.q.-a.s polarisée est dit périodique de degré 1 si son application de référence est un homéomorphisme qui identifie $\tilde{\Delta}$ à un pavage périodique de V par des polytopes convexes entiers (au sens défini en 2.1).*

En particulier, pour une v.q.-a.s polarisée (X, L) dont le type $\tilde{\Delta}$ est périodique de degré 1, l'ensemble des sommets des polytopes de $\tilde{\Delta}$ n'est autre que Λ . Autrement dit, X contient une unique orbite fermée de G , qui est alors un A -torseur.

3.4. Espaces des modules de couples stables

La classification des variétés et des couples quasi-abéliens stables peut se déduire des résultats de 3.2 et 3.3 : grâce au théorème 3.19, on se ramène au cas linéarisé, où les objets considérés s'obtiennent par recollement de fermés irréductibles qu'on a déterminés dans la proposition 3.14(i). On ne détaille pas ici cette classification, pour laquelle on renvoie à [3, Sec.1.2], mais on en énonce deux conséquences importantes :

PROPOSITION 3.25. — *Pour toute v.q.-a.s polarisée (X, L) , on a $H^i(X, L^n) = 0$ lorsque $i \geq 1$ et $n \geq 1$.*

Pour une v.q.-a linéarisée, c'est la proposition 3.14(ii) ; le cas général en est déduit dans [3, Thm.2.5.1, Thm.5.4.1].

Il en résulte que pour tout couple quasi-abélien stable $(\mathcal{X}, \mathcal{D})$ sur S , le degré de $(\mathcal{X}_{\bar{s}}, \mathcal{D}_{\bar{s}})$ est constant sur chaque composante connexe de S .

THÉORÈME 3.26. — *Soit $(\mathcal{X}_\eta, \mathcal{D}_\eta)$ un couple quasi-abélien stable sur le point générique d'un trait S . Alors, quitte à faire un changement de base fini ramifié, on peut étendre $(\mathcal{X}_\eta, \mathcal{D}_\eta)$ en un couple quasi-abélien stable $(\mathcal{X}, \mathcal{D})$ sur S ; une telle extension est unique à isomorphisme près. Lorsque le type de $(\mathcal{X}_\eta, \mathcal{D}_\eta)$ est périodique de degré 1, il en est de même du type de la fibre spéciale.*

Cet analogue du théorème de réduction semi-stable est d’abord établi pour les couples toriques stables [3, Sec.2.8]. Le cas général est déduit dans [3, Sec.5.7] du cas torique stable et de la construction de dégénérescences des couples abéliens obtenue dans [5] (voir le théorème 2.7 pour les dégénérescences maximales).

Le théorème 3.26 est l’un des principaux ingrédients dans la preuve du résultat principal de ce texte [3, Sec.1.2.H] :

THÉORÈME 3.27. — (i) *Les couples quasi-abéliens stables de dimension g et de type périodique de degré 1 forment un champ algébrique propre $\overline{\mathcal{A}}_g^{\text{mod}}$ dont le morphisme diagonal est fini. Les couples de type fixé forment un sous-champ localement fermé.*

(ii) *Le champ $\overline{\mathcal{A}}_g^{\text{mod}}$ admet un espace des modules grossier $\overline{A}_g^{\text{mod}}$. C’est un espace algébrique propre muni d’une stratification par le type, indexée par les classes d’isomorphie des pavages périodiques des \mathbb{R}^r , $r = 0, 1, \dots, g$, par des polytopes convexes entiers.*

(iii) *La strate de $\overline{A}_g^{\text{mod}}$ associée à $r = 0$ n’est autre que A_g ; son adhérence dans $\overline{A}_g^{\text{mod}}$ en est une composante irréductible dite principale, dont la normalisation est la compactification toroïdale $\overline{A}_g^{\text{Vor}}$ associée à la deuxième subdivision de Voronoi (voir 1.4).*

Indiquons les grandes lignes de la démonstration. On commence par obtenir un résultat similaire pour les couples toriques stables linéarisés dont le type est une subdivision d’un polytope convexe entier donné ; dans ce cas, on construit directement le champ des modules par des méthodes de géométrie torique [3, Sec.2]. Ces méthodes ne se transposent pas telles quelles au cas périodique de degré 1, mais elles permettent de décrire les déformations infinitésimales des couples stables (car tout schéma semi-abélien sur un anneau local artinien est extension d’un schéma abélien par un tore). L’algébricité du champ $\overline{\mathcal{A}}_g^{\text{mod}}$ se déduit alors d’un critère d’Artin [12, Thm.10.10]. Et sa propriété résulte du critère valuatif et du théorème 3.26 ; l’existence de son espace des modules grossier s’obtient en appliquant le résultat principal de [10].

Remarque 3.28. — L’énoncé (iii) ci-dessus apparaît dans [3, Thm.5.11.6] sous une forme beaucoup plus simple : la composante principale de $\overline{A}_g^{\text{mod}}$ est isomorphe à $\overline{A}_g^{\text{Vor}}$. Mais la preuve de ce résultat, qui équivaut bien sûr à la normalité de la composante principale, n’est pas tout à fait complète.

Le théorème 3.27 et les propriétés des jacobiniennes compactifiées relatives énoncées dans l’exemple 3.10 entraînent aussitôt l’existence d’une version modulaire du morphisme $\bar{\mathfrak{t}}_{\mathbb{C}} : \overline{M}_{g,\mathbb{C}} \rightarrow \overline{A}_{g,\mathbb{C}}^{\text{Vor}}$.

THÉORÈME 3.29. — *En associant à toute courbe stable $\pi : \mathcal{C} \rightarrow S$ de genre g le couple quasi-abélien stable $(\mathbf{Jac}^{g-1}(\mathcal{C}/S), \Theta) \rightarrow S$, on définit un morphisme de Torelli compactifié*

$$\bar{\mathfrak{t}} : \overline{M}_g \rightarrow \overline{A}_g^{\text{mod}}$$

dont l’image est contenue dans la composante principale.

Remarque 3.30. — Rappelons que les strates de $\overline{A}_g^{\text{Vor}}$ sont paramétrées par les classes d’isomorphie des décompositions de Delaunay des \mathbb{R}^r où $r = 0, 1, \dots, g$. On montre que lorsque $g \leq 3$, tout pavage périodique de \mathbb{R}^g par des polytopes convexes entiers est une décomposition de Delaunay, mais qu’il n’en est plus ainsi dès que $g \geq 4$; dans ce dernier cas, il en résulte que $\overline{A}_g^{\text{mod}}$ admet des composantes irréductibles autres que la composante principale.

Ces composantes sont analysées plus en détail dans [2] à l’aide d’une formule calculant la dimension de la strate associée à un type donné. Il y est montré en particulier que $\overline{A}_g^{\text{mod}}$ est irréductible lorsque $g \leq 3$, mais que la dimension de $\overline{A}_g^{\text{mod}}$ croît au moins comme 2^g , bien plus vite que la dimension $g(g+1)/2$ de la composante principale.

Ainsi, \overline{A}_g est loin d’être dense dans $\overline{A}_g^{\text{mod}}$ lorsque g est grand. En fait, il n’est pas rare qu’une compactification modulaire d’un espace de paramètres introduise de nouvelles composantes irréductibles ; c’est le cas, par exemple, du schéma de Hilbert ponctuel $X^{[n]}$ qui compactifie les n -uplets de points non ordonnés, deux à deux distincts, d’une variété X de dimension au moins 3.

Remarque 3.31. — Les résultats de [3] concernent, plus généralement, les couples abéliens stables de dimension g et degré d arbitraire ; on y montre qu’ils admettent un espace des modules grossier, noté $\text{AP}_{g,d}$. En associant à chaque couple la variété abélienne polarisée sous-jacente, on définit un morphisme

$$\text{AP}_{g,d} \rightarrow A_{g,d}$$

qui est projectif de dimension relative $d - 1$; ici $A_{g,d}$ désigne l’espace des modules grossier des variétés abéliennes de dimension g munies d’une polarisation de degré d . Et $\text{AP}_{g,d}$ admet aussi une compactification modulaire $\overline{\text{AP}}_{g,d}^{\text{mod}}$ paramétrant les couples quasi-abéliens stables dont le type est *périodique de degré d* , c’est-à-dire possède exactement d composantes connexes, chacune étant identifiée par l’application de référence à un pavage de V , périodique pour un sous-groupe d’indice d de Λ .

Cependant, la construction d’une compactification modulaire de $A_{g,d}$ est une question ouverte lorsque $d \geq 2$.

De même, on dispose d’espaces des modules grossiers $A_{g,d,n}$ qui paramètrent les variétés abéliennes de dimension g munies d’une polarisation de degré d et d’une *structure de niveau n* , c’est-à-dire d’un isomorphisme de $(\mathbb{Z}/n\mathbb{Z})^{2g}$ sur leur sous-groupe de n -torsion. Lorsque $n \geq 3$, il s’agit même d’espaces des modules fins (voir [15]).

Mais on ne connaît pas de compactification des espaces $A_{g,d,n}$ qui soit modulaire au sens du théorème 3.27 (voir cependant [18, 17] pour des compactifications de certains $A_{g,d,n}$, modulaires en un sens plus faible). On ignore aussi comment définir des structures de niveau n pour les couples quasi-abéliens stables.

RÉFÉRENCES

- [1] V. ALEXEEV – *Log canonical singularities and complete moduli of stable pairs*. arXiv : alg-geom/9608013.
- [2] V. ALEXEEV – *On extra components in the functorial compactification of A_g* . Moduli of abelian varieties (Texel Island, 1999), 1–9, Birkhäuser, Basel, 2001.
- [3] V. ALEXEEV – *Complete moduli in the presence of semiabelian group action*. Ann. Math. **155** (2002), 611–708.
- [4] V. ALEXEEV – *Compactified Jacobians and Torelli map*. Publ. Res. Inst. Math. Sci. **40** (2004), no. 4, 1241–1265.
- [5] V. ALEXEEV, I. NAKAMURA – *On Mumford’s construction of degenerating abelian varieties*. Tohoku Math. J. **51** (1999), 399–420.
- [6] A. ASH, D. MUMFORD, M. RAPOPORT, Y. TAI – *Smooth compactification of locally symmetric varieties*. Math Sci Press, Brookline, 1975.
- [7] S. BOSCH, W. LÜTKEBOHMERT, M. RAYNAUD – *Néron models*. Springer-Verlag, Berlin, 1990.
- [8] G. FALTINGS, C.-L. CHAI – *Degeneration of Abelian Varieties*. Springer-Verlag, Berlin, 1990.
- [9] R. HARTSHORNE – *Algebraic Geometry*. Springer-Verlag, Berlin, 1977.
- [10] S. KEEL, S. MORI – *Quotients by groupoids*. Ann. Math. **145** (1997), 193–213.
- [11] J. KOLLÁR, N. I. SHEPHERD–BARRON, *Threefolds and deformations of surface singularities*. Invent. Math. **91** (1988), 299–338.
- [12] G. LAUMON, L. MORET–BAILLY – *Champs algébriques*. Springer-Verlag, Berlin, 2000.
- [13] D. MUMFORD – *Abelian Varieties*. Oxford University Press, Oxford, 1970.
- [14] D. MUMFORD – *An analytic construction of degenerate abelian varieties over complete rings*. Compos. Math. **24** (1972), 239–272.
- [15] D. MUMFORD, J. FOGARTY, F. KIRWAN – *Geometric Invariant Theory*. Third edition. Springer-Verlag, Berlin, 1994.
- [16] I. NAKAMURA – *On moduli of stable quasi abelian varieties*. Nagoya Math. J. **58** (1975), 149–214.
- [17] I. NAKAMURA – *Stability of degenerate abelian varieties*. Invent. Math. **136** (1999), 659–715.
- [18] Y. NAMIKAWA – *A new compactification of Siegel spaces and degenerations of abelian varieties I, II*. Math. Ann. **221** (1976), 97–141, 201–241.
- [19] Y. NAMIKAWA – *Toroidal Compactification of Siegel Spaces*. Lecture Notes in Math. **812**, Springer-Verlag, Berlin, 1980.
- [20] T. ODA, C. S. SESHADRI – *Compactifications of the generalized jacobian variety*. Trans. Amer. Math. Soc. **253** (1979), 1–90.
- [21] M. C. OLSSON – *Modular description of the main components in Alexeev’s compactifications*. <http://www.ma.utexas.edu/~molsson/>

- [22] M. RAYNAUD – *Spécialisation du foncteur de Picard*. Publ. Math. IHÉS **38** (1970), 27–76.
- [23] M. RAYNAUD – *Compactification du module des courbes*. Séminaire Bourbaki (23ème année, 1970/1971), Exp. No. 385, pp. 47–61. Lecture Notes in Math. **244**, Springer-Verlag, Berlin, 1971.
- [24] C. TRAVERSO – *Seminormality and Picard group*. Ann. Scuola Norm. Sup. Pisa (3) **24** (1970), 585–595.

Michel BRION

Université Grenoble I

Institut Joseph Fourier

UMR 5582 du CNRS

B.P. 74

F–38402 Saint-Martin-d’Hères Cedex

E-mail : Michel.Brion@ujf-grenoble.fr

**EXPLOSION POUR L'ÉQUATION DE SCHRÖDINGER
AU RÉGIME DU « LOG LOG »
[d'après Merle-Raphael]**

par **Nicolas BURQ**

INTRODUCTION

On se propose dans cet exposé de présenter quelques résultats récents sur l'explosion pour l'équation de Schrödinger non linéaire :

$$(1) \quad i \frac{\partial}{\partial t} u + \Delta u + |u|^{\frac{4}{d}} u = 0, \quad (t, x) \in \mathbb{R}^{1+d}, \quad \Delta u = \sum_{j=1}^d \frac{\partial^2 u}{\partial x_j^2}.$$

Cette équation apparaît dans plusieurs modèles mathématiques de phénomènes physiques : la propagation d'ondes dans un milieu non linéaire, la propagation dans les fibres optiques ou la condensation de Bose–Einstein (équation de Gross–Pitaevskii). Elle possède (au moins formellement) trois lois de conservation :

– Conservation de la masse

$$(2) \quad \|u(t)\|_{L^2(\mathbb{R}^d)} = \|u(0)\|_{L^2(\mathbb{R}^d)}.$$

– Conservation de l'énergie

$$(3) \quad E(u)(t) = \int_{\mathbb{R}^d} \left(\frac{1}{2} |\nabla_x u|^2 - \frac{d}{2d+4} |u|^{\frac{2d+4}{d}} \right) dx = E(u)(0).$$

– Conservation du moment cinétique

$$(4) \quad \mathbf{Im} \left(\int \nabla_x u \bar{u}(t, x) dx \right) = \mathbf{Im} \left(\int \nabla_x u \bar{u}(0, x) dx \right).$$

Ces invariants sont reliés aux invariances de l'équation dans l'espace d'énergie H^1 : si $u(t, x)$ est solution de (1) alors

– $u(t_0 + t, x_0 + x)$ aussi (invariance par translation),

– $u(t, x)e^{i\gamma}$ aussi (invariance de phase),

– $\lambda^{\frac{d}{2}} u(\lambda^2 t, \lambda x)$ aussi (invariance d'échelle),

– $u(t, x - \beta t)e^{i\frac{\beta}{2} \cdot (x - \frac{\beta}{2} t)}$, $\beta \in \mathbb{R}^d$ aussi (invariance galiléenne).

Une quatrième symétrie peut-être moins évidente (et qui n'agit pas sur H^1) est l'invariance conforme : si $u(t, x)$ est solution de (1) alors il en est de même de

$$(5) \quad v(t, x) = \frac{1}{|t|^{\frac{d}{2}}} \bar{u}\left(\frac{1}{t}, \frac{x}{t}\right) e^{i\frac{x^2}{4t}}.$$

On peut remarquer que l'application

$$u \mapsto u_\lambda, \quad u_\lambda(x) = \lambda^{\frac{d}{2}} u(\lambda x)$$

préserve la norme L^2 : l'équation (1) est dite L^2 -critique (un autre choix de puissance dans la non-linéarité donnerait une autre puissance de λ pour l'invariance d'échelle).

L'équation (1) est un système hamiltonien de dimension infinie. La partie *linéaire* de l'équation, $i\partial_t + \Delta$, possède des propriétés dispersives et un des points importants de l'étude de (1) consiste à comprendre l'interaction entre ces propriétés de dispersion et l'effet de focalisation dû à la non-linéarité. Dans ce contexte des équations dispersives, l'équation de Schrödinger non linéaire est, avec l'équation de Korteweg–de Vries critique (voir les travaux de Martel et Merle [MM00, MM02a, MM02b, MM04] qui ont inspiré le travail de Merle et Raphael et l'exposé au séminaire Bourbaki de Tzvetkov [Tzv04] sur le sujet), un modèle important pour lequel on est capable d'exhiber des solutions *explosives*.

Les questions auxquelles on va s'intéresser dans cet exposé sont les suivantes :

- Existe-t-il des solutions explosives autres que celles (explicites) qui sont connues depuis les années 60 ?
- Peut-on classifier les types d'explosion possibles ?
- Quels types d'explosion sont stables ?

Remerciements. Je remercie P. Gérard pour les discussions que j'ai eues avec lui sur le sujet de cet exposé et P. Raphael qui a passé du temps à m'expliquer de nombreux points de leur preuve et dont les notes de cours sur le sujet [Rap] ont été une source d'inspiration.

1. LE CARACTÈRE BIEN POSÉ DANS H^1 , CRITÈRES D'EXPLOSION ET DE NON-EXPLOSION

1.1. Non-explosion : normes L^2 petites

Le caractère localement bien posé de (1) pour des données initiales L^2 ou H^1 (norme quelconque) est connu depuis les travaux de Ginibre et Velo [GV79] :

THÉORÈME 1.1. — *Pour tout $C > 0$ il existe $T > 0$ tel que pour toute donnée initiale $u_0 \in H^1(\mathbb{R}^d)$, $\|u_0\|_{H^1} \leq C$, il existe une unique solution $u \in C([0, T[; H^1)$ de l'équation (1) vérifiant $u|_{t=0} = u_0$.*

On peut remarquer que le temps d'existence de la solution est minoré par une fonction de la norme H^1 de la donnée initiale. Si on note T le temps maximal d'existence on a ainsi, si $T < +\infty$

$$\lim_{t \rightarrow T-0} \|u(t)\|_{H^1} = +\infty.$$

On dit qu'on a *explosion* en temps fini de la solution. On peut aussi avoir explosion en temps infini si

$$T = +\infty \text{ et } \limsup_{t \rightarrow T-0} \|u(t)\|_{H^1} = +\infty.$$

Pour des données initiales petites dans L^2 , la solution est globale en temps : ceci est une conséquence facile de l'inégalité de Gagliardo–Nirenberg suivante :

PROPOSITION 1.2. — *Il existe $C > 0$ tel que pour tout $u \in H^1(\mathbb{R}^d)$, on a*

$$\int_{\mathbb{R}^d} |u|^{\frac{4}{d}+2} dx \leq C \int_{\mathbb{R}^d} |\nabla u|^2 dx \left(\int_{\mathbb{R}^d} |u|^2 dx \right)^{\frac{2}{d}}.$$

En effet, on déduit de l'inégalité précédente que si la norme L^2 de la donnée initiale (et donc de la solution d'après (2)) est petite, alors

$$E(u(t)) \geq \frac{1}{4} \int_{\mathbb{R}^d} |\nabla u(t)|^2 dx$$

donc notre solution reste bornée dans H^1 (compte tenu de la conservation de la norme L^2) et $T = +\infty$.

On peut préciser la condition « norme L^2 petite » (voir Weinstein [Wei83]).

PROPOSITION 1.3. — *On considère le problème de minimisation suivant :*

$$m = \inf_{0 \neq v \in H^1} \frac{\left(\int_{\mathbb{R}^d} |\nabla v|^2 dx \right) \left(\int_{\mathbb{R}^d} |v|^2 dx \right)^{\frac{2}{d}}}{\int_{\mathbb{R}^d} |v|^{\frac{4}{d}+2} dx}.$$

Alors m est atteint pour la famille à trois paramètres de fonctions

$$\lambda^{\frac{d}{2}} Q(\lambda x + x_0) e^{i\gamma}, \quad (\lambda, x_0, \gamma) \in \mathbb{R}^{*,+} \times \mathbb{R}^d \times \mathbb{R},$$

où Q est l'unique solution positive radiale et exponentiellement décroissante à l'infini du système

$$\Delta Q - Q + Q^{\frac{4}{d}+1} = 0, \quad Q(r) \rightarrow 0, r \rightarrow +\infty.$$

(On appelle Q « l'état fondamental »).

On en déduit l'inégalité de Gagliardo–Nirenberg précisée :

$$E(v) \geq \frac{1}{2} \int |\nabla v|^2 \left(1 - \left(\frac{\|v\|_{L^2(\mathbb{R}^d)}}{\|Q\|_{L^2(\mathbb{R}^d)}} \right)^{\frac{4}{d}} \right).$$

On voit immédiatement que si $\|u_0\|_{L^2(\mathbb{R}^d)} < \|Q\|_{L^2(\mathbb{R}^d)}$, les deux quantités conservées, énergie et masse, impliquent que la norme H^1 reste bornée et donc que la solution existe globalement (en temps) :

PROPOSITION 1.4. — *Pour tout $u_0 \in H^1$ telle que $\|u_0\|_{L^2(\mathbb{R}^d)} < \|Q\|_{L^2(\mathbb{R}^d)}$, la solution de (1) de donnée initiale u_0 existe pour tout temps.*

Remarque 1.5. — On peut également montrer en utilisant les estimations de Strichartz dues dans ce contexte à Ginibre et Velo [GV79] que, pour toute donnée initiale $u_0 \in L^2$, il existe une solution locale en temps de (1) et que la solution est globale ($T = +\infty$) si la norme de la donnée initiale est petite dans L^2 . La question de savoir si la borne garantissant l'existence globale est $\|Q\|_{L^2(\mathbb{R}^d)}$ (comme au niveau H^1) est ouverte.

1.2. Ondes solitaires et explosion à masse critique

Il est remarquable de constater que non seulement Q fournit un critère de non-explosion, mais qu'il en établit aussi le caractère optimal. En effet, la fonction $e^{it}Q(x)$ (le soliton de l'équation de Schrödinger) est clairement solution de (1). En utilisant l'invariance conforme (5), on peut définir

$$S(t) = \frac{1}{|t|^{\frac{d}{2}}} Q\left(\frac{x}{t}\right) e^{i\left(\frac{1}{t} - \frac{x^2}{4t}\right)},$$

qui est la solution de (1) de donnée initiale à $t = -1$ égale à $Q(x)e^{i\left(\frac{x^2}{4} - 1\right)}$. Alors $S(t)$ vérifie

1. $E(S(t)) > 0$,
2. $\|S(t)\|_{L^2(\mathbb{R}^d)} = \|Q\|_{L^2(\mathbb{R}^d)}$,
3. $\|\nabla S(t)\|_{L^2(\mathbb{R}^d)} \sim \frac{C}{|t|}$, $t \rightarrow 0, t > 0$,
4. $S(t)$ se concentre en $x = 0$ quand t tend vers 0 : au sens de la convergence faible des mesures,

$$|S(t)|^2 dx \rightarrow \|Q\|_{L^2(\mathbb{R}^d)}^2 \delta_{x=0}.$$

En fait ces propriétés caractérisent la fonction $S(t)$.

THÉORÈME 1.6 (Merle [Mer93]). — *Soit $u_0 \in H^1$ telle que $\|u_0\|_{L^2(\mathbb{R}^d)} = \|Q\|_{L^2(\mathbb{R}^d)}$. On suppose que la solution de (1) explose en temps fini. Alors aux symétries de translation, phase, dilatation et invariance conforme près, $u(t, x) = S(t, x)$.*

1.3. Explosion à énergie négative

Une question naturelle, compte tenu du théorème 1.6 est de savoir si pour $\|u_0\|_{L^2(\mathbb{R}^d)} > \|Q\|_{L^2(\mathbb{R}^d)}$ il existe effectivement des solutions explosives de (1) (et éventuellement de les décrire). Grâce à l'identité du viriel due à Zakharov et Shabat [ZS71],

$$(6) \quad \frac{d^2}{dt^2} \left(\int |x|^2 |u(t, x)|^2 dx \right) = 4 \frac{d}{dt} \operatorname{Im} \left(\int x \nabla u \bar{u} dx \right) = 16E_0,$$

on a une première réponse très simple. En effet, on peut alors vérifier que si la donnée initiale est dans l'espace du viriel

$$\Sigma = \{u \in H^1; xu \in L^2\},$$

alors la solution reste dans cet espace tant qu'elle existe (en tant que solution dans H^1) et vérifie (6). En particulier, si $E_0 < 0$, la solution explose en temps fini (le terme de gauche dans (6) est positif mais a une dérivée seconde constante strictement négative).

Cet argument a pu être généralisé au cadre de données initiales qui ne sont plus dans Σ mais seulement dans H^1 :

THÉORÈME 1.7. — *Soit $u_0 \in H^1$ telle $E_0 < 0$. Alors la solution de (1) de donnée initiale u_0 explose en temps fini si*

1. $d = 1$ (Ogawa, Tsutsumi [OT91]) ou
2. $d \geq 2$ et u_0 est radiale (Nawa [Naw99]).

Un inconvénient majeur de ce résultat est qu'il fournit juste une obstruction à l'existence globale. Il ne dit rien sur l'explosion proprement dite ni sur le comportement de la solution.

1.4. Solutions auto-similaires

Compte tenu de l'invariance par changement d'échelle de l'équation, il est naturel de chercher des solutions sous la forme

$$U_b(t, x) = \frac{1}{(2b(T-t))^{\frac{d}{4}}} Q_b \left(\frac{x}{\sqrt{2b(T-t)}} \right) e^{-i \frac{\log(T-t)}{2b}}$$

avec Q_b solution de

$$\Delta Q_b - Q_b + ib \left(\frac{d}{2} Q_b + y \cdot \nabla Q_b \right) + Q_b |Q_b|^{\frac{4}{d}} = 0.$$

De telles solutions ne sont jamais dans L^2 à cause d'une divergence de type logarithmique à l'infini :

$$|Q_b(x)| \sim \frac{C(b)}{|x|^{\frac{d}{2}}}, \quad |x| \rightarrow +\infty.$$

Néanmoins ces solutions joueront un rôle important dans la suite.

Un autre résultat (élémentaire) relié à cette invariance par changement d'échelle est le suivant :

PROPOSITION 1.8. — *On considère u une solution de (1) de donnée initiale $u_0 \in H^1$ et explosant en temps fini, T . Alors il existe $C > 0$ (ne dépendant que de $\|u_0\|_{L^2(\mathbb{R}^d)}$) tel que*

$$(7) \quad \forall t \in [0, T[, \quad \|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \geq \frac{C}{\sqrt{T-t}}.$$

En effet, fixons $t \in [0, T[$ et considérons

$$v^t(s, y) = \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-\frac{d}{2}} u(t + \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-2} s, \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-1} y).$$

La fonction v^t est la solution de (1) de donnée initiale (à $s = 0$)

$$v_0^t(y) = \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-\frac{d}{2}} u(t, \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-1} y)$$

qui vérifie

$$\|v_0^t\|_{L^2(\mathbb{R}^d)} = \|u_0\|_{L^2(\mathbb{R}^d)}, \quad \|\nabla v_0^t\|_{L^2(\mathbb{R}^d)} = 1.$$

Donc d’après la théorie locale dans H^1 , il existe s_0 ne dépendant que de $\|u_0\|_{L^2(\mathbb{R}^d)}$, tel que v^t existe pour $s \in [0, s_0]$. Par conséquent,

$$t + \|\nabla u(t)\|_{L^2(\mathbb{R}^d)}^{-2} s_0 \leq T$$

ce qui est (7)

1.5. Dynamique à l’explosion

La plupart des résultats sur la dynamique à l’explosion concernent le régime perturbatif :

$$u_0 \in \mathcal{B}_\alpha = \{u_0 \in H^1; \quad \|Q\|_{L^2(\mathbb{R}^d)}^2 \leq \|u_0\|_{L^2(\mathbb{R}^d)}^2 \leq \|Q\|_{L^2(\mathbb{R}^d)}^2 + \alpha\}, \quad 0 < \alpha \ll 1$$

On connaît deux différents types de comportement à l’explosion pour les solutions de (1)

1. L’explosion en $\|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \sim \frac{1}{T-t}$ dont le modèle est $S(T-t)$ (voir aussi les solutions construites par Bourgain et Wang [BW97]).
2. Des simulations numériques (voir les travaux de Landman, Papanicolaou, Sulem et Sulem [LPSS88] et la monographie de Sulem-Sulem [SS99]) ont prédit dans les années 80 l’existence de solutions explosant au régime du « log log » :

$$\|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \sim \left(\frac{\log |\log(t)|}{T-t} \right)^{\frac{1}{2}}.$$

Ces solutions apparaissent (numériquement) beaucoup plus stables que les précédentes. Le premier résultat mathématique confirmant l’existence de telles solutions est dû à Perelman [Per01] : elle construit en dimension 1 une telle solution et démontre sa stabilité dans un certain espace strictement inclus dans H^1 .

La situation a été complètement clarifiée dans une série de papiers par Merle et Raphael [MR03, MR04, MR05b, MR05a, Rap05].

THÉORÈME 1.9. — *On suppose $d = 1$. Il existe des constantes $\alpha > 0, C_1, C_2, C_3 > 0$ telles que si $u_0 \in \mathcal{B}_\alpha$ et si $u(t)$ est la solution de (1) de donnée initiale u_0 (qui existe sur un intervalle maximal $t \in [0, T[, T \leq +\infty)$, alors*

1. si

$$E_0^G = E_0(u_0) - \frac{(\mathbf{Im}(\int \nabla u_0 \overline{u_0} dx))^2}{2\|u_0\|_{L^2(\mathbb{R}^d)}^2} < 0,$$

alors $u(t)$ explose en temps fini ($T < +\infty$) et on a la borne suivante sur la vitesse d’explosion :

$$(8) \quad \|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \leq C_1 \left(\frac{\log |\log(T-t)|}{T-t} \right)^{\frac{1}{2}}.$$

2. Si $E_0^G = 0$ et si u_0 n’est pas un soliton (à translation, dilatation et changement d’échelle près) alors $u(t)$ explose en temps fini.
3. L’ensemble \mathcal{O} des données initiales $u_0 \in \mathcal{B}_\alpha$ telles que $u(t)$ explose en temps fini avec la borne supérieure (8) est ouvert dans H^1 .

4. Si u explose en temps fini et si (8) n'est pas vérifiée, alors

$$\|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \geq \frac{C_2}{(T-t)\sqrt{E_0^G}}$$

5. Si u explose en temps fini, alors

$$(9) \quad \|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \geq C_3 \left(\frac{\log |\log(T-t)|}{T-t} \right)^{\frac{1}{2}}.$$

Le résultat précédent se généralise au cas des dimensions d'espace $d = 2, 3, 4$ sous l'hypothèse (conjecturale) :

HYPOTHÈSE SPECTRALE 1.10. — On considère les deux opérateurs de Schrödinger suivants :

$$\mathcal{L}_1 = -\Delta + \frac{2}{d} \left(\frac{4}{d} + 1 \right) Q^{\frac{4}{d}-1} y \cdot \nabla Q, \quad \mathcal{L}_2 = -\Delta + \frac{2}{d} Q^{\frac{4}{d}-1} y \cdot \nabla Q.$$

On note

$$H(\varepsilon, \varepsilon) = (\mathcal{L}_1 \mathbf{Re}\varepsilon, \mathbf{Re}\varepsilon)_{L^2(\mathbb{R}^d)} + (\mathcal{L}_2 \mathbf{Im}\varepsilon, \mathbf{Im}\varepsilon)_{L^2(\mathbb{R}^d)}, \quad \varepsilon \in H^1,$$

et pour toute distribution f ,

$$f_1 = \frac{d}{2} f + y \cdot \nabla f, \quad f_2 = (f_1)_1.$$

Alors il existe deux constantes $\delta > 0$ et $\kappa < 2$ telles que pour toute fonction $\varepsilon \in H^1$ vérifiant les conditions d'orthogonalité

$$(10) \quad \begin{aligned} (\mathbf{Re}\varepsilon, Q)_{L^2(\mathbb{R}^d)} &= (\mathbf{Re}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)} = (\mathbf{Re}\varepsilon, yQ)_{L^2(\mathbb{R}^d)} = 0, \\ (\mathbf{Im}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)} &= (\mathbf{Im}\varepsilon, Q_2)_{L^2(\mathbb{R}^d)} = (\mathbf{Im}\varepsilon, \nabla Q)_{L^2(\mathbb{R}^d)} = 0, \end{aligned}$$

$$H(\varepsilon, \varepsilon) \geq \delta \left(\int |\nabla \varepsilon|^2 dx + \int |\varepsilon|^2 e^{-\kappa|x|} dx \right).$$

(On peut remarquer qu'à partir de la dimension 3, le second terme à droite dans l'inégalité précédente est contrôlé par le premier.)

Cette propriété spectrale des opérateurs \mathcal{L}_j a été démontrée (voir [MR05a]) en dimension $d = 1$ ($\kappa = \frac{9}{5}$) et vérifiée numériquement par Fibich, Merle et Raphael [FMR04] pour $d = 2, 3, 4$. Elle est fautive pour $d = 5, 6$.

1.6. Instabilité structurelle de la loi du « log log »

Même si, compte tenu des simulations numériques de Landman, Papanicolaou, Sulem et Sulem [LPSS88] et plus récemment des résultats du théorème 1.9, et en particulier du point (3), l'explosion au régime du « log log » présente de manifestes propriétés de

stabilité, ce phénomène est lié à une certaine dégénérescence de l'équation de Schrödinger non linéaire (1) et est en fait *structurellement instable* au sens suivant : Considérons en dimension $d = 2$ le système de Zakharov

$$(11) \quad \begin{cases} i\partial_t u + \Delta u - nu = 0, \\ \frac{1}{c_0^2} \partial_t^2 n - \Delta n - \Delta |u|^2 = 0. \end{cases}$$

L'équation (1) est en fait une approximation de ce système dans la limite $c_0 \rightarrow +\infty$ (au moins au niveau formel). Le système de Zakharov possède également une loi du viriel similaire à (6) et on peut en déduire l'explosion pour une classe de données initiales (voir les travaux de Merle [Mer96a]). On peut également exhiber une famille de solutions explosives explosant comme $S(t)$ [GM94b, GM94a]. Cependant, l'analogie avec (1) s'arrête là : Merle [Mer96b] a en effet démontré que toutes les solutions explosives explosent au moins en $1/(T - t)$:

$$\|\nabla u(t)\|_{L^2(\mathbb{R}^d)} \geq \frac{C}{T - t}.$$

En résumé, le régime du « log log » n'existe pas pour le système de Zakharov et il est donc conséquence d'une certaine dégénérescence dans la limite $c_0 \rightarrow +\infty$.

L'objet de cet exposé est de présenter quelques-unes des idées permettant de démontrer les bornes supérieure (8) et inférieure (9) dans le cas d'énergie négative $E_0^G < 0$.

2. MODULATION AUTOUR DE L'ÉTAT FONDAMENTAL

Une première remarque est que l'hypothèse $E_0^G < 0$ implique que, quitte à faire une transformation galiléenne, on peut supposer que

$$(12) \quad E(u) < 0, \quad \mathbf{Im} \left(\int \nabla u \bar{u} \right) = 0.$$

En effet, si u_0 vérifie $E_G(u_0) < 0$, alors si

$$(u_0)_\beta = u_0 e^{i\frac{\beta}{2} \cdot x}, \quad \beta = -2 \frac{\mathbf{Im} \left(\int \nabla u_0 \bar{u}_0 \right)}{\|u_0\|_{L^2(\mathbb{R}^d)}^2},$$

un calcul direct montre que la solution de (1) de donnée initiale $(u_0)_\beta$ vérifie (12). D'autre part, l'invariance galiléenne montre que la description de l'explosion de u_β est équivalente à la description de l'explosion de u . On supposera donc dans la suite que (12) est vérifiée. L'analyse que Merle et Raphael conduisent est de type perturbative (autour de l'état fondamental). La première étape consiste donc à rendre rigoureux le fait que la solution u est proche d'une certaine modulation de cet état fondamental. Dans le contexte de l'équation de Schrödinger, cette approche remonte aux travaux de Weinstein [Wei83, Wei85] (voir aussi Buslaev et Perelman [BP93, BP95]). Le premier résultat est une propriété de stabilité orbitale de l'état fondamental.

PROPOSITION 2.1. — Il existe $\alpha^* > 0$ et une fonction $\delta : \alpha \in]0, \alpha^*] \rightarrow \mathbb{R}^+$ qui tend vers 0 quand α tend vers 0 et tels que pour toute fonction $v \in H^1$ vérifiant

$$(13) \quad \int Q^2 dx \leq \int |v|^2 dx \leq \int Q^2 dx + \alpha,$$

et

$$(14) \quad E(v) \leq \alpha \int |\nabla v|^2 dx,$$

il existe $(\gamma_0, x_0) \in \mathbb{R}^{1+d}$ tels que

$$\|Q - e^{i\gamma_0} \lambda_0^{\frac{d}{2}} v(\lambda_0 x + x_0)\|_{H^1} \leq \delta(\alpha), \quad \lambda_0 = \frac{\|\nabla Q\|_{L^2(\mathbb{R}^d)}}{\|\nabla v\|_{L^2(\mathbb{R}^d)}}.$$

La preuve de cette proposition est assez classique : on fait un raisonnement par l'absurde, un argument de concentration-compacité [Lio84] permet d'extraire une sous-suite convergente qui est, par passage à la limite, d'énergie négative, de masse critique (égale à celle de l'état fondamental), mais n'est pas une translatée-dilatée de l'état fondamental, ce qui contredit la proposition 1.3

On peut remarquer que l'hypothèse (13) est la plus contraignante, c'est elle qui limite l'analyse à un régime de type « perturbation », tandis que l'hypothèse (14) est automatiquement vérifiée pour tout temps pour $E_0 < 0$ ou, si on sait *a priori* que la solution explose, pour t assez proche du temps d'explosion T (puisqu'alors l'hypothèse est vérifiée car le terme de gauche est constant tandis que celui de droite tend vers $+\infty$). Cette hypothèse (13) peut être relaxée (voir [MR05c]).

On va maintenant préciser cette décomposition. L'idée générale est d'utiliser un profil auto-similaire pour décrire la solution. Cependant, comme ces profils ne sont pas de norme L^2 finie, il est nécessaire de les tronquer au bon endroit. Pour cela, on introduit des profils auto-similaires proches de l'état fondamental Q : on fixe un paramètre $0 < \eta < 1$ (qui sera choisi petit par la suite) et pour $b \neq 0$ on définit

$$R_b = \frac{2}{|b|} \sqrt{1 + \eta}, \quad R_b^- = \sqrt{1 - \eta} \frac{2}{|b|}.$$

On fixe aussi une troncature radiale ϕ_b égale à 0 pour $|x| \geq R_b$ et à 1 pour $|x| \leq R_b^-$.

PROPOSITION 2.2. — Il existe $C, \eta^* > 0$ et des fonctions $\epsilon^*(\eta) > 0, b^*(\eta) > 0$ définies pour $0 < |\eta| \leq \eta^*$ et tendant vers 0 quand η tend vers 0 tels que pour tout $b < b^*(\eta)$, il existe une unique solution radiale de l'équation

$$(15) \quad \begin{aligned} \Delta Q_b - Q_b + ib \left(\frac{d}{2} Q_b + y \cdot \nabla Q_b \right) + Q + b|Q_b|^{\frac{4}{d}} &= 0, \\ P_b = Q_b e^{i\frac{b|x|^2}{4}} &> 0 \text{ si } |x| \leq R_b, \\ Q_b(0) \in [Q(0) - \epsilon^*(\eta), Q(0) + \epsilon^*(\eta)], \quad Q_b(R_b) &= 0. \end{aligned}$$

De plus, la fonction $\tilde{Q}_b = Q_b \phi_b(x)$ est solution de

$$(16) \quad \Delta \tilde{Q}_b - \tilde{Q}_b + ib \left(\frac{d}{2} \tilde{Q}_b + y \cdot \nabla \tilde{Q}_b \right) + \tilde{Q}_b + b |\tilde{Q}_b|^{\frac{4}{d}} = \Psi_b,$$

où Ψ_b vérifie

$$\forall \alpha \in \mathbb{N}, \forall \beta \in \mathbb{N}^d, |\beta| \leq 1, \quad \sup_x |x|^\alpha \partial_x^\beta |\Psi_b(x)| \leq e^{-\frac{c}{b}}.$$

On peut maintenant introduire une décomposition géométrique de la solution prenant en compte sa proximité par rapport à la variété de dimension $d + 3$ des fonctions de la forme

$$e^{i\gamma} \lambda^{\frac{d}{2}} \tilde{Q}_b(\lambda x + x_0),$$

où (γ, λ, b, x) sont dans un petit voisinage de $(\gamma_0, 0, 0, x_0)$.

PROPOSITION 2.3. — *Il existe des fonctions C^1 :*

$$(\lambda, \gamma, x, b) : [t_0, T[\rightarrow]0, +\infty[\times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$$

telles que pour tout $t_0 \leq t < T$, si on note

$$\varepsilon(t, y) = e^{i\gamma(t)} \lambda^{\frac{d}{2}}(t) u(t, \lambda(t)y + x(t)) - \tilde{Q}_{b(t)}(y)$$

alors $\varepsilon(t)$ vérifie les conditions d'orthogonalité suivantes :

$$(17) \quad \begin{aligned} & \left(\mathbf{Re} \varepsilon, |y|^2 \mathbf{Re} \tilde{Q}_b \right)_{L^2(\mathbb{R}^d)} + \left(\mathbf{Im} \varepsilon, |y|^2 \mathbf{Im} \tilde{Q}_b \right)_{L^2(\mathbb{R}^d)} = 0 \\ & \left(\mathbf{Re} \varepsilon, y \mathbf{Re} \tilde{Q}_b \right)_{L^2(\mathbb{R}^d)} + \left(\mathbf{Im} \varepsilon, y \mathbf{Im} \tilde{Q}_b \right)_{L^2(\mathbb{R}^d)} = 0 \\ & \left(\mathbf{Re} \varepsilon, (\mathbf{Im} \tilde{Q}_b)_1 \right)_{L^2(\mathbb{R}^d)} - \left(\mathbf{Im} \varepsilon, (\mathbf{Re} \tilde{Q}_b)_1 \right)_{L^2(\mathbb{R}^d)} = 0 \\ & \left(\mathbf{Re} \varepsilon, (\mathbf{Im} \tilde{Q}_b)_2 \right)_{L^2(\mathbb{R}^d)} - \left(\mathbf{Im} \varepsilon, (\mathbf{Re} \tilde{Q}_b)_2 \right)_{L^2(\mathbb{R}^d)} = 0 \end{aligned}$$

et de plus

$$(18) \quad \left| 1 - \lambda(t) \frac{\|\nabla u(t)\|_{L^2(\mathbb{R}^d)}}{\|\nabla Q\|_{L^2(\mathbb{R}^d)}} \right| + \|\varepsilon(t)\|_{H^1} + |b(t)| \leq \delta(\alpha^*), \quad \lim_{\alpha^* \rightarrow 0} \delta(\alpha^*) = 0.$$

(On rappelle que $(f)_1 = (\frac{d}{2}f + y \cdot \nabla f)$ et $(f)_2 = ((f)_1)_1$.)

Compte tenu de la proposition 2.1, la preuve de ce résultat repose essentiellement sur le théorème des fonctions implicites. On peut remarquer que si on fait tendre b vers 0 (formellement) dans les conditions d'orthogonalités (17), en utilisant que $\tilde{Q}_b \rightarrow Q$, on obtient

$$(\mathbf{Re} \varepsilon, y^2 Q) = (\mathbf{Re} \varepsilon, y^2 Q) = (\mathbf{Im} \varepsilon, Q_1) = (\mathbf{Im} \varepsilon, Q_2) = 0,$$

et que les trois dernières relations qu'on obtient apparaissent dans l'hypothèse spectrale.

3. LA BORNE SUPÉRIEURE

3.1. Les équations en variables dilatées

On introduit le changement de variables

$$s = \int_{t(u_0)}^t \frac{dt'}{\lambda^2(t')}.$$

Compte tenu de (7), $s(T) = +\infty$.

On peut maintenant écrire les lois de conservation de l'énergie et du moment linéarisées autour de \tilde{Q}_b . On obtient :

LEMME 3.1. —

(19)

$$\begin{aligned} |2(\mathbf{Re}\varepsilon, \mathbf{Re}\tilde{Q}_b) + 2(\mathbf{Im}\varepsilon, \mathbf{Im}\tilde{Q}_b)| &\leq C \left(\int |\nabla\varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} \right) + e^{-\frac{(1-C\eta)\pi}{b}} + C\lambda^2|E_0| \\ |(\mathbf{Im}\varepsilon, \nabla\mathbf{Re}\tilde{Q}_b)| &\leq C \left(\int |\nabla\varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} \right)^{1/2}. \end{aligned}$$

On remarquera que le facteur exponentiel dans la première équation vient du calcul explicite de l'énergie de \tilde{Q}_b .

On peut également écrire les équations vérifiées par les paramètres de modulation (λ, γ, x, b) et ε . En particulier (avec $\tilde{\gamma} = -s - \gamma$),

$$\begin{aligned} (20) \quad \frac{\partial}{\partial s} b \partial_b (\mathbf{Re}\tilde{Q}_b) + \frac{\partial}{\partial s} (\mathbf{Re}\varepsilon) - M_-(\varepsilon) + b \left(\frac{d}{2} \mathbf{Re}\varepsilon + y \cdot \nabla \mathbf{Re}\varepsilon \right) \\ = \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) (\mathbf{Re}\tilde{Q}_b)_1 + \frac{\partial}{\partial s} \tilde{\gamma} \mathbf{Im}\tilde{Q}_b + \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \nabla (\mathbf{Re}\tilde{Q}_b) \\ + \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) \left(\frac{d}{2} \mathbf{Re}\varepsilon + y \cdot \nabla \mathbf{Re}\varepsilon \right) + \frac{\partial}{\partial s} \tilde{\gamma} \mathbf{Im}\varepsilon + \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \nabla (\mathbf{Re}\varepsilon)_1 + \mathbf{Im}(\Psi_b) - R_1(\varepsilon) \end{aligned}$$

$$\begin{aligned} (21) \quad \frac{\partial}{\partial s} b \partial_b (\mathbf{Im}\tilde{Q}_b) + \frac{\partial}{\partial s} (\mathbf{Im}\varepsilon) + M_+(\varepsilon) + b \left(\frac{d}{2} \mathbf{Im}\varepsilon + y \cdot \nabla \mathbf{Im}\varepsilon \right) \\ = \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) (\mathbf{Im}\tilde{Q}_b)_1 - \frac{\partial}{\partial s} \tilde{\gamma} \mathbf{Re}\tilde{Q}_b + \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \nabla (\mathbf{Im}\tilde{Q}_b) \\ + \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) \left(\frac{d}{2} \mathbf{Im}\varepsilon + y \cdot \nabla \mathbf{Im}\varepsilon \right) - \frac{\partial}{\partial s} \tilde{\gamma} \mathbf{Re}\varepsilon + \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \nabla (\mathbf{Im}\varepsilon)_1 + \mathbf{Im}(\Psi_b) - R_2(\varepsilon) \end{aligned}$$

où

$$M_+(\varepsilon) = -\Delta \mathbf{Re}\varepsilon + \mathbf{Re}\varepsilon - \left(\frac{4(\mathbf{Re}\tilde{Q}_b)^2}{d|\tilde{Q}_b|^2} + 1 \right) |\tilde{Q}_b|^{\frac{4}{d}} \mathbf{Re}\varepsilon - \left(\frac{4(\mathbf{Re}\tilde{Q}_b \mathbf{Im}\tilde{Q}_b)}{d|\tilde{Q}_b|^2} \right) |\tilde{Q}_b|^{\frac{4}{d}} \mathbf{Im}\varepsilon,$$

$$M_-(\varepsilon) = -\Delta \mathbf{Im} \varepsilon + \mathbf{Im} \varepsilon - \left(\frac{4(\mathbf{Re} \tilde{Q}_b)^2}{d|\tilde{Q}_b|^2} + 1 \right) |\tilde{Q}_b|^{\frac{4}{d}} \mathbf{Im} \varepsilon - \left(\frac{4(\mathbf{Re} \tilde{Q}_b \mathbf{Im} \tilde{Q}_b)}{d|\tilde{Q}_b|^2} \right) |\tilde{Q}_b|^{\frac{4}{d}} \mathbf{Re} \varepsilon,$$

et les termes $R_{1;2}(\varepsilon)$ sont des termes de reste (d'ordre supérieur en ε). Si on fait le produit scalaire de ces équations avec les diverses quantités apparaissant dans les relations d'orthogonalité (17), on obtient

LEMME 3.2. —

$$(22) \quad \left| \frac{\partial \lambda}{\partial s} + b \right| + \left| \frac{\partial}{\partial s} b \right| \leq C \left(\int |\nabla \varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} \right) + e^{-\frac{(1-C\eta)\pi}{b}} + C\lambda^2 |E_0|$$

$$(23) \quad \left| \frac{\partial}{\partial s} \tilde{\gamma} - \frac{(\mathbf{Re} \varepsilon, L_+ Q_2)}{\|Q_1\|_{L^2(\mathbb{R}^d)}^2} \right| + \left| \frac{\partial x}{\partial s} \right| \leq \delta(\alpha^*) \left(\int |\nabla \varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} \right)^{1/2} \\ + \int |\nabla \varepsilon|^2 + e^{-\frac{(1-C\eta)\pi}{b}} + C\lambda^2 |E_0|$$

où

$$L_+ = -\Delta + 1 - \left(1 + \frac{4}{d}\right) Q^{\frac{4}{d}}, \quad \lim_{\alpha^* \rightarrow 0} \delta(\alpha^*) = 0.$$

3.2. L'inégalité du viriel locale

Le point clef de l'analyse de Merle et Raphael consiste à localiser l'inégalité du viriel (6). Le point de départ est la remarque suivante : d'après (6), on a formellement (en intégrant par parties)

$$(24) \quad 4E_0 t + c_0 = \mathbf{Im} \left(\int x \nabla u \bar{u} dy \right) \\ = \mathbf{Im} \left(\int x \nabla \tilde{Q}_b \overline{\tilde{Q}_b} dy \right) + \mathbf{Im} \left(\int y \cdot \nabla \varepsilon \bar{\varepsilon} dy \right) \\ + 2(\mathbf{Re} \varepsilon, (\mathbf{Im} \tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} - 2(\mathbf{Im} \varepsilon, (\mathbf{Re} \tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)}.$$

Mais le choix de conditions d'orthogonalité (17) garantit que le terme linéaire en ε disparaît. On peut par ailleurs calculer explicitement

$$\mathbf{Im} \left(\int y \cdot \nabla \tilde{Q}_b \overline{\tilde{Q}_b} dy \right) = -\frac{b}{2} \|y \tilde{Q}_b\|_{L^2(\mathbb{R}^d)} \sim -Cb;$$

on obtient alors

$$4E_0 t + c_0 \sim -Cb + \mathbf{Im} \left(\int y \cdot \nabla \varepsilon \bar{\varepsilon} dy \right).$$

On repasse en variables s et on dérive l'identité obtenue par rapport à s . On obtient

$$4\lambda^2 E_0 + C \frac{\partial}{\partial s} b = \frac{d}{ds} \mathbf{Im} \left(\int y \cdot \nabla \varepsilon \bar{\varepsilon} dy \right).$$

En d'autres termes, pour calculer l'expression $\frac{d}{ds} \mathbf{Im} \left(\int y \cdot \nabla \varepsilon \bar{\varepsilon} dy \right)$ qui n'a aucun sens (car ε est seulement dans H^1), il suffit de calculer $4\lambda^2 E_0 + C \frac{\partial}{\partial s} b$. C'est exactement le calcul que Merle et Raphael font :

PROPOSITION 3.3. — Il existe $\delta_0, C > 0$ tels que

$$(25) \quad \frac{\partial}{\partial s} b \geq \delta_0 \left(\int |\nabla \varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\delta|y|} dy \right) - C\lambda^2 E_0 - e^{-\frac{C}{b}}.$$

Pour démontrer cette proposition, on écrit l'équation vérifiée par $\frac{\partial}{\partial s} b$, et après quelques calculs algébriques, on obtient

$$(26) \quad \begin{aligned} \frac{\partial}{\partial s} b & \left((\partial_b(\mathbf{Im}\tilde{Q}_b), (\mathbf{Re}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} - (\partial_b(\mathbf{Re}\tilde{Q}_b), (\mathbf{Im}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} \right. \\ & \left. - (\mathbf{Im}\varepsilon, (\partial_b(\mathbf{Re}\tilde{Q}_b)))_{L^2(\mathbb{R}^d)} + (\mathbf{Re}\varepsilon, (\partial_b(\mathbf{Im}\tilde{Q}_b)))_{L^2(\mathbb{R}^d)} \right) \\ & = H(\varepsilon, \varepsilon) - 2\lambda^2 E_0 \\ & \quad - \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \left((\mathbf{Im}\varepsilon, \nabla(\mathbf{Re}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Re}\varepsilon, \nabla(\mathbf{Im}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} \right) \\ & \quad - \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) \left((\mathbf{Im}\varepsilon, (\mathbf{Re}\tilde{Q}_b)_2)_{L^2(\mathbb{R}^d)} - (\mathbf{Re}\varepsilon, (\mathbf{Im}\tilde{Q}_b)_2)_{L^2(\mathbb{R}^d)} \right) \\ & \quad - \frac{\partial}{\partial s} \tilde{\gamma} \left((\mathbf{Re}\varepsilon, (\mathbf{Re}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} + (\mathbf{Im}\varepsilon, (\mathbf{Im}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} \right) \\ & \quad - (\mathbf{Re}\varepsilon, (\mathbf{Re}\Psi_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im}\varepsilon, (\mathbf{Im}\Psi_b)_1)_{L^2(\mathbb{R}^d)} + R(\varepsilon) \end{aligned}$$

où $R(\varepsilon)$ est un terme de reste contenant tous les termes d'ordre supérieur en ε qui est contrôlé par

$$|R(\varepsilon)| \leq \delta(\alpha^*) \left(\int |\nabla \varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\delta|y|} dy \right).$$

La contribution de

$$-(\mathbf{Re}\varepsilon, (\mathbf{Re}\Psi_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im}\varepsilon, (\mathbf{Im}\Psi_b)_1)_{L^2(\mathbb{R}^d)}$$

est, compte tenu des propriétés de décroissance de la fonction Ψ_b , bornée par

$$\epsilon \int |\varepsilon|^2 e^{-\kappa|y|} dy + C e^{-\frac{C}{b}}.$$

Les conditions d'orthogonalité (17) annulent le terme

$$- \left(\frac{\frac{\partial}{\partial s} \lambda}{\lambda} + b \right) \left((\mathbf{Im}\varepsilon, (\mathbf{Re}\tilde{Q}_b)_2)_{L^2(\mathbb{R}^d)} - (\mathbf{Re}\varepsilon, (\mathbf{Im}\tilde{Q}_b)_2)_{L^2(\mathbb{R}^d)} \right).$$

On peut ensuite, en remarquant que $Q_b = Q + o(1)$ ($b \rightarrow 0$) et en utilisant (23), remplacer le terme

$$\frac{\partial}{\partial s} \tilde{\gamma} \left((\mathbf{Re}\varepsilon, (\mathbf{Re}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} + (\mathbf{Im}\varepsilon, (\mathbf{Im}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} \right)$$

par

$$\frac{(\mathbf{Re}\varepsilon, L_+ Q_2)}{\|Q_1\|_{L^2(\mathbb{R}^d)}^2} (\mathbf{Re}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)} + \mathcal{O} \left(\delta(\alpha^*) \int |\nabla \varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} + e^{-\frac{C}{b}} \right).$$

D'autre part, en réutilisant (23), on a

$$(27) \quad \left| \frac{\partial x}{\partial s} \cdot \left((\mathbf{Im}\varepsilon, \nabla(\mathbf{Re}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Re}\varepsilon, \nabla(\mathbf{Im}\tilde{Q}_b)_1)_{L^2(\mathbb{R}^d)} \right) \right| \leq \delta(\alpha^*) \left(\int |\nabla\varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\delta|y|} dy \right).$$

On obtient ainsi

$$(28) \quad \frac{\partial}{\partial s} b \geq H(\varepsilon, \varepsilon) - \frac{1}{\|Q_1\|_{L^2(\mathbb{R}^d)}^2} (\mathbf{Re}\varepsilon, L_+ Q_2)_{L^2(\mathbb{R}^d)} (\mathbf{Re}\varepsilon, Q_1) - \delta(\alpha^*) \int |\nabla\varepsilon|^2 + |\varepsilon|^2 e^{-\kappa|y|} - e^{-\frac{c}{b}}.$$

On remarque ensuite que

$$H(\varepsilon, \varepsilon) - \frac{1}{\|Q_1\|_{L^2(\mathbb{R}^d)}^2} (\mathbf{Re}\varepsilon, L_+ Q_2)_{L^2(\mathbb{R}^d)} (\mathbf{Re}\varepsilon, Q_1) = H(\tilde{\varepsilon}, \tilde{\varepsilon})$$

où $\tilde{\varepsilon} = \varepsilon - \frac{(\mathbf{Re}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)}}{\|Q_1\|_{L^2(\mathbb{R}^d)}^2} Q_1$. Mais d'après l'hypothèse spectrale,

$$(29) \quad \begin{aligned} H(\tilde{\varepsilon}, \tilde{\varepsilon}) &\geq \delta \left(\int |\nabla\tilde{\varepsilon}|^2 dy + \int |\tilde{\varepsilon}|^2 e^{-\kappa|y|} dy \right) - (\mathbf{Re}\tilde{\varepsilon}, Q)_{L^2(\mathbb{R}^d)}^2 \\ &\quad - (\mathbf{Re}\tilde{\varepsilon}, yQ)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\tilde{\varepsilon}, Q_1)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\tilde{\varepsilon}, Q_2)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\tilde{\varepsilon}, \nabla Q)_{L^2(\mathbb{R}^d)}^2 \\ &\geq \frac{\delta}{2} \left(\int |\nabla\varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\kappa|y|} dy \right) - C(\mathbf{Re}\varepsilon, Q)_{L^2(\mathbb{R}^d)}^2 \\ &\quad - (\mathbf{Re}\varepsilon, yQ)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\varepsilon, Q_2)_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\varepsilon, \nabla Q)_{L^2(\mathbb{R}^d)}^2. \end{aligned}$$

Il reste, pour conclure, à vérifier qu'on contrôle les termes négatifs dans le terme de droite de l'équation précédente. Les directions

$$(\mathbf{Re}\varepsilon, yQ)_{L^2(\mathbb{R}^d)}^2, (\mathbf{Im}\varepsilon, Q_1)_{L^2(\mathbb{R}^d)}^2 \text{ et } (\mathbf{Im}\varepsilon, Q_2)_{L^2(\mathbb{R}^d)}^2$$

sont traitées grâce aux conditions d'orthogonalité (17). Par exemple :

$$(30) \quad \begin{aligned} (\mathbf{Re}\varepsilon, yQ)_{L^2(\mathbb{R}^d)} &= (\mathbf{Re}\varepsilon, y\mathbf{Re}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)} + (\mathbf{Im}\varepsilon, y\mathbf{Im}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)} \\ &\quad + (\mathbf{Re}\varepsilon, Q_1 - \mathbf{Re}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\varepsilon, y\mathbf{Im}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)} \end{aligned}$$

la condition d'orthogonalité annule le premier terme tandis que (puisque \tilde{Q}_b est proche de Q) le second est contrôlé, ce qui donne

$$(31) \quad \begin{aligned} |(\mathbf{Re}\varepsilon, yQ)_{L^2(\mathbb{R}^d)}| &\leq \left| (\mathbf{Re}\varepsilon, Q_1 - \mathbf{Re}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)}^2 - (\mathbf{Im}\varepsilon, y\mathbf{Im}(\tilde{Q}_b))_{L^2(\mathbb{R}^d)} \right| \\ &\leq \delta(\alpha^*) \left(\int |\nabla\varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\kappa|x|} dy \right) \end{aligned}$$

(rappelons que α^* petit implique b petit aussi d'après (18)).

Les deux autres termes, $(\mathbf{Re}\varepsilon, Q)_{L^2}^2$ et $(\mathbf{Im}\varepsilon, \nabla Q)_{L^2}^2$, sont estimés en utilisant (19) (et le fait que Q est proche de \tilde{Q}_b).

3.3. Une quasi-fonctionnelle de Lyapunov

On montre dans cette partie comment on déduit de l'inégalité différentielle (25), vérifiée par b , une information sur la dynamique. On commence par remarquer que la fonction b ne peut pas passer du signe $+$ au signe $-$. En effet s'il existait un point où $b(s_0) = 0$, $\frac{\partial}{\partial s}b(s_0) \leq 0$, d'après (25), on aurait $\varepsilon = 0$ ce qui contredit l'hypothèse d'énergie négative. D'après (22) et en utilisant (25), on obtient

$$\left| \frac{\frac{\partial}{\partial s}\lambda}{\lambda} + b \right| \leq C \frac{\partial}{\partial s}b + e^{-\frac{c}{|b|}}.$$

On intègre (et on utilise que b est petit et de signe fixe pour $s \geq s_0$ dans la dernière inégalité)

$$(32) \quad \left| \log \left(\frac{\lambda(s)}{\lambda(s_0)} \right) + \int_{s_0}^s b(s) ds \right| \leq C(b(s) - b(s_0)) + \int_{s_0}^s e^{-\frac{c}{|b|}} ds \leq \frac{1}{4} + \frac{1}{2} \left| \int_{s_0}^s b(s) ds \right|.$$

On en déduit que $\int_{s_0}^{+\infty} b(s) = +\infty$ (et donc que b est strictement positive pour s assez grand). En effet, sinon, $|\log(\lambda(s))|$ serait borné inférieurement, donc $\lambda(s)$ serait minoré par une constante $\lambda_0 > 0$, ce qui, compte tenu de (25), impliquerait

$$\frac{\partial}{\partial s}b \geq -\lambda^2 E_0 - e^{-\frac{c}{|b|}} \geq -\frac{1}{2}\lambda^2 E_0$$

et donc contredirait que b est borné (et même petit d'après (18)).

On peut maintenant intégrer l'inéquation

$$\frac{\partial}{\partial s}b \geq e^{-\frac{c}{b}} \Rightarrow -\frac{\frac{\partial}{\partial s}b}{b^2} e^{\frac{c}{b}} \leq 1 \Rightarrow b(s) \geq \frac{C}{\log(s)}.$$

Si on revient à (32), on obtient

$$|\log(\lambda)| \geq C \int_{s_0}^s \frac{ds}{\log(s)} \geq C s^{\frac{1}{2}} \Rightarrow \log |\log(\lambda)| \geq C \log(s).$$

On en déduit d'abord qu'il y a bien explosion en temps fini

$$|\lambda(s)| \leq e^{-s^\alpha} \Rightarrow T = \int_0^{+\infty} \lambda^2(s) ds < +\infty$$

puis en réutilisant (32)

$$(33) \quad \begin{aligned} \frac{T - t(s)}{\lambda^2(s)} &= \int_s^{+\infty} \frac{\lambda^2(\sigma)}{\lambda^2(s)} d\sigma \\ &\leq \int_s^{+\infty} e^{\frac{1}{4} - \frac{1}{2} \int_s^\sigma b(\tau) d\tau} d\sigma \\ &\leq C \log(s) \leq C \log(|\log(\lambda)|) \end{aligned}$$

où on a utilisé la minoration $b(\tau) \geq \frac{1}{\log(\tau)}$ pour obtenir l'avant-dernière inégalité. Finalement on a bien obtenu la majoration annoncée

$$\frac{T - t(s)}{\lambda^2(s) \log(|\log(\lambda)|)} \leq C \Rightarrow \|\nabla u(t(s))\|_{L^2(\mathbb{R}^d)} \sim \lambda^{-1}(s) \leq C' \sqrt{\frac{\log(|\log(T - t(s))|)}{T - t(s)}}.$$

4. LA BORNE INFÉRIEURE

Nous avons obtenu dans la section précédente une inégalité

$$b(s) \geq \frac{C}{\log(s)}$$

et ceci est la clef pour obtenir la borne supérieure. Nous voulons maintenant obtenir l'inégalité inverse. Curieusement, le point de départ est toujours l'inégalité (25), qu'ici nous allons utiliser pour contrôler des termes dans un calcul de flux que nous ferons plus tard.

4.1. Inégalité du viriel précisée

La présence dans l'inégalité (25) du terme $e^{-\frac{C}{b}}$ est due à la présence dans (26) du terme induit par le fait que le profil auto-similaire Q_b a été tronqué pour donner \tilde{Q}_b :

$$(34) \quad -(\mathbf{Re}\varepsilon, (\mathbf{Re}\Psi_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im}\varepsilon, (\mathbf{Im}\Psi_b)_1)_{L^2(\mathbb{R}^d)}.$$

En fait, en raffinant l'analyse précédente, on peut démontrer l'inégalité suivante :

$$(35) \quad \frac{\partial}{\partial s} b \geq \delta_0 \left(\int |\nabla \varepsilon|^2 dy + \int |\varepsilon|^2 e^{-\delta|y|} dy \right) - C\lambda^2 E_0 \\ - e^{-\frac{\pi(1+\varepsilon)}{b}} - (\mathbf{Re}\varepsilon, (\mathbf{Re}\Psi_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im}\varepsilon, (\mathbf{Im}\Psi_b)_1)_{L^2(\mathbb{R}^d)}.$$

Pour faire disparaître le terme (34), il faut raffiner le profil de modulation : pour $|y| \leq 2/|b|$, le profil \tilde{Q}_b (un objet non linéaire) modulé modélise bien la solution, tandis que pour $2/|b| \leq |y| \leq e^{\varepsilon/|b|}$ nous allons introduire la radiation pour modéliser la solution :

LEMME 4.1. — *Il existe une unique solution radiale, ζ_b de l'équation suivante :*

$$(36) \quad \Delta \zeta_b - \zeta_b + ib(\zeta_b)_1 = \Psi_b \\ \|\nabla \zeta_b\|_{L^2(\mathbb{R}^d)} < +\infty.$$

Cette solution vérifie

$$\lim_{|y| \rightarrow +\infty} |y|^d |\zeta_b(y)|^2 = \Gamma_b, \quad e^{-\frac{(1-C\eta)\pi}{b}} \leq \Gamma_b \leq e^{-\frac{(1+C\eta)\pi}{b}}.$$

On peut remarquer que la fonction ζ_b n'est pas dans L^2 à cause d'une divergence de type logarithmique. Il faut donc la tronquer. On fixe donc maintenant $A = e^{\frac{\epsilon}{b}}$ ($\epsilon > 0$ sera choisi petit) et pour $\chi \in C_0^\infty(\mathbb{R})$ égale à 1 pour $|r| \leq 1$, $\chi_A(r) = \chi(r/A)$, et à 0 pour $|r| \geq 2$

$$\tilde{\zeta}_b = \zeta_b \chi_A(r), \quad \tilde{\varepsilon} = \varepsilon - \tilde{\zeta}_b.$$

PROPOSITION 4.2. — *Il existe $\delta_0, C, c > 0$ tels que*

$$(37) \quad \frac{\partial}{\partial s} f \geq \delta_0 \left(\int |\nabla \tilde{\varepsilon}|^2 dy + \int |\tilde{\varepsilon}|^2 e^{-\delta|y|} dy \right) + c\Gamma_b - C\lambda^2 E_0 - C \int_A^{2A} |\varepsilon|^2$$

avec

$$f(s) = \frac{b}{4} \|y \tilde{Q}_b\|_{L^2(\mathbb{R}^d)}^2 + \frac{1}{2} \mathbf{Im} \left(\int y \cdot \nabla \tilde{\zeta}_b \overline{\tilde{\zeta}_b} \right) + (\mathbf{Im} \varepsilon, (\mathbf{Re} \tilde{\zeta}_b)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Re} \varepsilon, (\mathbf{Im} \tilde{\zeta}_b)_1)_{L^2(\mathbb{R}^d)}.$$

Remarque 4.3. — Si on compare les diverses inégalités de viriel obtenues, on peut remarquer que, dans cette dernière, $f(s) \sim b(s)$ (car ζ est exponentiellement petit en b), et d'autre part que le terme exponentiel en b apparaît maintenant avec le bon signe, le prix à payer étant le terme $C \int_A^{2A} |\varepsilon|^2$.

La démonstration de ce résultat est assez proche de celle de la proposition 3.3. Indiquons seulement comment on s'est débarrassé du terme (34). L'introduction de la radiation $\tilde{\zeta}_b$ l'a remplacé (modulo des termes négligeables) par un terme du type

$$(38) \quad -(\mathbf{Re} \varepsilon, (\mathbf{Re} F)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im} \varepsilon, (\mathbf{Im} F)_1)_{L^2(\mathbb{R}^d)}$$

où

$$F = [\Delta, \chi_A] \zeta_b + iby \cdot \chi_A \zeta_b$$

est supporté dans $\{A \leq |y| \leq 2A\}$, dont on déduit

$$\left| (\mathbf{Re} \varepsilon, (\mathbf{Re} F)_1)_{L^2(\mathbb{R}^d)} - (\mathbf{Im} \varepsilon, (\mathbf{Im} F)_1)_{L^2(\mathbb{R}^d)} \right| \leq \epsilon \Gamma_b + \frac{1}{\epsilon} \int_A^{2A} |\varepsilon|^2.$$

4.2. Dispersion L^2 à l'infini

Pour traiter le terme $-\int_A^{2A} |\varepsilon|^2$, il est nécessaire de comprendre comment l'excès de masse de la donnée initiale u_0 est évacué et dispersé à l'infini (en d'autres termes comprendre comment croît la norme L^2 de la restriction de ε à l'ensemble $\{|y| \geq Cste\}$). On choisit ϕ une fonction positive, croissante, égale à 0 pour $r \leq 1/2$, à 1 pour $r \geq 3$ et telle que $\phi'(r) \geq 1/4$ pour $r \in [1, 2]$. On définit $\phi_A(x) = \phi(|x|/A)$. On a alors

PROPOSITION 4.4. — *Il existe $C > 0$ tel que*

$$(39) \quad \frac{d}{ds} \left(\int \phi_A(x) |\varepsilon|^2 \right) \geq \frac{b}{C} \int_A^{2A} |\varepsilon|^2 - \frac{C}{b^2} \lambda^2 E_0 - \Gamma_b^{1+\epsilon} - \Gamma_b^\epsilon \int |\nabla \varepsilon|^2.$$

Pour démontrer ce résultat, on fait le produit scalaire de (22) avec $\phi_A(x)\mathbf{Re}\varepsilon$ et de (23) avec $\phi_A(x)\mathbf{Im}\varepsilon$; et on intègre par parties. On obtient

$$(40) \quad \frac{d}{ds} \left(\int \phi_A |\varepsilon|^2 \right) = \int \frac{\partial}{\partial s} (\phi_A) |\varepsilon|^2 + b \int y \cdot \nabla (\phi_A) |\varepsilon|^2 + 2\mathbf{Im} \left(\int \nabla \phi_A \cdot \nabla \varepsilon \bar{\varepsilon} \right) \\ - \left(\frac{\partial}{\partial s} \lambda \lambda + b \right) \int y \cdot \nabla \phi_A |\varepsilon|^2 - \frac{\frac{\partial}{\partial s} x}{\lambda} \cdot \int \nabla \phi_A |\varepsilon|^2.$$

Le terme principal est

$$(41) \quad b \int y \cdot \nabla (\phi_A) |\varepsilon|^2 \geq \frac{b}{C} \int \phi' \left(\frac{y}{A} \right) |\varepsilon|^2 \geq \frac{b}{4C} \int_A^{2A} |\varepsilon|^2.$$

Les autres termes sont des termes de reste responsables des termes de reste dans la proposition.

On peut maintenant utiliser la conservation de la norme L^2 :

$$(42) \quad \|u_0\|_{L^2(\mathbb{R}^d)}^2 = \|Q\|_{L^2(\mathbb{R}^d)}^2 + \alpha \\ = \|\phi_A^{1/2} \varepsilon\|_{L^2(\mathbb{R}^d)}^2 + \|(1 - \phi_A)^{1/2} \varepsilon\|_{L^2(\mathbb{R}^d)}^2 + \|\tilde{Q}_b\|_{L^2(\mathbb{R}^d)}^2 \\ + 2(\mathbf{Re}\varepsilon, \mathbf{Re}\tilde{Q}_b)_{L^2(\mathbb{R}^d)} + 2(\mathbf{Im}\varepsilon, \mathbf{Im}\tilde{Q}_b)_{L^2(\mathbb{R}^d)}$$

soit

$$(43) \quad \frac{d}{ds} \left(\int \phi_A(x) |\varepsilon|^2 \right) \\ = -\frac{d}{ds} \left(\|Q_b\|_{L^2(\mathbb{R}^d)}^2 - \|Q\|_{L^2(\mathbb{R}^d)}^2 + 2(\mathbf{Re}\varepsilon, \mathbf{Re}\tilde{Q}_b)_{L^2(\mathbb{R}^d)} + 2(\mathbf{Im}\varepsilon, \mathbf{Im}\tilde{Q}_b)_{L^2(\mathbb{R}^d)} \right).$$

Le point important est que

$$\|Q_b\|_{L^2(\mathbb{R}^d)}^2 - \|Q\|_{L^2(\mathbb{R}^d)}^2 + 2(\mathbf{Re}\varepsilon, \mathbf{Re}\tilde{Q}_b)_{L^2(\mathbb{R}^d)} + 2(\mathbf{Im}\varepsilon, \mathbf{Im}\tilde{Q}_b)_{L^2(\mathbb{R}^d)} \sim b^2.$$

On peut donc, en additionnant à l'inéquation que nous venons d'obtenir l'inéquation (37) multipliée par εb obtenir une inéquation du type (avec $C_0 > 0$)

$$\frac{d}{ds} (-C_0 b^2 + \varepsilon b^2) \geq Cb \left(\Gamma_b + \int |\nabla \tilde{\varepsilon}|^2 + |\tilde{\varepsilon}|^2 e^{-\delta|y|} + \int_A^{2A} |\varepsilon|^2 - 2\lambda^2 E_0 \right),$$

ce qui donne une inégalité différentielle sur b du type

$$\frac{d}{ds} (b^2) \leq -Cb \left(\Gamma_b + \int |\nabla \tilde{\varepsilon}|^2 + |\tilde{\varepsilon}|^2 e^{-\delta|y|} + \int_A^{2A} |\varepsilon|^2 - 2\lambda^2 E_0 \right)$$

qui est le point essentiel de la preuve de la borne inférieure (9).

Remarque 4.5. — Dans cette dernière étape, nous avons utilisé de manière importante la conservation de la norme L^2 , ce qui *de facto* élimine la possibilité d'adopter cette approche pour une solution auto-similaire (et pour une bonne raison puisque dans ce cas la borne inférieure est fausse).

Remarque 4.6. — On remarquera aussi que nous avons utilisé le fait que l'énergie est négative, ce qui permet de ne pas se préoccuper du terme $-\lambda^2 E_0$ qui a toujours le bon signe. En fait, dans le régime du « log log », ce terme est négligeable car il a pour ordre de grandeur

$$\lambda^2 \ll e^{-\alpha s^\beta} \ll b \sim \frac{1}{\log(s)}.$$

C'est ce fait qui permet de « bootstrapper » le régime du « log log » (voir le point (3) du théorème 1.9), même pour des énergies positives.

RÉFÉRENCES

- [BP93] V.S. Buslaev and G.S. Perel'man. Scattering for the nonlinear Schrödinger equation : States close to a soliton. *St. Petersburg Math. J.*, 4(6) :1111–1142, 1993.
- [BP95] V.S. Buslaev and G.S. Perel'man. On the stability of solitary waves for nonlinear Schrödinger equations. *American Mathematical Society. Transl., Ser. 2*, 164(22) :75–98, 1995.
- [BW97] J. Bourgain and W. Wang. Construction of blowup solutions for the nonlinear Schrödinger equation with critical nonlinearity. *Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser.*, 25(1-2) :197–215, 1997.
- [FMR04] G. Fibich, F. Merle, and P. Raphael. Numerical proof of a spectral property related to the singularity formation for the L^2 critical non linear schrödinger equation. Preprint, 2004.
- [GM94a] L. Glangetas and F. Merle. Concentration properties of blow-up solutions and instability results for Zakharov equation in dimension two. II. *Comm. Math. Phys.*, 160(2) :349–389, 1994.
- [GM94b] L. Glangetas and F. Merle. Existence of self-similar blow-up solutions for Zakharov equation in dimension two. I. *Comm. Math. Phys.*, 160(1) :173–215, 1994.
- [GV79] J. Ginibre and G. Velo. On a class of nonlinear Schrödinger equations. I. The Cauchy problem, general case. *J. Funct. Anal.*, 32(1) :1–32, 1979.
- [JP93] R. Johnson and X. B. Pan. On an elliptic equation related to the blow-up phenomenon in the nonlinear Schrödinger equation. *Proc. Roy. Soc. Edinburgh Sect. A*, 123(4) :763–782, 1993.
- [Lio84] P.-L. Lions. The concentration-compactness principle in the calculus of variations. The locally compact case. I. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 1(2) :109–145, 1984.

- [LPSS88] M. J. Landman, G. C. Papanicolaou, C. Sulem, and P.-L. Sulem. Rate of blow-up for solutions of the nonlinear Schrödinger equation at critical dimension. *Phys. Rev. A (3)*, 38(8) :3837–3843, 1988.
- [Mer93] F. Merle. Determination of blow-up solutions with minimal mass for nonlinear Schrödinger equations with critical power. *Duke Math. J.*, 69(2) :427–454, 1993.
- [Mer96a] F. Merle. Blow-up results of virial type for Zakharov equations. *Comm. Math. Phys.*, 175(2) :433–455, 1996.
- [Mer96b] F. Merle. Lower bounds for the blowup rate of solutions of the Zakharov equation in dimension two. *Comm. Pure Appl. Math.*, 49(8) :765–794, 1996.
- [MM00] Y. Martel and F. Merle. A Liouville theorem for the critical generalized Korteweg-de Vries equation. *J. Math. Pures Appl. (9)*, 79(4) :339–425, 2000.
- [MM02a] Y. Martel and F. Merle. Blow up in finite time and dynamics of blow up solutions for the L^2 -critical generalized KdV equation. *J. Amer. Math. Soc.*, 15(3) :617–664 (electronic), 2002.
- [MM02b] Y. Martel and F. Merle. Stability of blow-up profile and lower bounds for blow-up rate for the critical generalized KdV equation. *Ann. of Math. (2)*, 155(1) :235–280, 2002.
- [MM04] Y. Martel and F. Merle. Review on blow up and asymptotic dynamics for critical and subcritical gKdV equations. In *Noncompact problems at the intersection of geometry, analysis, and topology*, volume 350 of *Contemp. Math.*, pages 157–177. Amer. Math. Soc., Providence, RI, 2004.
- [MR03] F. Merle and P. Raphael. Sharp upper bound on the blow-up rate for the critical nonlinear Schrödinger equation. *Geom. Funct. Anal.*, 13(3) :591–642, 2003.
- [MR04] F. Merle and P. Raphael. On universality of blow-up profile for L^2 critical nonlinear Schrödinger equation. *Invent. Math.*, 156(3) :565–672, 2004.
- [MR05a] F. Merle and P. Raphael. The blow-up dynamic and upper bound on the blow-up rate for critical nonlinear Schrödinger equation. *Ann. of Math. (2)*, 161(1) :157–222, 2005.
- [MR05b] F. Merle and P. Raphael. Profiles and quantization of the blow up mass for critical nonlinear Schrödinger equation. *Comm. Math. Phys.*, 253(3) :675–704, 2005.
- [MR05c] F. Merle and P. Raphael. On one blow up point solutions to the critical nonlinear Schrödinger equation *J. Hyperbolic Differ. Equ.*, 2(4) : 919–962, 2005
- [Naw99] H. Nawa. Asymptotic and limiting profiles of blowup solutions of the nonlinear Schrödinger equation with critical power. *Comm. Pure Appl. Math.*, 52(2) :193–270, 1999.

- [OT91] T. Ogawa and Y. Tsutsumi. Blow-up of H^1 solution for the nonlinear Schrödinger equation. *J. Differential Equations*, 92(2) :317–330, 1991.
- [Per01] G. Perelman. On the formation of singularities in solutions of the critical nonlinear Schrödinger equation. *Ann. Henri Poincaré*, 2(4) :605–673, 2001.
- [Rap] P. Raphael. On the blow up phenomenon for the l^2 critical non linear schrödinger equation. *Préprint, 2004*
- [Rap05] P. Raphael. Stability of the log-log bound for blow up solutions to the critical non linear Schrödinger equation. *Math. Ann.*, 331(3) :577–609, 2005.
- [SS99] C. Sulem and P.-L. Sulem. *The nonlinear Schrödinger equation*, volume 139 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1999. Self-focusing and wave collapse.
- [Tzv04] N. Tzvetkov. On the long time behaviour of kdv type equations (after Martel-Merle). Sémin. Bourbaki (2003-04), Exp. n° 933, *Astérisque*, 299 :219–248, 2005.
- [Wei85] M.I. Weinstein. Modulational stability of ground states of nonlinear Schrödinger equations. *SIAM J. Math. Anal.*, 16(3) :472–491, 1985.
- [Wei83] M.I. Weinstein. Nonlinear Schrödinger equations and sharp interpolation estimates. *Comm. Math. Phys.*, 87(4) :567–576, 1982/83.
- [ZS71] V.E. Zakharov and A.B. Shabat. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Ž. Èksper. Teoret. Fiz.*, 61(1) :118–134, 1971.

Nicolas BURQ

Université Paris XI

UMR 8628 du CNRS

Département de Mathématiques

Bâtiment 425

F–91405 Orsay Cedex

E-mail : nicolas.burq@math.u-psud.fr

**PROJECTIVITY OF KÄHLER MANIFOLDS -
KODAIRA'S PROBLEM**
[after C. Voisin]

by **Daniel HUYBRECHTS**

There are various geometric structures that can be studied on a topological manifold M . Depending on one's geometric taste, it is important to know whether M can be endowed with a symplectic form, whether (special) Riemannian metrics can be found or whether M carries an algebraic structure. Often, the existence of a certain geometric structure imposes topological conditions on M . In other words, it may happen that a given topological manifold does simply not allow one's favorite geometry. E.g. if M is compact and $b_2(M) = 0$ the manifold M cannot be symplectic, or if $b_1(M) = 1$ no Kähler metrics can exist.

In order to fully understand the relation between two sorts of geometries, it is important to know whether they impose the same topological obstructions. In other words, does the existence of one of the two on a given manifold topological M imply the existence of the other one? This is a report on the work of Claire Voisin [11, 12] that sheds light on an old question, usually attributed to Kodaira, that asks for the topological relation between Kähler geometry and projective geometry.

In the following we let M be a compact manifold that can be endowed with the structure of a complex manifold. Once a complex structure is chosen, one studies Riemannian metrics g that are 'compatible' with it. One possible compatibility condition is to require that g be hermitian, i.e. that the complex structure thought of as an almost complex structure I is orthogonal with respect to g . It is not difficult to see that a hermitian structure can always be found. It is, however, a completely different matter to find a hermitian structure g such that its fundamental form $\omega := g(I, \cdot)$ is closed, i.e. g satisfies the Kähler condition. Indeed, the classical theory of Kähler manifolds shows that the existence of a Kähler metric imposes strong conditions on the topology of M , which are not satisfied by arbitrary complex or symplectic manifolds. For instance, the odd Betti numbers of a compact Kähler manifold are even, Kähler manifolds are formal and their fundamental groups satisfy further conditions. (In contrast, if only one of the two structures, complex or symplectic, is required, then any finitely presentable group can be realized.)

On the other hand, Kähler manifolds are quite common. Indeed, any complex submanifold of the complex projective space \mathbb{P}^n admits a Kähler metric - the restriction of the Fubini-Study metric is an example. Conversely, one might wonder whether a

compact complex manifold that admits a Kähler structure can always be realized as a complex submanifold of \mathbb{P}^n or, in other words, whether the complex structure is projective. This is obviously not the case, general complex tori \mathbb{C}^n/Γ ($n \geq 2$) and general K3 surfaces provide counter-examples. In fact, a famous theorem of Kodaira proves that a Kähler manifold is projective if and only if the Kähler metric can be chosen such that the cohomology class of its fundamental form ω is integral, i.e. $[\omega] \in H^2(X, \mathbb{Z})$ (see [6, Thm. 4]).

In these examples one observes that although the given complex structure is not projective, it becomes projective after a small deformation. Kodaira proved that in fact any Kähler surface can be deformed to a projective surface (see [7, Thm. 23] and [8]). Thus, as deforming the complex structure does not change the diffeomorphism type of the manifold, there is no topological difference between compact Kähler surfaces and algebraic surfaces. (Let us also mention that in fact any compact surface X with even $b_1(X)$ is Kähler, i.e. for surfaces the condition to be Kähler is a topological condition. This fails in higher dimensions, due to a famous example of Hironaka [5] of a compact Kähler manifold that deforms to complex manifold which is no longer Kähler.) Note in passing that a similar result holds true for symplectic manifolds: clearly, any given symplectic form ω can be deformed to a symplectic form with integral cohomology class.

Kodaira's problem, which apparently has never been stated by himself in this form, asks for the higher-dimensional version of his result: *Can any compact Kähler manifold be deformed to a projective manifold?*

More in the spirit of the general philosophy explained above, one could ask whether the topological manifold underlying a compact Kähler manifold may also be endowed with the structure of a projective manifold. This question had been open for a very long time. As Kodaira's arguments to prove the two-dimensional case use a great deal of classification theory of surfaces, there was little hope to generalize them to higher dimensions.

Recent work of Claire Voisin fills this gap [11, 12, 13]. She succeeded in showing that topology makes a difference between compact Kähler manifolds and those that are projective. In other words, there exist compact topological manifolds that admit the structure of a Kähler manifold without carrying also the structure of a projective manifold. More precisely, Voisin shows the stronger statement:

THEOREM 0.1 ([11]). — *In any dimension ≥ 4 there exists a compact Kähler manifold X whose rational cohomology ring $H^*(X, \mathbb{Q})$ cannot be realized as the rational cohomology ring of a projective manifold.*

Voisin originally worked with the integral cohomology ring $H^*(X, \mathbb{Z})$, but Deligne then pointed out the stronger version above.

One could wonder whether the answer to these questions would be different if the topological manifold satisfies further conditions, e.g. if it is in addition simply-connected. Some of these questions have been addressed and answered by Voisin in [11, 12] and we will comment on them on the way.

Although the examples are obtained by particular constructions, the principal ideas of [11, 12] are of a more general nature and might be applicable in other situations.

The i -th cohomology of a compact Kähler manifold is naturally endowed with a Hodge structure of weight i , which can be polarized (on the primitive part) if the manifold is projective. The idea is to show that there exist compact Kähler manifolds whose cohomology does not admit Hodge structures that are compatible with both, the given cup-product and a polarization. Roughly, there are three steps **A-C**, the first two of which are purely Hodge-theoretical and only the last one has a geometric flavor.

(A) Certain algebraic structures on a rational vector space A are not compatible with any *polarizable* Hodge structure (of weight k) on A .

Remark 0.2. — In the examples, the algebraic structure will be a specific endomorphism $\Phi : A \rightarrow A$, but others are in principle possible. That the algebraic structure is not compatible with any polarizable Hodge structure means in the case of an endomorphism Φ that one cannot find a Hodge structure on A such that Φ becomes an endomorphism of it and such that the Hodge structure can be polarized.

(B) Suppose $\bigoplus H^\ell$ is a graded \mathbb{Q} -algebra whose direct summands H^ℓ are Hodge structures of weight ℓ and such that the multiplications $H^{\ell_1} \otimes H^{\ell_2} \rightarrow H^{\ell_1+\ell_2}$ are homomorphisms of Hodge structures. Suppose furthermore that this \mathbb{Q} -algebra structure allows us to detect a subspace $A \subset H^k$ such that: i) $A \subset H^k$ is a Hodge substructure. ii) An algebraic structure as in **(A)** is compatible with this Hodge structure. Then H^k does not admit a polarization.

Remark 0.3. — Subspaces that are defined purely in terms of the \mathbb{Q} -algebra structure do define Hodge substructures. We shall also need a refined version of this, which is due to Deligne.

The compatibility in ii) is more difficult to check, but relies on the same principle. For an endomorphism Φ the idea goes as follows: Firstly, find two Hodge substructures $A, A' \subset H^k$ and a Hodge substructure $\Delta \subset A \oplus A' \subset H^k$ which is the graph of an isomorphism $A \cong A'$. Secondly, prove that under the induced isomorphism of Hodge structures $A \oplus A \cong A \oplus A'$ the graph of Φ is a Hodge substructure.

(C) Construct compact Kähler manifolds such that the above principles apply to its cohomology ring $\bigoplus H^\ell(X, \mathbb{Q})$. Then $H^*(X, \mathbb{Q})$ should not be realizable by a smooth projective variety.

Remark 0.4. — This works best for Hodge structures of weight one ($k = 1$). In this case $H^1(X, \mathbb{Q})$ of a smooth projective variety X admits a polarized Hodge structure. For the Hodge structure of weight two on $H^2(X, \mathbb{Q})$ one needs an extra argument, for only the primitive part of it admits a polarization.

This report roughly follows these three steps. Some of the algebraic structures in Section 2 might seem rather ad hoc, as their geometric origin is only explained in Section 3. However, I found it helpful for my own understanding to completely separate the arguments that explain why certain \mathbb{Q} -algebras cannot be realized as the cohomology of a projective manifold from the part that contains the construction of compact Kähler manifolds that do realize these \mathbb{Q} -algebras.

Acknowledgements: I wish to thank Claire Voisin for patiently answering my questions and for her valuable comments on a first draft of these notes. I am grateful to C.-F. Bödigheimer, U. Görtz, M. Lehn, R. Thomas, P. Stellari, J. Stix, B. Totaro, and T. Wedhorn for their help, comments, and suggestions.

1. HODGE STRUCTURES (OF WEIGHT ONE AND TWO)

1.1. Recollections

A Hodge structure of weight k on a \mathbb{Q} -vector space A is given by a direct sum decomposition

$$(1) \quad A_{\mathbb{C}} := A \otimes_{\mathbb{Q}} \mathbb{C} = \bigoplus_{p+q=k} A^{p,q} \text{ such that } \overline{A^{p,q}} = A^{q,p}.$$

A direct sum decomposition (1) can also be described in terms of a representation $\rho : \mathbb{C}^* \rightarrow \mathrm{Gl}(A_{\mathbb{R}})$ such that the \mathbb{C} -linear extension of $\rho(z)$ satisfies $\rho(z)|_{A^{p,q}} = z^p \bar{z}^q \cdot \mathrm{id}$. The Hodge classes of a Hodge structure of weight $2k$ on A are the elements in $A^{k,k} \cap A$.

We shall be particularly interested in Hodge structures of weight one and two.

Remark 1.1. — Recall that Hodge structures of weight one with $A^{p,q} = 0$ for $pq \neq 0$ which are integral, i.e. $A = \Gamma_{\mathbb{Q}}$ for some lattice Γ , are in bijection with complex tori. Indeed, to a Hodge structure of weight one on $\Gamma_{\mathbb{Q}}$ given by $\Gamma_{\mathbb{C}} = A^{1,0} \oplus A^{0,1}$ one associates the complex torus $A^{1,0}/\Gamma$, where Γ is identified with its image under the projection $A_{\mathbb{C}} \rightarrow A^{1,0}$.

A \mathbb{Q} -linear map $\varphi : A \rightarrow A'$ is a *morphism (of weight m) of Hodge structures*

$$A_{\mathbb{C}} = \bigoplus_{p+q=k} A^{p,q} \text{ and } A'_{\mathbb{C}} = \bigoplus_{r+s=\ell} A'^{r,s}$$

of weight k and $\ell = k + 2m$, respectively, if $\varphi(A^{p,q}) \subset A'^{p+m,q+m}$. If the two Hodge structures correspond to $\rho : \mathbb{C}^* \rightarrow \mathrm{Gl}(A_{\mathbb{R}})$ and $\rho' : \mathbb{C}^* \rightarrow \mathrm{Gl}(A'_{\mathbb{R}})$, respectively, then this condition is equivalently expressed by $\varphi(\rho(z)v) = |z|^{2m} \rho'(z)\varphi(v)$ for all $v \in A$ and $z \in \mathbb{C}^*$.

A *Hodge substructure* of a Hodge structure of weight k on A is given by a subspace $A' \subset A$ such that $A'_\mathbb{C} = \bigoplus (A^{p,q} \cap A'_\mathbb{C})$ or, equivalently, such that $A'_\mathbb{C} \subset A_\mathbb{C}$ is invariant under the representation $\rho : \mathbb{C}^* \rightarrow \text{Gl}(A_\mathbb{R})$ that corresponds to the given Hodge structure on A .

The *tensor product* $A \otimes_{\mathbb{Q}} A'$ of two \mathbb{Q} -vector spaces A and A' endowed with Hodge structures of weight k and ℓ , respectively, comes with a natural Hodge structure of weight $(k + \ell)$:

$$(A \otimes_{\mathbb{Q}} A')^{r,s} := \bigoplus_{p+p'=r, q+q'=s} A^{p,q} \otimes_{\mathbb{C}} A'^{p',q'}.$$

In other words, the Hodge structure is given by $\rho \otimes \rho'$.

Note that $A_2 := \bigwedge^2 A_1$ of a Hodge structure of weight one A_1 is naturally a Hodge structure of weight two with $A_2^{2,0} := \bigwedge^2 A_1^{1,0}$, $A_2^{1,1} := A_1^{1,0} \otimes A_1^{0,1}$, and $A_2^{0,2} = \bigwedge^2 A_1^{0,1}$.

A *polarization* of a Hodge structure of weight one $A_\mathbb{C} = A^{1,0} \oplus A^{0,1}$ is a skew-symmetric form $q \in \bigwedge^2 A^*$ such that

$$(2) \quad A_\mathbb{C} \times A_\mathbb{C} \longrightarrow \mathbb{C}, \quad (v, w) \longmapsto iq(v, \bar{w})$$

(where q is extended \mathbb{C} -linearly) satisfies the Hodge–Riemann relations:

- i) $A^{1,0}$ and $A^{0,1}$ are orthogonal with respect to (2).
- ii) The restriction of (2) to $A^{1,0}$ and to $A^{0,1}$ is positive, respectively negative, definite.

Remark 1.2. — With this definition a polarization is always rational. Furthermore, the form q considered as an element of the induced weight-two Hodge structure on $\bigwedge^2 A^*$ is of type $(1, 1)$. Since it is rational, q is a Hodge class (of weight two). Note that any Hodge substructure of a weight-one polarized Hodge structure is naturally polarized.

Example 1.3. — Let X be a compact Kähler manifold of dimension n . The Hodge decomposition

$$H^1(X, \mathbb{C}) = H^{1,0}(X) \oplus H^{0,1}(X)$$

defines a Hodge structure of weight one on $H^1(X, \mathbb{Q})$.

Suppose X is projective and $\omega \in H^2(X, \mathbb{Z})$ is the class of a hyperplane section, then $q(\alpha) = \int_X \alpha^2 \omega^{n-1}$ is a polarization of the natural Hodge structure of weight one on $H^1(X, \mathbb{Q})$.

If we drop the condition that q be rational, then any Kähler class on a compact Kähler manifold X would yield a form on the Hodge structure of weight one on $H^1(X, \mathbb{Q})$ that satisfies the Hodge–Riemann relations i) and ii).

The notion of a polarization exists for Hodge structures of arbitrary weight, but we shall only need it for weight one, explained above, and for weight two. For a Hodge structure of weight two $A_\mathbb{C} = A^{2,0} \oplus A^{1,1} \oplus A^{0,2}$ a *polarization* is a symmetric bilinear form $q \in S^2 A^*$ such that:

- i) The $A^{p,q}$ are pairwise orthogonal with respect to $(v, w) \mapsto q(v, \bar{w})$.
- ii) For $0 \neq v \in A^{p,q}$ one has $-i^{p-q}q(v, \bar{v}) > 0$.

Example 1.4. — If X is compact Kähler of dimension n , then $H^2(X, \mathbb{Q})$ comes with a natural Hodge structure of weight two $H^2(X, \mathbb{C}) = H^{2,0}(X) \oplus H^{1,1}(X) \oplus H^{0,2}(X)$ given by the Hodge decomposition. If X is projective and $\omega \in H^2(X, \mathbb{Z})$ is the class of a hyperplane section, then

$$q(\alpha) = \int_X \alpha^2 \omega^{n-2}$$

defines a polarization on the *primitive cohomology*

$$H^2(X, \mathbb{Q})_p := \{\alpha \in H^2(X, \mathbb{Q}) \mid \alpha \wedge \omega^{n-1} = 0\}.$$

Note that due to the Hodge–Riemann bilinear relation $H^{1,1}(X, \mathbb{R}) \cong H^{1,1}(X, \mathbb{R})_p \oplus \mathbb{R}\omega$ does not contain any q -isotropic subspace of dimension ≥ 2 . Also, $H^2(X, \mathbb{R})$ does not contain Hodge substructures of dimension ≥ 2 which are q -isotropic.

1.2. Detecting Hodge structures algebraically

The following observation is the key to a general principle, due to Deligne, which allows one to identify Hodge substructures algebraically.

LEMMA 1.5. — *Let $H_{\mathbb{C}} = \bigoplus_{p+q=k} H^{p,q}$ be a Hodge structure of weight k on a \mathbb{Q} -vector space H given by a representation $\rho : \mathbb{C}^* \rightarrow \text{Gl}(H_{\mathbb{R}})$ and let $Z \subset H_{\mathbb{C}}$ be an algebraic subset which is invariant under $\rho(\mathbb{C}^*)$. Suppose the span $\langle Z' \rangle$ of an irreducible component $Z' \subset Z$ is of the form $H' \otimes_{\mathbb{Q}} \mathbb{C}$ with $H' \subset H$ a \mathbb{Q} -subspace. Then H' is a Hodge substructure of H .*

Proof. — Since \mathbb{C}^* is connected, the \mathbb{C}^* -action leaves invariant the irreducible components of Z . Hence, also $\langle Z' \rangle$ is \mathbb{C}^* -invariant. But for $\langle Z' \rangle = H' \otimes_{\mathbb{Q}} \mathbb{C}$ this is equivalent to saying that $H' \subset H$ is a Hodge substructure. \square

In [11, 12] the lemma is applied in various situations. The algebraic set Z is always defined by algebraic conditions on homomorphisms of Hodge structures and thus automatically invariant under \mathbb{C}^* . Usually, one starts with several Hodge structures of weight ℓ on \mathbb{Q} -vector spaces H^ℓ and homomorphisms of Hodge structures $H^{\ell_1} \otimes H^{\ell_2} \rightarrow H^{\ell_1+\ell_2}$, $a \otimes b \mapsto a \cdot b$. (Think of the cohomology of a smooth projective variety or of a compact Kähler manifold.)

We shall in particular encounter algebraic subsets of the form

$$Z_1 = \{\alpha \in H_{\mathbb{C}}^k \mid \alpha^2 = 0\} \text{ or } Z_2 = \left\{ a \in H_{\mathbb{C}}^k \mid \text{rk} \left(H_{\mathbb{C}}^{\ell} \xrightarrow{a} H_{\mathbb{C}}^{k+\ell} \right) \leq m \right\}.$$

Let us sketch the argument that shows that these sets are \mathbb{C}^* -invariant in the example $Z = Z_2$. By definition of the Hodge structure on $H^{\ell_1} \otimes H^{\ell_2}$ and the hypothesis that the multiplication $a \otimes b \mapsto a \cdot b$ is a morphism of Hodge structures, one has $\rho(z)(a) \cdot b = \rho(z)(a \cdot (\rho(z^{-1})(b)))$. Thus, the endomorphism given by multiplication with $\rho(z)(a)$ and a , respectively, differ by automorphisms $\rho(z) \in \text{Gl}(H_{\mathbb{R}}^{k+\ell})$ and $\rho(z^{-1}) \in \text{Gl}(H_{\mathbb{R}}^{\ell})$. In particular, $\text{rk}(\rho(z)(a) \cdot) = \text{rk}(a \cdot)$ and hence $a \in Z$ if and only if $\rho(z)(a) \in Z$.

Note that it might well happen that $\langle Z \rangle$ is defined over \mathbb{Q} , but not $\langle Z' \rangle$.

Let us illustrate the use of Deligne’s principle in a concrete situation that will be at the heart of the subsequent discussion. Suppose we are given a graded \mathbb{Q} -algebra $\bigoplus H^k$, an integer $\ell \in \mathbb{Z}$ and a subspace $0 \neq H' \subset H^\ell$. Then define for $i \geq 1$ the \mathbb{Q} -subspace

$$(3) \quad P_i := \left\{ a \in H^2 \mid \left(\bigotimes^i H' \xrightarrow{\cdot a} H^{\ell i + 2} \right) = 0 \right\}.$$

We shall later fix in addition an integer $m > 1$ and consider the two subspaces

$$P_1 \subset P_m \subset H^2$$

and the algebraic subset of $P_{m\mathbb{C}}$:

$$(4) \quad Z := \{ a \in P_{m\mathbb{C}} \mid \text{Ker} \left(H'_\mathbb{C} \xrightarrow{\cdot a} H^{\ell+2}_\mathbb{C} \right) \neq 0 \}.$$

Then Z contains $P_{1\mathbb{C}}$ and we denote its image in $(P_m/P_1)_\mathbb{C}$ by \bar{Z} (which is again algebraic). Furthermore, let $e \in Z \cap P_m$ be such that $\mathbb{C}\bar{e} \subset \bar{Z}$ is an irreducible component of \bar{Z} .

COROLLARY 1.6. — *Suppose each H^k is endowed with a Hodge structure of weight k such that the multiplications are morphisms of Hodge structures and such that $H' \subset H^\ell$ is a Hodge substructure. Then*

- i) *the $P_i \subset H^2$ are Hodge substructures,*
- ii) *the element $\bar{e} \in P_m/P_1$ is of type $(1, 1)$, i.e. a Hodge class, and*
- iii) *$\text{Ker}(H' \xrightarrow{\cdot e} H^{\ell+2})$ is a Hodge substructure of H^ℓ .*

Proof. — The P_i can be viewed as the kernels of the morphisms of Hodge structures $H^2 \rightarrow \left(\bigotimes^i H' \right)^* \otimes H^{\ell i + 2}$ and are, therefore, Hodge substructures of H^2 .

Deligne’s principle shows that $\mathbb{Q}\bar{e} \subset P_m/P_1$ is a Hodge substructure. Since any weight two Hodge structure of rank one is of pure type, one finds $\bar{e} \in (P_m/P_1)^{1,1}$.

In order to prove iii), use the morphism of Hodge structures $P_m/P_1 \otimes H' \rightarrow H^{\ell+2}$. \square

Remark 1.7. — i) The actual description of P_m is of no importance here. We only used $P_1 \subset P_m$ and the condition on e . Note that $e \in P_m$ itself might be of mixed type, e.g. it could be arbitrarily modified by rational classes in $P_1^{2,0} \oplus P_1^{0,2}$.

ii) In the applications only the cases $\ell = 1$ and $\ell = 2$ will be considered and, moreover, for $\ell = 1$ we will have $H' = H^1$.

2. THE IMPOSSIBLE ONES

The aim is to exhibit two specific Hodge structures of weight one respectively two which resist polarization. Section 2.1 explains Step **A** of the program, whereas Section 2.2 corresponds to Step **B**.

2.1. Special endomorphisms excluding polarization

Let us start out with an endomorphism $\Phi \in \text{End}(A)$ of a \mathbb{Q} -vector space A of dimension $2n$. For any field $\mathbb{Q} \subset K$ we shall denote by Φ_K its K -linear extension. We also use the naturally induced endomorphisms Φ^* and $\bigwedge^2 \Phi^*$ of A^* and $\bigwedge^2 A^*$ respectively.

Denote the set of all eigenvalues of Φ by $EV(\Phi) := \{\mu_1, \dots, \mu_{2n}\}$ and by K_Φ the splitting field of the characteristic polynomial of Φ , i.e. $K_\Phi = \mathbb{Q}(\mu_1, \dots, \mu_{2n})$.

Henceforth, we shall assume that:

HYPOTHESIS 2.1. — i) $\mu_i \notin \mathbb{R}$ for all i , and ii) $G := \text{Gal}(K_\Phi/\mathbb{Q})$ acts as the symmetric group S_{2n} on $EV(\Phi)$.

Example 2.2. — It is not difficult to find explicit examples of endomorphisms Φ satisfying these conditions:

– Let $A = \mathbb{Q}^2$, hence $n = 1$, and $\Phi = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then $\{\mu_1, \mu_2\} = \{\pm i\}$.

– Let $A = \mathbb{Q}^4$ and $\Phi = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. The characteristic polynomial of Φ is $x^4 - x + 1$

whose Galois group is the symmetric group (see [1, Ch.14.6]) and which clearly has no real eigenvalues.

Remark 2.3. — Clearly, ii) implies that $\Phi_{\mathbb{C}} \in \text{End}(A_{\mathbb{C}})$ can be diagonalized. It also yields $\mu_{i_1} \cdot \dots \cdot \mu_{i_k} \neq \mu_{j_1} \cdot \dots \cdot \mu_{j_k}$ for any two distinct multi-indices $i_1 < \dots < i_k$ and $j_1 < \dots < j_k$.

LEMMA 2.4. — Under the assumptions of 2.1 the induced endomorphism $\bigwedge^k \Phi \in \text{End}(\bigwedge^k A)$ does not admit any non-trivial invariant subspace.

Proof. — Clearly, the eigenvalues of $\bigwedge^k \Phi$ are $\mu_{i_1} \cdot \dots \cdot \mu_{i_k}$, $i_1 < \dots < i_k$. Thus, if $W \subset \bigwedge^k A$ is invariant under $\bigwedge^k \Phi$, then the eigenvalues of $\psi := \bigwedge^k \Phi|_W$ are also of this form. In particular, also ψ can be diagonalized over K_Φ . Suppose $W \neq 0$. Then there exists an eigenvector $v \in W_{K_\Phi}$ with eigenvalue say $\mu_1 \cdot \dots \cdot \mu_k$.

Being defined over \mathbb{Q} , the extension of ψ (and of $\bigwedge^k \Phi$) to an endomorphism of W_{K_Φ} (respectively $\bigwedge^k A_{K_\Phi}$) commutes with the action of the Galois group G on the scalars K_Φ . Hence, with $\mu_1 \cdot \dots \cdot \mu_k$ also $\mu_{\sigma(1)} \cdot \dots \cdot \mu_{\sigma(k)}$ is an eigenvalue of ψ for any $\sigma \in G$.

By Remark 2.3, this shows that all $\mu_{i_1} \cdot \dots \cdot \mu_{i_k}$, $i_1 < \dots < i_k$, which are pairwise distinct, occur as eigenvalues of ψ . Hence, $\dim(W) = \dim \bigwedge^k A$ or, equivalently, $W = \bigwedge^k A$. □

PROPOSITION 2.5. — Suppose $\bigwedge^2 \Phi$ respects a Hodge structure of weight two on $A_2 := \bigwedge^2 A$ given by $\bigwedge^2 A_{\mathbb{C}} = A_2^{2,0} \oplus A_2^{1,1} \oplus A_2^{0,2}$ with $A_2^{2,0} \neq 0$. If Φ satisfies 2.1, then

$$A_2^{1,1} \cap \bigwedge^2 A = \{0\},$$

which is equivalent to saying that all Hodge classes of A_2 are trivial.

Proof. — As $\bigwedge^2 \Phi_{\mathbb{C}}$ preserves the bidegree (p, q) of elements in $\bigwedge^2 A_{\mathbb{C}}$, the rational subspace $W := A_2^{1,1} \cap \bigwedge^2 A$ is $\bigwedge^2 \Phi$ -invariant. Due to the lemma one either has $W = \bigwedge^2 A$, which is excluded by $A_2^{2,0} \neq 0$, or $W = 0$, which proves the assertion. \square

COROLLARY 2.6. — Suppose $n \geq 2$. A Hodge structure of weight one $A_{\mathbb{C}} = A^{1,0} \oplus A^{0,1}$ that is preserved by $\Phi_{\mathbb{C}}$ does not admit a polarization.

Proof. — A polarization of the Hodge structure $A_{\mathbb{C}} = A^{1,0} \oplus A^{0,1}$ would be given by a special Hodge class q in the induced Hodge structure of weight two on $\bigwedge^2 A^*$. However, there are no non-trivial ones due to the proposition. (Use that Φ^* as well satisfies 2.1.) The assumption $n \geq 2$ is needed in order to ensure that $A_2^{2,0} \neq 0$. \square

Remark 2.7. — Observe that Φ preserves the Hodge structure if and only if its graph $\Gamma_{\Phi} \subset A \oplus A$ is a Hodge substructure.

Example 2.8. — If Φ satisfies i) and ii) of 2.1, one easily constructs Hodge structures of weight one that are preserved by Φ . This will be needed when it actually comes to constructing examples.

Pick n distinct eigenvalues $\lambda_1, \dots, \lambda_n \in EV(\Phi)$ such that $\lambda_i \neq \bar{\lambda}_j$ for all i, j (note that due to i) no eigenvalue is real) and let $A^{1,0} = \bigoplus_{i=1}^n \mathbb{C}v_i$, where the $v_i \in A_{\mathbb{C}}$ are eigenvectors with eigenvalue λ_i .

With Φ being defined over \mathbb{Q} , the complex conjugate $\bar{\lambda}$ of an eigenvalue $\lambda \in EV(\Phi)$ is again an eigenvalue. Thus, with $A^{0,1} := \overline{A^{1,0}}$ one has $A_{\mathbb{C}} = A^{1,0} \oplus A^{0,1}$.

2.2. Identifying the special endomorphisms algebraically

We continue the discussion of Section 1.2 and combine it with endomorphisms Φ of the type studied in Section 2.1.

Thus, consider a \mathbb{Q} -vector space A of dimension $2n \geq 4$ together with an endomorphism Φ and let $H^* = \bigoplus_{k=0}^{4n} H^k$ be a graded \mathbb{Q} -algebra.

To bring both structures together, we assume that there is a graded inclusion

$$\bigwedge^* (A \oplus A) \subset H^*$$

satisfying the following conditions. (We shall apply Corollary 1.6 with $\ell = 1$, $m = 4n - 2$, and $H' = H^1$.)

HYPOTHESIS 2.9. — i) $A \oplus A = H^1$,

ii) $H^2 = \bigwedge^2(A \oplus A) \oplus P \oplus R$, where $P := P_{4n-2}$ is defined as in (3) and R is some subspace,

iii) $P = P_1 \oplus \bigoplus_{i=1}^4 e_i \mathbb{Q}$, and

iv) the kernel of the multiplication $H^1 \xrightarrow{\cdot e_i} H^3$, for $i = 1, \dots, 4$, equals the subspaces $A \oplus \{0\}$, $\{0\} \oplus A$, $\Delta := \{(a, a) \mid a \in A\}$, and the graph Γ_Φ of Φ , respectively. The sum $\sum \text{Im}(\cdot e_i) \subset H^3$ is direct.

Remark 2.10. — Roughly, e_1 and e_2 will be used to detect certain Hodge substructures, e_3 to identify them, and e_4 to view Φ as a homomorphism between them. The auxiliary space R is later only needed in order to construct odd-dimensional examples. Due to Remark 3.4 one could even restrict to the case $P_1 = 0$.

PROPOSITION 2.11. — Suppose H^* and Φ meet the conditions of 2.9 and 2.1, respectively. Then H^* cannot be realized as the rational cohomology ring $H^*(X, \mathbb{Q})$ of a projective manifold X .

Proof. — Suppose X is a projective manifold that does realize H^* . In the following we will simply identify $H^*(X, \mathbb{Q})$ with H^* . Thus, each H^k inherits the natural Hodge structure of weight k from $H^k(X, \mathbb{Q})$ and the multiplications $H^{\ell_1} \otimes H^{\ell_2} \rightarrow H^{\ell_1+\ell_2}$ are morphisms of Hodge structures.

Corollary 1.6 applies and shows that $A \oplus \{0\}$, $\{0\} \oplus A$, Δ , and the graph Γ_Φ are Hodge substructures of $H^1(X, \mathbb{Q})$. Indeed, the only thing that needs to be checked is that the $\mathbb{C}\bar{e}_i$ define irreducible components of $\bar{Z} \subset (P/P_1)_{\mathbb{C}}$ (the image of Z as in (4)). This follows from iv): Suppose $\sum a_i e_i \in Z$. Then there exists $0 \neq a \in H^1$ that is annihilated by it. Thus, $a_i(a \cdot e_i) = 0$ for $i = 1, \dots, 4$. If e.g. $a_i \neq 0 \neq a_j$, then $a \in \text{Ker}(\cdot e_i) \cap \text{Ker}(\cdot e_j)$. The description of the kernels shows that this is impossible.

With the identification of the two Hodge structures on $A \oplus \{0\}$ and $\{0\} \oplus A$ via Δ , the graph Γ_Φ allows to view Φ as an endomorphism of the Hodge structure on $A \oplus \{0\}$.

By Corollary 2.6 this Hodge structure does not admit a polarization. Hence, also the Hodge structure $H^1(X, \mathbb{Q})$, of which $A \oplus \{0\}$ is a Hodge substructure, cannot be polarized. This yields a contradiction to the projectivity of X . \square

We shall next present a similar result based on an analysis of Hodge structures of weight two.

Let as before A be a \mathbb{Q} -vector space of dimension $2n \geq 4$ together with an endomorphism Φ and let $H^* = \bigoplus_{k=0}^{4n} H^k$ be a graded \mathbb{Q} -algebra. We assume that there is a graded inclusion $\bigwedge^{2*}(A \oplus A) \subset H^{2*}$ and consider $B_1 := \bigwedge^2 A \oplus \{0\}$, $B_2 := \{0\} \oplus \bigwedge^2 A$, and $H' := B_1 \oplus B_2$ as subspaces of H^2 . We shall use the notation of Corollary 1.6 with $\ell = 2$, $m = 2n - 1$.

HYPOTHESIS 2.12. — i) $H^2 = B_1 \oplus B_2 \oplus P$ with $P := P_{2n-1}$ as in (3),

ii) $P = Q_1 \oplus Q_2 \oplus \mathbb{Q}e_1 \oplus \mathbb{Q}e_2$ for some vector space Q and $P_1 = 0$,

iii) $\{\alpha \in H_{\mathbb{C}}^2 \mid \alpha^2 = 0\} = \{\alpha \in B_{1\mathbb{C}} \mid \alpha^2 = 0\} \cup \{\alpha \in B_{2\mathbb{C}} \mid \alpha^2 = 0\}$,

iv) $\alpha^2 a^{2n-2} = 0$ for all $\alpha \in B_1$ and $a \in P$, and

v) the kernel of the multiplication $B_1 \oplus B_2 \xrightarrow{\cdot e_i} H^4$, $i = 1, 2$, is the diagonal $\Delta := \{(a, a)\} \subset \bigwedge^2 A \oplus \bigwedge^2 A$ for $i = 1$ and the graph $\Gamma_{\bigwedge^2 \Phi}$ for $i = 2$. Similarly, $\text{Ker}(\cdot a_i) = B_i$ for any $0 \neq a_i \in Q_i$. The sum $Q_1 \cdot B_2 + Q_2 \cdot B_1 + \text{Im}(\cdot e_1) + \text{Im}(\cdot e_2)$ is direct.

PROPOSITION 2.13. — Suppose H^* and Φ meet the requirements of 2.12 and 2.1, respectively. Then H^* cannot be realized as the rational cohomology ring $H^*(X, \mathbb{Q})$ of a projective manifold X .

Proof. — Suppose X is a projective manifold whose rational cohomology ring $H^*(X, \mathbb{Q})$ can be identified with H^* .

Due to iii) and Lemma 1.5, B_1 , B_2 , and hence H' are Hodge substructures of H^2 . Thus, Corollary 1.6 applies and shows that P is a Hodge substructure. (Note that $\bigwedge^2 A$ is spanned by vectors α with $\alpha^2 = 0$.)

Due to v) the algebraic set $\bar{Z} \subset P$ (see notation in Corollary 1.6) contains $\mathbb{C}\bar{e}_1$ and $\mathbb{C}\bar{e}_2$ as two irreducible components. Indeed, if $\sum a_i + \sum \eta_i e_i \in Z$ with $a_i \in Q_i$, then some $0 \neq b = b_1 + b_2 \in B_1 \oplus B_2$ is annihilated by it. Since the sum of the multiplications is direct, this yields $a_2 \cdot b_1 = a_1 \cdot b_2 = \eta_i(a \cdot e_i) = 0$. In particular, $a_1 \neq 0$ implies $b_2 = 0$ and $a_2 \neq 0$ implies $b_1 = 0$. Thus, if $\eta_1 = \eta_2 = 0$, then either $a_1 \neq 0$ or $a_2 \neq 0$. Similarly, if $\eta_1 \neq 0 = \eta_2$, then $b_1 = b_2 \neq 0$ and, therefore, $a_1 = a_2 = 0$. Finally, the case $\eta_1 \neq 0 \neq \eta_2$ is excluded by $\Delta \cap \Gamma_{\bigwedge^2 \Phi} = \{(0, 0)\}$, which follows from $\mu_i \cdot \mu_j \neq 1$ for all $i \neq j$ and $n \geq 2$. (The argument shows that the other irreducible components are $Q_{i\mathbb{C}}$.)

Thus, by iii) of Corollary 1.6, the diagonal and the graph of $\bigwedge^2 \Phi$ are Hodge substructures of $B_1 \oplus B_2$. In other words, $\bigwedge^2 \Phi$ is an endomorphism of the Hodge structure of $\bigwedge^2 A$ induced by B_1 (or, equivalently, by B_2).

Clearly, $\bigwedge^2 A$ contains a subspace V of dimension at least two such that $0 = \alpha^2 \in H^4$ for all $\alpha \in V$. (For instance, take $V = \langle v_1 \wedge v_2, v_1 \wedge v_3 \rangle$ if $A = \bigoplus \mathbb{Q}v_i$.)

Hence, by the Hodge–Riemann bilinear relations this excludes $V \subset H^{1,1}(X)$ (see Example 1.4). Therefore, $\bigwedge^{2,0} A \neq 0$ and, hence, the Hodge structure $\bigwedge^2 A$ does not contain any Hodge class (see Proposition 2.5).

This shows that all Hodge classes of H^2 are contained in P . In particular, any hyperplane class $[\omega]$ is contained in P . On the other hand, due to iv) one has $\alpha^2 \cdot [\omega]^{2n-2} = 0$ for all $\alpha \in B_1$, but $H^2(X, \mathbb{Q})$ can clearly not contain a Hodge substructure of dimension ≥ 2 which is isotropic with respect to the polarization (see Example 1.4). This yields the contradiction. \square

3. CONSTRUCTION OF EXAMPLES

So far we have explained how Voisin is able to exclude certain Hodge structures on \mathbb{Q} -algebras from being realized by the cohomology of a projective manifold. It remains to find compact Kähler manifolds which do realize these structures and which, therefore, are topologically different from any projective manifold.

The first two examples are obtained as blow-ups of well-known Kähler manifolds and the following general facts will be used tacitly throughout (see [3, 4, 14]). Let $\pi : \tilde{X} \rightarrow X$ be the blow-up of a compact complex manifold X along a submanifold $i : Y \hookrightarrow X$ of codimension $c \geq 2$. The exceptional divisor $j : E = \pi^{-1}(Y) \hookrightarrow \tilde{X}$ is isomorphic to $\mathbb{P}(\mathcal{N}_{Y/X})$ and $\pi|_E$ equals the projection $\pi_Y : \mathbb{P}(\mathcal{N}_{Y/X}) \rightarrow Y$. In the following, cohomology will be considered with coefficients in \mathbb{Q} .

- If X is Kähler, then \tilde{X} is Kähler.
- If a submanifold $Z \subset X$ intersects Y transversally, then the proper transform, which is by definition the closure of $\pi^{-1}(Z \setminus Y)$, is the blow-up $\tilde{Z} \rightarrow Z$ along $Y \cap Z$.
- The natural morphisms $\pi^* : H^k(X) \rightarrow H^k(\tilde{X})$ and

$$H^{k-2(\ell+1)}(Y) \xrightarrow{\pi_Y^*} H^{k-2(\ell+1)}(E) \xrightarrow{\cdot h^\ell} H^{k-2}(E) \xrightarrow{j_*} H^k(\tilde{X}),$$

where $h := c_1(\mathcal{O}_{\pi_Y}(1))$, induce isomorphisms

$$H^k(\tilde{X}) \cong H^k(X) \oplus \bigoplus_{i=k-2(c-1)}^{k-2} H^i(Y).$$

In particular, $H^2(\tilde{X}) \cong H^2(X) \oplus \mathbb{Q}e$ if $e := [E] \in H^2(\tilde{X})$ and Y is connected.

- Moreover,

$$\varphi_e : H^k(X) \xrightarrow{\pi^*} H^k(\tilde{X}) \xrightarrow{\cdot e} H^{k+2}(\tilde{X})$$

equals

$$H^k(X) \xrightarrow{i^*} H^k(Y) \xrightarrow{\pi_Y^*} H^k(E) \xrightarrow{j_*} H^{k+2}(\tilde{X}).$$

In particular, $\text{Ker}(H^k(X) \xrightarrow{\varphi_e} H^{k+2}(\tilde{X})) = \text{Ker}(H^k(X) \xrightarrow{i^*} H^k(Y))$.

- If $Y = Y_1 \sqcup Y_2$ and accordingly $E = E_1 \sqcup E_2$, then for $k = 1$ the sum $\sum \text{Im}(\varphi_{e_i}) \subset H^3(\tilde{X}) = H^3(X) \oplus H^1(Y_1) \oplus H^1(Y_2)$ is direct and similar for $k = 2$ the sum $\sum \text{Im}(\varphi_{e_i}) \subset H^4(\tilde{X}) \cong H^4(X) \oplus \bigoplus H^2(Y_i) \oplus \bigoplus H^0(Y_i)$ is direct. (Note that the degree zero terms only occur if $c \geq 3$.) This principle can be generalized to the case that Y_1, Y_2 intersect transversally and that $\pi : \tilde{X} \rightarrow X$ is obtained from first blowing-up along Y_1 and then along the proper transform of Y_2 .

3.1. Voisin’s first example

Let Φ be an endomorphism of a \mathbb{Q} -vector space A of dimension $2n \geq 4$ satisfying Hypothesis 2.1. By passing to $k\Phi$ for some $0 \neq k \in \mathbb{Z}$ if necessary, we may assume that Φ^* preserves a maximal lattice $\Gamma \subset A^*$. Consider the complex torus $T := A^{1,0^*}/\Gamma$, where $A_{\mathbb{C}} = A^{1,0} \oplus A^{0,1}$ is a Hodge structure as in Example 2.8. Then there exist natural

isomorphisms $H^1(T, \mathbb{Q}) \cong A$ and $H^{1,0}(T) \cong A^{1,0}$. The endomorphism Φ^* induces an endomorphism of T which shall also be denoted Φ^* .

Remark 3.1. — The complex tori T and $T \times T$ are not projective due to Corollary 2.6, but they are, as all other complex tori, deformation equivalent and hence homeomorphic to abelian varieties.

Voisin’s first example constructed in [11] is a compact Kähler manifold X obtained as a blow-up of $T \times T$.

Consider the following submanifolds of $T \times T$:

$$\Delta_1 := \{(x, -x)\}, \Delta_2 := \{(x, -\Phi^*(x))\}, T_1 := \{0\} \times T, T_2 := T \times \{0\},$$

which meet pairwise transversally. (E.g., via the first projection the tangent space of $\Delta_1 \cap \Delta_2$ in an intersection point $z = (x, y)$ is identified with $\text{Ker}(\text{id} - \Phi^*)$, but 1 is not an eigenvalue of Φ .)

Let $z_1, \dots, z_M \in T \times T$ be the finitely many intersection points of all the pairwise intersections. Then consider the blow-up $\pi_1 : \widetilde{T \times T} \rightarrow T \times T$ in these points. The proper transforms of the four submanifolds $\widetilde{\Delta}_1, \widetilde{\Delta}_2, \widetilde{T}_1, \widetilde{T}_2$ are pairwise disjoint submanifolds of $\widetilde{T \times T}$. Thus, the blow-up $\pi_2 : X \rightarrow \widetilde{T \times T}$ along the union $\widetilde{\Delta}_1 \cup \widetilde{\Delta}_2 \cup \widetilde{T}_1 \cup \widetilde{T}_2$ is a compact Kähler manifold.

We shall denote by $F_1, \dots, F_M \subset X$ the proper transform of the exceptional divisors of π_1 and by $E_1 \rightarrow \widetilde{T}_1, E_2 \rightarrow \widetilde{T}_2, E_3 \rightarrow \widetilde{\Delta}_1, E_4 \rightarrow \widetilde{\Delta}_2$ the exceptional divisors of π_2 . Their cohomology classes shall be called $f_1, \dots, f_M, e_1, \dots, e_4 \in H^2(X, \mathbb{Q})$. It is the second blow-up π_2 and its exceptional classes e_1, \dots, e_4 that are important; the first blow-up π_1 is only needed in order to ensure the smoothness of X .

The composition $\pi := \pi_1 \circ \pi_2 : X \rightarrow T \times T$ induces a graded inclusion $\bigwedge^*(A \oplus A) = H^*(T \times T, \mathbb{Q}) \subset H^*(X, \mathbb{Q})$.

PROPOSITION 3.2. — *The conditions i)-iv) of 2.9 are satisfied.*

Proof. — The condition i) is obvious, as X and $T \times T$ are homeomorphic away from subsets of real codimension ≥ 2 . Since $H^2(T \times T, \mathbb{Q}) \cong \bigwedge^2 H^1(T \times T, \mathbb{Q})$, one has $H^2(X, \mathbb{Q}) \cong \bigwedge^2(A \oplus A) \oplus \bigoplus_{i=1}^M \mathbb{Q}f_i \oplus \bigoplus_{i=1}^4 \mathbb{Q}e_i$.

A class in $\bigwedge^{4n-2} H^1(X, \mathbb{Q}) = \bigwedge^{4n-2} H^1(T \times T, \mathbb{Q})$ can be thought of as a linear combination of fundamental classes of subsets of real codimension $4n - 2$ in $T \times T$ in general position, whose pull-back clearly avoids the exceptional divisors $F_1, \dots, F_M, E_1, \dots, E_4$ which all live over subsets of real codimension > 2 . This yields ii) with $P = \langle f_1, \dots, f_M, e_1, \dots, e_4 \rangle$ and $R = 0$.

A similar argument yields iii), where $P_1 = \langle f_1, \dots, f_M \rangle$. Finally, condition iv) is proved by applying the above general remarks on the cohomology of a blow-up and by using the explicit description of Δ_1, Δ_2, T_1 , and T_2 . □

Together with Proposition 2.11 this yields

COROLLARY 3.3. — *The rational homotopy type of the compact Kähler manifold X of dimension $2n \geq 4$ is not realized by any projective manifold.* \square

Note that this time the result has been phrased in terms of the rational homotopy type rather than in terms of the rational cohomology. Both statements are equivalent due to [2] and the fact that the fundamental group is abelian in our situation.

Remark 3.4. — One could also avoid the initial point blow-ups and instead successively blow-up $T_1, T_2, \Delta_1, \Delta_2$, respectively their proper transforms. The above arguments remain valid, only that in this case $P_1 = 0$.

In order to fully prove Theorem 0.1 it remains to construct examples of odd dimension. These are obtained as products $X' := X \times \mathbb{P}^1$, where X is one of the compact Kähler manifolds above. Once more the conditions i)-iv) of 2.9 are satisfied, but this time $R = H^2(\mathbb{P}^1, \mathbb{Q})$. The rest of the argument is unaffected by this modification.

Remark 3.5. — In [11] it is first shown that the integral cohomology $H^*(X, \mathbb{Z})$ of the above constructed Kähler manifold cannot be realized by a projective manifold. The proof of this weaker statement does not rely on Deligne’s principle, but uses the Albanese morphism instead.

One finds in [11] also an example, due to Deligne, of a compact Kähler manifold whose complex cohomology $H^*(X, \mathbb{C})$ cannot be realized by a projective manifold. The manifold X is again obtained as a blow-up of $T \times T$.

3.2. Simply-connected examples

One might wonder whether the fundamental group is responsible for the fact that the above constructed compact Kähler manifold is topologically different from any projective manifold. This question leads Voisin to her second example, which is simply-connected. Roughly, the simply-connected Kähler manifold is obtained from the first one by dividing by the $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ -action, which is induced by the standard involution on the two factors.

On the one hand, the construction is simpler in the sense that blowing-up T_1 and T_2 can be avoided, which was needed before to detect certain Hodge substructures. As it turns out, the analogous Hodge structures in the simply-connected case can be described directly. (As the examples will be simply-connected, one cannot work with Hodge structures of weight one. Therefore, Voisin analyses the weight-two Hodge structure on $H^2(X, \mathbb{Q})$ instead.) On the other hand, due to the (mild) singularities of T/\pm , the construction is slightly more involved, as we first have to desingularize.

In [11] Voisin proceeds as follows. Start with a torus $T = A^{1,0^*}/\Gamma$ as in Section 3.1. In particular, T comes with an endomorphism Φ^* . Next, consider the quotient T/\pm of T by the standard involution $z \mapsto \pm z$ and its desingularization $K \rightarrow T/\pm$ obtained by a simple blow-up of all the two-torsion points. Equivalently, one may first blow-up the two-torsion points $\tilde{T} \rightarrow T$ and then take the quotient $K = \tilde{T}/\pm$ by the induced involution. The latter description shows that K is smooth and Kähler. (Indeed, a

general result of Varouchas [10] proves that for a surjection $\pi : X \rightarrow X'$ whose fibres are all of dimension $\dim(X) - \dim(X')$ the manifold X' is Kähler if X is so.) Viewing K as the desingularization of T/\pm , shows that it is simply-connected, for T/\pm is.

The endomorphism $-\Phi^*$ of T descends to an endomorphism $-\bar{\Phi}^*$ of T/\pm and we consider its graph $\Gamma_{-\bar{\Phi}^*} \subset (T/\pm) \times (T/\pm)$.

In the last step, one first blows up $K \times K$ along the anti-diagonal $\Delta_1 := \{(a, -a)\}$ and then along the proper transform Γ' of $\Gamma_{-\bar{\Phi}^*}$. (Note that Γ' is smooth. This can be seen by passing via $\tilde{T} \times \tilde{T} \rightarrow T \times T$.)

Thus, the resulting variety X is indeed a Kähler manifold. We let $\pi : X \rightarrow K \times K$ be the composition of the two blow-ups. The two exceptional divisors $E_1 \rightarrow \Delta$ and $E_2 \rightarrow \Gamma'$ yield distinguished cohomology classes $e_1, e_2 \in H^2(X, \mathbb{Z})$.

PROPOSITION 3.6. — *Let $n \geq 3$. Then the conditions i)-v) of 2.12 are satisfied.*

Proof. — Since the involution of T acts trivially on $H^2(T, \mathbb{Q})$, one has $H^2(T/\pm, \mathbb{Q}) \cong H^2(T, \mathbb{Q}) = A$ and $H^2(K, \mathbb{Q}) = A \oplus \bigoplus \mathbb{Q}f_j$, where f_i are the classes corresponding to the exceptional divisors F_i over the two-torsion points.

Thus, $H^2(X, \mathbb{Q}) = H^2(K \times K, \mathbb{Q}) \oplus \mathbb{Q}e_1 \oplus \mathbb{Q}e_2 = H^2((T/\pm) \times (T/\pm), \mathbb{Q}) \oplus Q_1 \oplus Q_2 \oplus \mathbb{Q}e_1 \oplus \mathbb{Q}e_2$, where Q_i is the pull-back of $\bigoplus \mathbb{Q}f_j$ under the i -th projection onto K .

It is easy to see that $P := Q_1 \oplus Q_2 \oplus \mathbb{Q}e_1 \oplus \mathbb{Q}e_2$ is indeed the subspace that is annihilated by $S^{2n-1}H^2((T/\pm) \times (T/\pm), \mathbb{Q})$. This proves i).

Since $\bigwedge^2 A$ is spanned by elements a with $a^2 = 0$ and no non-trivial linear combination of $f_{1j} := \pi_1^* f_j, f_{2j} := \pi_2^* f_j, e_1$, and e_2 has this property, condition iii) follows. It is here that one needs the assumption $n \geq 3$. The verification of condition v) is straightforward; use the explicit description of the classes e_1 and e_2 .

To conclude, we have to verify condition iv). One can show that for all $\alpha \in B_1$ expressions of the form $\alpha^2 \cdot P(f_{ij}, e_1, e_2)$ with P a polynomial of degree $2n - 2$ are indeed trivial. Here are a few of the necessary arguments. Firstly, $f_{ij}^k = 0$ for all $k > n$. Secondly, the classes $f_{ij} \cdot e_k$ and $e_1 \cdot e_2$ are supported over finitely many points in $(T/\pm) \times (T/\pm)$ and, hence as α is pulled back from there, one has $\alpha \cdot (f_{ij} \cdot e_k) = \alpha \cdot (e_1 \cdot e_2) = 0$. Thirdly, $\alpha \cdot f_{1j} = 0$. Thus, the only combinations that need to be checked are $\alpha^2 \cdot e_i^{2n-2}$. We may assume that $E_i = \mathbb{P}(\Omega_T)$ and that $\pi|_{E_i}$ is the natural projection $p : E \rightarrow T$. Then one shows that $e_i|_{E_i} = c_1(\mathcal{O}_p(-1))$ and thus reduces to $0 = p^* \alpha_T^2 \cdot c_1(\mathcal{O}_p(-1))^{2n-3}$, which follows from $c_1(\mathcal{O}_p(-1))^k = 0$ for $k \geq n$ and the assumption $n \geq 3$. □

Together with Proposition 2.13 this yields

COROLLARY 3.7. — *The rational homotopy type of the compact simply-connected Kähler manifold X of dimension $2n \geq 6$ is not realized by any projective manifold.* □

Odd-dimensional examples can again be produced by taking products with \mathbb{P}^1 . In 2.12 only i) and iii) have to be modified. In i) one has $H^2 = B_1 \oplus B_2 \oplus P \oplus R$ with $R = H^2(\mathbb{P}^1, \mathbb{Q})$ and in iii) $R_{\mathbb{C}}$ will provide another irreducible component. The

arguments are not affected by this modification. This yields C. Voisin’s second counter-example:

THEOREM 3.8 ([11]). — *In any dimension ≥ 6 there exists a simply-connected compact Kähler manifold which does not have the rational homotopy type of a projective manifold.*

Once more, instead of working with the rational homotopy type one could equivalently say that $H^*(X, \mathbb{Q})$ is not realized as the cohomology ring of a projective manifold (see [2]).

Remark 3.9. — Inspired by Voisin’s examples, Oguiso studies in [9] simply-connected compact Kähler manifolds of dimension $d \geq 4$ which are not projective, but rigid, i.e. which do not allow any deformations at all and, therefore, cannot be deformed to projective ones in particular. In the case of simply-connected examples one can no longer work with Hodge structures of weight one. Thus, K3 surfaces (or, more generally, compact hyperkähler manifolds) with their very special but rich Hodge structures of weight two provide a rich reservoir of potentially interesting examples. Roughly, the special endomorphisms of tori used by Voisin are in [9] replaced by special automorphisms of K3 surfaces which are described completely by their action on the second cohomology.

However, the methods in [9] fall short of proving that the examples do not have the rational homotopy type of projective manifolds. It seems likely, nevertheless, that four-dimensional simply-connected examples could eventually be produced in this way.

3.3. The birational Kodaira problem

Right after [11] had appeared modifications of the original problem have been proposed. For many problems in complex algebraic geometry it is natural not to restrict to projective or Kähler manifolds, but to allow manifolds that are birational or bimeromorphic to those. Passing to a bimeromorphic model often changes the topology drastically, but in a somewhat controlled manner. So, modifying Kodaira’s problem in this sense seems natural also from a topological point of view.

More precisely, the compact Kähler manifolds constructed in [11] are both bimeromorphic to compact Kähler manifolds which do have the homotopy type of projective manifolds. E.g. in the first example, described in Section 3.1, the Kähler manifold X was constructed as a blow-up of a torus whose underlying manifold carries also the structure of a projective manifold. In other words, after a controlled topological modification the original topological manifold underlying X has been transformed to one that does carry a projective structure. So, one could ask whether this is true for any Kähler manifold. Again, the answer is negative.

THEOREM 3.10 ([12]). — *There exist compact Kähler manifolds X of dimension $2n \geq 10$ such that no complex manifold bimeromorphic to it has the rational homotopy type of a projective manifold.*

The principal ideas in [12] are similar to those in [11]. Roughly, one tries to detect certain Hodge structures in terms of the multiplicative structure of the cohomology ring and to derive a contradiction to the existence of a polarization on the (primitive) second cohomology of a projective manifold. Technically, the arguments are more involved and we only give an idea of the actual construction.

The construction of the birational counter-examples in [12] starts again with the same torus T of dimension $n \geq 4$ and an endomorphism Φ satisfying 2.1. If \mathcal{P} denotes the Poincaré bundle on $T \times \widehat{T}$, then let $E := \mathcal{P} \oplus \mathcal{P}^{-1}$, $\mathcal{P}_\Phi := (\Phi, \text{id})^* \mathcal{P}$, and $E_\Phi := (\Phi, \text{id})^* E$. In the next step one considers the fibre product $\mathbb{P}(E) \times_{T \times \widehat{T}} \mathbb{P}(E_\Phi)$ and its quotient Q by the action of $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ given by natural lifts of $(-\text{id}, \text{id})$ and $(\text{id}, -\text{id})$. Then any Kähler desingularization X of Q will work. Note that these examples are bimeromorphic to a $\mathbb{P}^1 \times \mathbb{P}^1$ -bundle over $K \times \widehat{K}$, where $K \rightarrow T/\pm$ is the desingularization considered in the simply-connected case.

The reason that one is able to control in this example all bimeromorphic models by cohomological methods is due to the fact that there exist only few subvarieties of positive dimension.

4. FURTHER COMMENTS

This is still not the end. Why not allowing topological changes that are not obtained by bimeromorphic maps? One could ask whether there always exists another complex structure on X (e.g. one obtained by a deformation) such that a bimeromorphic model of this new one has the rational homotopy type of a projective manifold. So, more formally, if one introduces the equivalence relation between complex manifolds generated by deformations and bimeromorphic correspondences, one might ask whether any compact Kähler manifold is equivalent to a projective manifold.

Continuing in this direction, one could allow singular varieties or certain ramified covers in order to enlarge the equivalence classes. Would the answer to Kodaira's problem be different then? Most of these questions are open for the time being, but see the comments in [13].

In another direction, it could be interesting to see whether the birational geometry does matter in these questions. The above counter-example for the birational Kodaira problem is, by construction, of Kodaira dimension $-\infty$. For the time being the techniques do not seem to produce examples of non-negative Kodaira dimension.

As has been mentioned, topologically there is no difference between compact Kähler surfaces and projective surfaces. Due to the examples of Voisin, the situation changes drastically in dimension ≥ 4 (or rather ≥ 6 if one prefers simply-connected manifolds). What seems open, however, is the three-dimensional case:

Does there exist a compact Kähler threefold which is not homeomorphic to a projective manifold?

Since we mentioned fundamental groups in the beginning, let us point out that the following problem is also still open:

Does there exist a group that is the fundamental group of a compact Kähler manifold, but not of a projective manifold?

A question of a more general nature is the following:

Are there topological, cohomological,... conditions that decide whether a compact Kähler manifold can also be endowed with a complex structure which is projective?

Nothing seems to be known in this direction and the examples show that if such conditions can be found at all, they cannot be formulated purely in terms of the fundamental group.

REFERENCES

- [1] M. Artin – *Algebra*. Prentice Hall (1991).
- [2] P. Deligne, Ph. Griffiths, J. Morgan, D. Sullivan – *Real homotopy theory of Kähler manifolds*. *Invent. math.* 2 (1975), 245-274.
- [3] J.-P. Demailly – *Complex analytic and algebraic geometry*. <http://www-fourier.ujf-grenoble.fr/~demailly/books.html>.
- [4] Ph. Griffiths, J. Harris – *Principles of algebraic geometry*. Wiley (1978).
- [5] H. Hironaka – *An example of a non-Kählerian complex-analytic deformation of Kählerian complex structures*. *Ann. Math.* 75 (1962), 190-208.
- [6] K. Kodaira – *On Kähler metrics of restricted type*. *Ann. of Math.* 60 (1954), 28-48.
- [7] K. Kodaira – *On compact analytic surfaces*. *Analytic Functions*, Princeton University Press (1960), 121-135.
- [8] K. Kodaira – *On compact complex analytic surfaces, I*. *Ann. of Math.* 71 (1960), 111-152.
- [9] K. Oguiso – *Bimeromorphic automorphism groups of non-projective hyperkähler manifolds - a note inspired by C.T. McMullen*. [math.AG/0312515](https://arxiv.org/abs/math/0312515).
- [10] J. Varouchas – *Stabilité de la classe des variétés kählériennes par certains morphismes propres*. *Invent. Math.* 77 (1984), 117-127.
- [11] C. Voisin – *On the homotopy types of compact Kähler and complex projective manifolds*. *Invent. Math.* 157 (2004), 329-343.
- [12] C. Voisin – *On the homotopy types of Kähler manifolds and the birational Kodaira problem*. *J. Diff. Geom.*, to appear.
- [13] C. Voisin – *Hodge theory and the topology of compact Kähler and complex projective manifolds*. Lecture Notes for the Seattle AMS Summer Institute

- [14] C. Voisin – *Théorie de Hodge et géométrie algébrique complexe*. Cours spécialisés. SMF (2002).

Daniel HUYBRECHTS

Universität Bonn

Mathematisches Institut

Beringstraße 1

D–53115 BONN

(Allemagne)

E-mail : `huybrech@math.uni-bonn.de`

GENRES DE TODD ET VALEURS AUX ENTIERS DES DÉRIVÉES DE FONCTIONS L

par **Christophe SOULÉ**

Le théorème de Hirzebruch-Riemann-Roch calcule la caractéristique d'Euler-Poincaré d'un fibré holomorphe E sur une variété complexe X projective et lisse :

$$(1) \quad \chi(E) = \int_X \text{ch}(E) \text{Td}(TX).$$

Dans cette identité le genre de Todd $\text{Td}(\cdot)$ est la classe caractéristique multiplicative associée à la série formelle

$$\text{Td}(x) = \frac{x}{1 - e^{-x}} = 1 + \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} + \dots,$$

que l'on peut aussi écrire

$$\text{Td}(x) = 1 - \sum_{m \geq 0} \zeta(-m) \frac{x^{m+1}}{m!},$$

où $\zeta(s)$ est la fonction zêta de Riemann. Ce lien entre le genre de Todd et les valeurs aux entiers de la fonction zêta peut paraître fortuit. Mais l'analogue de l'égalité (1) en géométrie d'Arakelov fait intervenir, en plus du genre de Todd classique, la classe caractéristique additive associée à la série

$$R(x) = \sum_{\substack{m \geq 1 \\ m \text{ impair}}} \left(2\zeta'(-m) + \left(1 + \frac{1}{2} + \dots + \frac{1}{m} \right) \zeta(-m) \right) \frac{x^m}{m!},$$

où $\zeta'(s)$ est la *dérivée* de la fonction zêta. Et si les nombres rationnels $\zeta(-m)$ (autrement dit les nombres de Bernoulli) apparaissent dans de nombreux calculs, il est par contre très rare de rencontrer les nombres réels $\zeta'(-m)$ avec m impair. Le but de cet exposé est de montrer comment, plus généralement, les avatars du genre de Todd en géométrie d'Arakelov permettent d'obtenir des formules, nouvelles ou déjà connues, impliquant les valeurs aux entiers des dérivées des fonctions L de la théorie des nombres. Si X est un schéma régulier, projectif sur un ouvert du spectre de l'anneau des entiers d'un corps de nombres, muni d'une action du schéma en groupes G des racines n -ièmes de l'unité, $n \geq 1$, Köhler et Roessler ont démontré un analogue équivariant du théorème de Riemann-Roch en géométrie d'Arakelov. Ce «théorème de Lefschetz arithmétique» [23] comporte aussi une correction au genre de Todd équivariant habituel. Ses coefficients

sont donnés par les valeurs aux entiers négatifs de la *dérivée* en s de la fonction zêta de Lerch définie, si $\operatorname{Re}(s) > 1$, par la formule

$$\zeta(z, s) = \sum_{m \geq 1} \frac{z^m}{m^s}, \quad z \in \mathbb{C}, \quad |z| = 1.$$

Si χ est un caractère de Dirichlet modulo n , cette fonction est liée à la fonction L de Dirichlet

$$L(\chi, s) = \sum_{m \geq 1} \frac{\chi(m)}{m^s}$$

par une transformée de Fourier sur le groupe $(\mathbb{Z}/n)^*$. Par ailleurs, une formule célèbre de Chowla et Selberg (étendue par Gross [19], Anderson [1] et Colmez [11] aux variétés abéliennes de type CM) calcule les périodes d'une courbe elliptique à multiplication complexe à l'aide des dérivées à l'origine des fonctions L de Dirichlet [12].

Köhler et Roessler ont montré que leur théorème de Lefschetz arithmétique fournit une nouvelle preuve des formules de Chowla-Selberg, Gross, Anderson et Colmez (sans retrouver cependant le calcul complet aux «mauvaises places»). Maillot et Roessler [30] ont abordé le cas d'une variété quelconque sur un corps de nombres, munie d'une action de G . Outre les résultats précédents, ils obtiennent un calcul des périodes du groupe $H^2(X)$ quand X est une surface, et $H^d(X)$ quand X est une hypersurface de dimension d . Leur résultat est la première confirmation, en dehors des variétés abéliennes, d'une conjecture de Gross et Deligne sur les motifs à multiplication complexe [19].

Après avoir introduit la géométrie d'Arakelov (§1) et énoncé les théorèmes de Riemann-Roch (§2, Th. 2.1) et de Lefschetz arithmétiques (§3, Th. 3.1), nous montrerons, en suivant [29], que cette formule se simplifie énormément quand on l'applique au complexe de De Rham d'une variété projective et lisse sur un corps de nombres X , munie d'une action de G . Le résultat principal (§4, Th. 4.4) est une réécriture de la transformée de Fourier de l'identité initiale. Il affirme qu'une certaine combinaison de logarithmes de périodes de X est un multiple entier explicite de $\frac{L'(\chi, 0)}{L(\chi, 0)}$, où χ est un caractère de Dirichlet impair et primitif modulo n . La section 5 détaille ce calcul, et on discute dans la section 6 le cas d'une variété abélienne. On fait alors le lien avec la conjecture de Gross et Deligne et avec la formule de Chowla-Selberg. Il faut toutefois signaler que les égalités ainsi obtenues ne sont pas aussi précises qu'on le souhaiterait. Elles souffrent en effet d'une ambiguïté, due au fait que le complexe de De Rham ne s'étend pas en général, de façon équivariante, à un modèle entier de X . L'appendice donne par contre un exemple d'une courbe elliptique de type CM pour laquelle la méthode fournit un résultat sans cette ambiguïté.

Pour terminer, nous évoquerons brièvement d'autres travaux reliant la géométrie d'Arakelov aux valeurs des dérivées de fonctions L , y compris celles associées aux formes modulaires [30] [25] [26]. C'est un domaine en plein essor.

Je tiens à remercier Maillot et Roessler pour m'avoir beaucoup aidé à préparer cet exposé. L'exemple traité en appendice leur est dû. Je remercie aussi Burgos et Kudla pour leurs commentaires sur ce manuscrit.

Notation : si M est un groupe abélien et si K est un corps, on notera M_K le K -espace vectoriel $M \otimes_{\mathbb{Z}} K$.

1. GÉOMÉTRIE D'ARAKELOV

Dans cette section et la suivante nous décrivons les principales notions de la géométrie d'Arakelov. Pour des exposés de synthèse plus détaillés, voir [36] [35] [4] et l'exposé [7] de ce séminaire.

1.1. Appelons *variété arithmétique* la donnée d'un schéma X régulier et projectif sur $S = \text{Spec}(\mathbb{Z})$. La conjugaison des coordonnées munit l'ensemble $X(\mathbb{C})$ des points complexes de X d'une involution $F_{\infty} : X(\mathbb{C}) \rightarrow X(\mathbb{C})$. Pour un entier $p \geq 0$ on note $Z^p(X)$ le groupe des cycles algébriques de codimension p sur X , i.e. les combinaisons formelles finies $\sum_{\alpha} n_{\alpha} Z_{\alpha}$, $n_{\alpha} \in \mathbb{Z}$, où les $Z_{\alpha} \subset X$ sont des fermés de codimension p dans X . On note aussi $D^{pp}(X_{\mathbb{R}})$ (resp. $A^{pp}(X_{\mathbb{R}})$) l'espace vectoriel réel des courants réels (resp. des formes différentielles réelles) de type (p, p) sur $X(\mathbb{C})$ sur lequel(le)s F_{∞}^* agit par multiplication par $(-1)^p$. Un *courant de Green* pour le cycle $Z \in Z^p(X)$ est un élément g de $D^{p-1, p-1}(X_{\mathbb{R}})$ tel que

$$(2) \quad dd^c g + \delta_Z = \omega,$$

où $\omega \in A^{pp}(X_{\mathbb{R}})$, $\delta_Z \in D^{pp}(X_{\mathbb{R}})$ est le courant d'intégration sur les points complexes de Z , et $dd^c = \frac{i}{2\pi} \partial \bar{\partial}$. Le *groupe de Chow arithmétique* $\widehat{\text{CH}}^p(X)$ est engendré par les couples (Z, g) , où $Z \in Z^p(X)$ et g est un courant de Green de Z . On a $(Z, g) + (Z', g') = (Z + Z', g + g')$ et on impose les relations

$$(\text{div}(f), -\log |f|^2 + \partial u + \bar{\partial} v) = 0$$

dans $\widehat{\text{CH}}^p(X)$, où u (resp. v) est un courant de type $(p-2, p-1)$ (resp. $(p-1, p-2)$) et $f \in k(Y)^*$ est n'importe quelle fonction rationnelle non nulle sur un sous-schéma fermé intègre $Y \subset X$ de codimension $p-1$. Le cycle $\text{div}(f)$ est le diviseur de f et $-\log |f|^2$ est le courant obtenu en associant à toute forme différentielle sur $X(\mathbb{C})$ l'intégrale sur $Y(\mathbb{C})$ de son produit avec la fonction intégrable $-\log |f|^2$.

À tout morphisme algébrique $f : X \rightarrow Y$ entre variétés arithmétiques sont associés des morphismes d'image inverse

$$f^* : \widehat{\text{CH}}^p(Y) \rightarrow \widehat{\text{CH}}^p(X).$$

De plus, il existe un produit d'*intersection arithmétique*

$$\widehat{\text{CH}}^p(X) \otimes \widehat{\text{CH}}^q(X) \rightarrow \widehat{\text{CH}}^{p+q}(X)_{\mathbb{Q}}.$$

Ce produit est compatible aux images inverses. On notera qu'au lieu de $\widehat{\text{CH}}^{p+q}(X)_{\mathbb{Q}}$ on peut aussi considérer, et c'est plus naturel, le \mathbb{Q} -espace vectoriel défini de la même façon que $\widehat{\text{CH}}^{p+q}(X)$ mais en prenant pour générateurs les couples (Z, g) , où $Z \in Z^{p+q}(X)_{\mathbb{Q}}$

et $g \in D^{p+q-1, p+q-1}(X_{\mathbb{R}})$ est un courant de Green de Z . De même pour les espaces vectoriels $\widehat{\text{CH}}^p(X)_K$ utilisés plus loin.

1.2. Un *fibré hermitien* sur X est la donnée d'un couple $\bar{E} = (E, h)$, où E est un fibré algébrique sur X et h une métrique hermitienne C^∞ sur le fibré holomorphe $E_{\mathbb{C}}$ sur $X(\mathbb{C})$ associé à E , cette métrique étant invariante par F_∞ . On peut associer à tout fibré hermitien \bar{E} des classes caractéristiques telles que les classes de Chern $\hat{c}_p(\bar{E}) \in \widehat{\text{CH}}^p(X)$, le caractère de Chern

$$\widehat{\text{ch}}(\bar{E}) \in \widehat{\text{CH}}^*(X)_{\mathbb{Q}} = \bigoplus_{p \geq 0} \widehat{\text{CH}}^p(X)_{\mathbb{Q}}$$

et la classe de Todd

$$\widehat{\text{Td}}(\bar{E}) \in \widehat{\text{CH}}^*(X)_{\mathbb{Q}}.$$

Si par exemple $\bar{L} = \det(\bar{E})$ est la puissance extérieure maximale de E , la première classe de Chern

$$\hat{c}_1(\bar{E}) = \hat{c}_1(\bar{L}) \in \widehat{\text{CH}}^1(X)$$

est la classe du couple $(\text{div}(s), -\log \|s\|^2)$, où s est n'importe quelle section rationnelle non nulle de s sur X et $\|s\|$ la norme de cette section.

Ces classes caractéristiques vérifient les propriétés usuelles de functorialité, normalisation et comportement par produit tensoriel. Si $E = E' \oplus E''$ est la somme directe de deux fibrés algébriques et si la métrique sur $E_{\mathbb{C}}$ est la somme directe orthogonale des métriques sur $E'_{\mathbb{C}}$ et $E''_{\mathbb{C}}$ on a

$$\hat{c}_p(\bar{E}) = \sum_{i+j=p} \hat{c}_i(\bar{E}') \hat{c}_j(\bar{E}''),$$

$$\widehat{\text{ch}}(\bar{E}) = \widehat{\text{ch}}(\bar{E}') \widehat{\text{ch}}(\bar{E}''),$$

et

$$\widehat{\text{Td}}(\bar{E}) = \widehat{\text{Td}}(\bar{E}') \widehat{\text{Td}}(\bar{E}'').$$

Mais ces formules ne sont plus valables en général pour une suite exacte

$$0 \rightarrow E' \rightarrow E \rightarrow E'' \rightarrow 0,$$

et ce quel que soit le choix des métriques sur les trois fibrés.

2. THÉORÈME DE RIEMANN-ROCH ARITHMÉTIQUE

2.1. Soit X une variété arithmétique. On définit comme suit une application \mathbb{Q} -linéaire d'intégration sur X

$$\int_X : \widehat{\text{CH}}^*(X)_{\mathbb{Q}} \rightarrow \mathbb{R}.$$

L'image de la classe α d'un couple (Z, g) est nulle sauf si $Z = \sum_x n_x x$ est un cycle de dimension zéro et g est un courant de degré maximum sur $X(\mathbb{C})$, auquel cas

$$\int_X \alpha = \sum_x n_x \log(\#k(x)) + \frac{1}{2} \int_{X(\mathbb{C})} g.$$

On a noté $\#k(x)$ le cardinal du corps résiduel au point fermé $x \in X$ (un corps fini), et défini de g comme étant celle d'une forme cohomologue à ce courant. Le morphisme

$$\int_S : \widehat{\text{CH}}^1(S) \rightarrow \mathbb{R}$$

est un isomorphisme.

2.2. Soit M un \mathbb{Z} -module de type fini et h un produit scalaire hermitien sur $M_{\mathbb{C}}$, invariant par la conjugaison complexe. On pose

$$\widehat{\text{deg}}(\bar{M}) = \int_S \hat{c}_1(\det(\bar{M})).$$

Si M_{tors} est le sous-groupe de torsion de M et si $M_{\mathbb{R}}$ est muni de la mesure euclidienne définie par h , on a

$$\widehat{\text{deg}}(\bar{M}) = \log(\#M_{\text{tors}}) - \log \text{vol} \left(\frac{M_{\mathbb{R}}}{M} \right).$$

2.3. Soit X une variété arithmétique et h_X une métrique hermitienne sur le fibré tangent $TX(\mathbb{C})$, invariante par F_{∞} . Notons $\omega_0 \in A^{11}(X_{\mathbb{R}})$ la forme telle que

$$\omega_0 = \frac{i}{2\pi} \sum_{\alpha, \beta} h_X \left(\frac{\partial}{\partial z_{\alpha}}, \frac{\partial}{\partial z_{\beta}} \right) dz_{\alpha} d\bar{z}_{\beta}$$

pour tout choix d'une carte locale (z_{α}) sur $X(\mathbb{C})$. On suppose que h_X est Kähler, c'est-à-dire $d\omega_0 = 0$.

Si \bar{E} est un fibré hermitien sur X , les groupes de cohomologie $H^q(X, E)$ sont de type fini. Pour tout entier $q \geq 0$, l'espace vectoriel complexe

$$H^q(X, E)_{\mathbb{C}} = H^q(X(\mathbb{C}), E_{\mathbb{C}})$$

est canoniquement isomorphe à celui des formes différentielles harmoniques de type $(0, q)$ sur $E_{\mathbb{C}}$. On note $A^{0q}(X(\mathbb{C}), E_{\mathbb{C}})$ l'espace vectoriel de toutes les différentielles C^{∞} de type $(0, q)$ à coefficients dans $E_{\mathbb{C}}$. Il est muni du produit scalaire hermitien h_{L^2} défini par

$$h_{L^2}(\alpha, \beta) = \int_{X(\mathbb{C})} (\alpha, \beta) \frac{\omega_0^d}{d!},$$

où $d = \dim_{\mathbb{C}} X(\mathbb{C})$ et (α, β) désigne le produit scalaire ponctuel associé au choix des métriques sur $X(\mathbb{C})$ et $E_{\mathbb{C}}$. La restriction de h_{L^2} aux formes harmoniques définit un produit scalaire hermitien sur $H^q(X, E)_{\mathbb{C}}$.

2.4. Soient

$$\bar{\partial} : A^{0,q}(X(\mathbb{C}), E_{\mathbb{C}}) \rightarrow A^{0,q+1}(X(\mathbb{C}), E_{\mathbb{C}})$$

l'opérateur de Cauchy-Riemann et $\bar{\partial}^*$ son adjoint pour la métrique L^2 . L'opérateur de Laplace $\Delta_q = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial}$ a un spectre discret formé de nombres réels positifs ou nuls. On note $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots$ ses valeurs propres strictement positives (répétées selon leur multiplicité) et

$$\zeta_q(s) = \sum_{n \geq 1} \lambda_n^{-s}$$

sa fonction zêta. Cette série converge si $\operatorname{Re}(s) > d$, elle admet un prolongement méromorphe au plan complexe, et n'a pas de pôle à l'origine ; on peut donc prendre sa dérivée en zéro. On pose

$$T(\bar{E}_{\mathbb{C}}) = \sum_{q \geq 0} (-1)^{q+1} q \zeta'_q(0)$$

(cf. [31]).

2.5. Si $E_{\mathbb{C}}$ est un fibré holomorphe sur $X(\mathbb{C})$ invariant par F_{∞} , on note $\operatorname{ch}(E_{\mathbb{C}})$ et $\operatorname{Td}(E_{\mathbb{C}})$ ses classes caractéristiques usuelles dans $\bigoplus_{p \geq 0} H^{pp}(X_{\mathbb{R}})$, où $H^{pp}(X_{\mathbb{R}})$ est le sous-espace de la cohomologie réelle de $X(\mathbb{C})$ de type (p, p) où F_{∞} agit par $(-1)^p$. On définit aussi une classe

$$R(E_{\mathbb{C}}) \in \bigoplus_{p \geq 0} H^{pp}(X_{\mathbb{R}}),$$

compatible aux images inverses et additive :

$$R(E'_{\mathbb{C}} \oplus E''_{\mathbb{C}}) = R(E'_{\mathbb{C}}) + R(E''_{\mathbb{C}}).$$

Si $L_{\mathbb{C}}$ est un fibré inversible et $x = c_1(L_{\mathbb{C}}) \in H^{11}(X_{\mathbb{R}})$ sa première classe de Chern on a $R(L_{\mathbb{C}}) = R(x)$, où $R(x)$ est la série formelle

$$R(x) = \sum_{\substack{m \text{ impair} \\ m \geq 1}} \left(2\zeta'(-m) + \left(1 + \frac{1}{2} + \dots + \frac{1}{m} \right) \zeta(-m) \right) \frac{x^m}{m!}.$$

2.6. Dans la situation de 2.2 on pose

$$\hat{\chi}(\bar{E}) = \sum_{q \geq 0} (-1)^q \widehat{\operatorname{deg}}(H^q(X, E), h_{L^2}).$$

Le théorème de Riemann-Roch arithmétique [18] [7] (dans sa version « à la Hirzebruch ») est le suivant :

THÉORÈME 2.1 ([18], Th. 7). — *On a l'égalité entre nombres réels*

$$\hat{\chi}(\bar{E}) - \frac{1}{2} T(\bar{E}_{\mathbb{C}}) = \int_X \widehat{\operatorname{ch}}(\bar{E}) \widehat{\operatorname{Td}}(X) - \frac{1}{2} \int_{X(\mathbb{C})} \operatorname{ch}(E_{\mathbb{C}}) \operatorname{Td}(TX(\mathbb{C})) R(TX(\mathbb{C})).$$

Remarques. —

- i) Si X est lisse sur \mathbb{Z} , $\widehat{\text{Td}}(X)$ est la classe de Todd $\widehat{\text{Td}}(TX, h_X)$. En général on la définit à l'aide du complexe tangent à X (cf. op. cit.). Elle dépend de la métrique h_X .
- ii) La preuve du théorème 2.1 repose pour une large part sur les travaux d'analyse globale de Bismut et de ses collaborateurs, dont l'article [5].

3. THÉORÈME DE LEFSCHETZ ARITHMÉTIQUE

3.1. Soit $n > 1$ un entier et $G = \text{Spec}(\mathbb{Z}[T]/(T^n - 1))$ le schéma en groupes des racines n -ièmes de l'unité. Supposons donnée une action

$$\mu : G \times X \rightarrow X$$

de G sur une variété arithmétique X . On note $Y = X^G$ le schéma des points fixes de G . C'est aussi une variété arithmétique. Soit $\bar{E} = (E, h)$ un fibré hermitien sur X tel que l'action de G se prolonge à E et préserve la métrique h . L'action de G sur la restriction de E à Y équivaut à la donnée d'une graduation

$$E|_Y = \bigoplus_{u \in \mathbb{Z}/n} E_u$$

([16], Prop. 4.7.3). Si V est un ouvert de Y , l'action de G fournit en effet un morphisme de $\mathcal{O}_Y(V)$ -modules :

$$\mu^* : E(V) \rightarrow E(V) \otimes_{\mathbb{Z}} \mathbb{Z}[T]/(T^n - 1).$$

Si $s \in E(V)$ on pose

$$\mu^*(s) = \sum_{u \in \mathbb{Z}/n} s_u \otimes T^u.$$

Le module $E_u(V) \subset E(V)$, $u \in \mathbb{Z}/n$, est celui des composantes s_u quand s décrit $E(V)$.

Les composantes $E_{u, \mathbb{C}}$ de la restriction de $E_{\mathbb{C}}$ à $Y(\mathbb{C})$ sont orthogonales pour la métrique h . On les munit de la métrique induite par h .

3.2. On fixe désormais une racine n -ième primitive de l'unité $\gamma \in \mathbb{C}^*$ et l'on note $g \in G(\mathbb{C})$ l'élément correspondant. Fixons une métrique de Kähler h_X sur $TX(\mathbb{C})$, invariante par G et F_{∞} . Pour tout entier $q \geq 0$, le G -module $H^q(X, E)$ définit une graduation

$$H^q(X, E) = \bigoplus_{u \in \mathbb{Z}/n} H^q(X, E)_u,$$

orthogonale pour la métrique L^2 . On pose

$$\widehat{\text{deg}}_g(H^q(X, E)) = \sum_{u \in \mathbb{Z}/n} \widehat{\text{deg}}(H^q(X, E)_u, h_{L^2}) \gamma^u$$

et

$$\hat{\chi}_g(\bar{E}) = \sum_{q \geq 0} (-1)^q \widehat{\text{deg}}_g(H^q(X, E)).$$

On a aussi une décomposition orthogonale

$$A^{0q}(X(\mathbb{C}), E_{\mathbb{C}}) = \bigoplus_{u \in \mathbb{Z}/n} A^{0q}(X(\mathbb{C}), E_{\mathbb{C}})_u$$

et, si $\zeta_{q,u}(s)$ est la fonction zêta de la restriction de Δ_q à $A^{0q}(X(\mathbb{C}), E_{\mathbb{C}})_u$, on pose

$$T_g(\bar{E}_{\mathbb{C}}) = \sum_{q \geq 0} (-1)^{q+1} q \sum_{u \in \mathbb{Z}/n} \zeta'_{q,u}(0) \gamma^u.$$

Les nombres $\hat{\chi}_g(\bar{E})$ et cette *torsion analytique équivariante* $T_g(\bar{E}_{\mathbb{C}})$ sont des nombres complexes.

3.3. Soit $K = \mathbb{Q}(\gamma) \subset \mathbb{C}$ le corps cyclotomique des racines n -ièmes de l'unité. Le caractère de Chern équivariant de la restriction de \bar{E} à Y est

$$\widehat{\text{ch}}_g(\bar{E}) = \sum_{u \in \mathbb{Z}/n} \widehat{\text{ch}}(\bar{E}_u) \gamma^u \in \widehat{\text{CH}}(Y)_K.$$

Désignons par $\Lambda^k(\bar{E})$, $k \geq 0$, les puissances extérieures de \bar{E} et posons

$$\lambda_{-1}(\bar{E}) = \sum_{k \geq 0} (-1)^k \Lambda^k(\bar{E}).$$

C'est un fibré hermitien virtuel, c'est-à-dire un élément du groupe abélien associé au monoïde des fibrés hermitiens, muni de la somme directe orthogonale. On note

$$\text{ch}_g(\lambda_{-1}(\bar{E})) = \sum_{k \geq 0} (-1)^k \widehat{\text{ch}}_g(\Lambda^k(\bar{E})).$$

Soit \bar{N}^\vee le fibré conormal à Y dans X , muni de la métrique induite par h_X . On munit $TY(\mathbb{C})$ de la métrique h_Y induite par h_X et l'on note $\widehat{\text{Td}}(Y) \in \widehat{\text{CH}}(Y)_{\mathbb{Q}}$ la classe de Todd arithmétique de (Y, h_Y) (cf. 2.6, Remarque i)). On pose

$$\widehat{\text{Td}}_g(X) = \widehat{\text{ch}}_g(\lambda_{-1}(\bar{N}^\vee))^{-1} \widehat{\text{Td}}(Y)$$

dans $\widehat{\text{CH}}(Y)_K$.

On définit de même des classes

$$\text{ch}_g(E_{\mathbb{C}}) = \sum_{u \in \mathbb{Z}/n} \text{ch}(E_{u,\mathbb{C}}) \gamma^u$$

et

$$\text{Td}_g(TX(\mathbb{C})) = \text{ch}_g(\lambda_{-1}(\bar{N}_{\mathbb{C}}^\vee))^{-1} \text{Td}(TY(\mathbb{C}))$$

dans $\bigoplus_{p \geq 0} H^{pp}(Y_{\mathbb{R}})_K$.

3.4. La genre R possède également un analogue équivariant. Pour tout nombre complexe z de module un, considérons la fonction zêta de Lerch

$$\zeta(z, s) = \sum_{m \geq 1} \frac{z^m}{m^s}.$$

Cette série converge si $\operatorname{Re}(s) > 1$, et admet un prolongement méromorphe au plan complexe. Si $\zeta'(z, s)$ est sa dérivée par rapport à s , on introduit les séries formelles

$$\tilde{R}(z, x) = \sum_{m \geq 0} \left(2\zeta'(z, -m) + \left(1 + \frac{1}{2} + \cdots + \frac{1}{m} \right) \zeta(z, -m) \right) \frac{x^m}{m!}$$

et

$$R(z, x) = \frac{1}{2} (\tilde{R}(z, x) - \tilde{R}(\bar{z}, -x)).$$

On note $R(z, \cdot)$ la classe caractéristique additive en cohomologie complexe telle que, si $L_{\mathbb{C}}$ est un fibré holomorphe inversible on ait

$$R(z, L_{\mathbb{C}}) = R(z, c_1(L_{\mathbb{C}})).$$

La restriction de $TX(\mathbb{C})$ à $Y(\mathbb{C})$ admet une graduation

$$TX(\mathbb{C})|_{Y(\mathbb{C})} = \bigoplus_{u \in \mathbb{Z}/n} TX(\mathbb{C})_u,$$

et l'on pose

$$R_g(TX(\mathbb{C})) = \sum_{u \in \mathbb{Z}/n} R(\gamma^u, TX(\mathbb{C})_u).$$

3.5.

THÉORÈME 3.1 ([23], Th. 7.14). — *Sous les hypothèses précédentes, on a l'égalité entre nombres complexes*

$$\hat{\chi}_g(\bar{E}) - \frac{1}{2} T_g(\bar{E}_{\mathbb{C}}) = \int_Y \widehat{\operatorname{ch}}_g(\bar{E}) \widehat{\operatorname{Td}}_g(X) - \frac{1}{2} \int_{Y(\mathbb{C})} \operatorname{ch}_g(E_{\mathbb{C}}) \operatorname{Td}_g(TX(\mathbb{C})) R_g(TX(\mathbb{C})).$$

Remarques. —

- i) La preuve du théorème 3.1 utilise à nouveau des résultats analytiques difficiles de Bismut et de ses collaborateurs [2][6].
- ii) Les constructions et les résultats précédents restent valables si, au lieu de $S = \operatorname{Spec}(\mathbb{Z})$, on prend pour base S des variétés arithmétiques un ouvert dans le spectre des entiers d'un corps de nombres. La seule différence est que les théorèmes 2.1 et 3.1 sont alors des égalités dans les groupes $\widehat{\operatorname{CH}}^1(S)_{\mathbb{Q}}$ et $\widehat{\operatorname{CH}}^1(S)_K$ respectivement.

Si par exemple F est un corps de nombres et que l'on fixe un plongement complexe $F \subset \mathbb{C}$, on pourra prendre $S = \operatorname{Spec}(F)$ et noter $X(\mathbb{C})$ les points complexes pour le plongement choisi d'une variété X projective et lisse sur F . L'involution F_{∞} agit trivialement sur $X(\mathbb{C})$ et l'intégration \int_X associée à un couple $(0, g)$ de dimension zéro l'intégrale de $\frac{1}{2} g$ sur $X(\mathbb{C})$. Si $\mathcal{S} \subset \mathbb{R}$ est le \mathbb{Q} -espace vectoriel

engendré par les éléments $\log |\alpha|$, $\alpha \in F^*$, le théorème 2.1 est alors une égalité dans

$$\widehat{\text{CH}}^1(S)_{\mathbb{Q}} = \mathbb{R}/\mathcal{S}$$

et le théorème 3.1 est une égalité dans

$$\widehat{\text{CH}}^1(S)_K = \mathbb{C}/\mathcal{C},$$

où \mathcal{C} est le K -espace vectoriel engendré dans \mathbb{C} par les éléments $\log |\alpha|$, $\alpha \in F^*$.

4. PÉRIODES DES VARIÉTÉS À MULTIPLICATION COMPLEXE

4.1. Soient $n \geq 1$ un entier, et $F \subset \mathbb{C}$ un corps de nombres plongé dans les complexes. On suppose que F contient le corps cyclotomique $K = \mathbb{Q}(\mu_n)$ des racines n -ièmes de l'unité. Posons $G = \text{Spec}(F[T]/(T^n - 1))$, choisissons une racine de l'unité n -ième primitive $\gamma \in \mathbb{C}^*$ et notons $g \in G(K)$ l'élément correspondant. Choisissons une métrique de Kähler G -invariante sur X ; soit $\bar{\Omega}$ le fibré hermitien des différentielles sur X . Maillot et Roessler ont eu l'idée d'appliquer le théorème de Lefschetz arithmétique (Théorème 3.1, Remarque ii)) au fibré hermitien virtuel $\lambda_{-1}(\bar{\Omega})$. À cause de la proposition suivante, son énoncé se simplifie considérablement :

PROPOSITION 4.1. — *On a*

- i) $T_g(\lambda_{-1}(\bar{\Omega}_{\mathbb{C}})) = 0$
- ii) $\text{ch}_g(\lambda_{-1}(\bar{\Omega}_{\mathbb{C}})) \text{Td}_g(TX(\mathbb{C})) = c^{\text{top}}(TY(\mathbb{C}))$
- iii) $\int_Y \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\Omega})) \widehat{\text{Td}}_g(X) = 0$.

La classe $c^{\text{top}}(TY(\mathbb{C}))$ est celle dont la restriction à toute composante de $Y(\mathbb{C})$ de dimension d est égale à $c_d(TY(\mathbb{C}))$. On déduit de la proposition 4.1 et du théorème 3.1 l'égalité

$$(3) \quad \hat{\chi}_g(\lambda_{-1}(\bar{\Omega})) = -\frac{1}{2} \int_{Y(\mathbb{C})} R_g(TX(\mathbb{C})) c^{\text{top}}(TY(\mathbb{C}))$$

dans \mathbb{C}/\mathcal{C} . Seule la composante $R_g^{(0)}$ de degré zéro de R_g intervient dans l'intégrale. Cette formule montre que $\hat{\chi}_g(\lambda_{-1}(\bar{\Omega}))$ ne dépend pas du choix de la métrique h_X .

4.2. Le groupe de Galois $G_K = \text{Gal}(K/\mathbb{Q})$ est canoniquement isomorphe à $(\mathbb{Z}/n)^*$ par l'application qui à $\sigma \in G_K$ associe l'élément $u \in (\mathbb{Z}/n)^*$ tel que

$$\sigma(\gamma) = \gamma^u.$$

Si à $u \in (\mathbb{Z}/n)^*$ on associe un des côtés de l'identité (3) où l'on a remplacé g par g^u , on obtient une fonction de $(\mathbb{Z}/n)^*$ à valeurs dans \mathbb{C}/\mathcal{C} . Nous allons calculer les transformées de Fourier sur $(\mathbb{Z}/n)^*$ de ces deux fonctions.

Soit

$$\chi : (\mathbb{Z}/n)^* \rightarrow S^1$$

un caractère de G_K , que nous supposons primitif et impair. On prolonge χ par zéro à l'ensemble \mathbb{Z}/n et l'on note aussi $\chi(m)$ l'image par χ de la classe modulo n de l'entier m . Soit

$$\tau(\chi) = \sum_{\sigma \in G_K} \sigma(\gamma) \chi(\sigma)$$

la somme de Gauss associée à χ et γ .

Pour tout $u \in \mathbb{Z}/n$, on définit comme suit des périodes

$$P_u(H^k(X)) \in \mathbb{C}^*/F^* .$$

Si $k \geq 0$ est un entier considérons la cohomologie de De Rham $H_{dR}^k(X)$. C'est un F -espace vectoriel, muni d'une action de G et donc d'une graduation

$$H_{dR}^k(X) = \bigoplus_{u \in \mathbb{Z}/n} H_{dR}^k(X)_u .$$

Le complexifié

$$H_{dR}^k(X)_u \otimes_F \mathbb{C} = H_B^k(X(\mathbb{C}), \mathbb{C})_u$$

admet une autre F -structure, à savoir la cohomologie de Betti (= cohomologie singulière) à coefficients dans F , notée $H_B^k(X(\mathbb{C}), F)_u$. Si v_{dR} est un générateur de la droite $\det_F H_{dR}^k(X)_u$ sur F et si v_B est un générateur de la droite $\det_F H_B^k(X(\mathbb{C}), F)_u$, l'élément $P_u(H^k(X)) \in \mathbb{C}^*/F^*$ est la classe du nombre complexe λ tel que

$$(4) \quad v_{dR} = \lambda v_B$$

dans $\det_{\mathbb{C}} H^k(X(\mathbb{C}), \mathbb{C})_u$. C'est une période au sens de [14], auquel on se référera pour la comparaison des cohomologies.

PROPOSITION 4.2. — *Dans \mathbb{C}/\mathcal{C} on a l'égalité*

$$\sum_{\sigma \in G_K} \chi(\sigma) \sum_{p \geq 0, q \geq 0} (-1)^{p+q} \widehat{\deg}_{\sigma(g)} H^q(X, \Lambda^p \Omega) = -\tau(\chi) \sum_{k \geq 0} (-1)^k \sum_{u \in (\mathbb{Z}/n)^*} \overline{\chi(u)} \log |P_u(H^k(X))| .$$

4.3. Soit $L(\chi, s)$, $s \in \mathbb{C}$, la fonction L de Dirichlet de χ . Si $\text{Re}(s) > 1$, elle est donnée par la série absolument convergente

$$L(\chi, s) = \sum_{m \geq 0} \frac{\chi(m)}{m^s} .$$

Posons $H^{pq}(X(\mathbb{C})) = H^q(X(\mathbb{C}), \Lambda^p \Omega_{\mathbb{C}})$.

PROPOSITION 4.3. — *Dans \mathbb{C}/\mathcal{C} on a l'égalité*

$$\begin{aligned} & \frac{1}{2} \sum_{\sigma \in G_K} \chi(\sigma) \int_{Y(\mathbb{C})} R_{\sigma(g)}(TX(\mathbb{C})) e^{\text{top}}(TY(\mathbb{C})) \\ &= \tau(\chi) \sum_{k \geq 0} (-1)^k \frac{L'(\bar{\chi}, 0)}{L(\bar{\chi}, 0)} \sum_{u \in (\mathbb{Z}/n)^*} \sum_{p+q=k} p \dim_{\mathbb{C}} H^{pq}(X(\mathbb{C}))_u \bar{\chi}(u) . \end{aligned}$$

On déduit de l'égalité (3) pour $\sigma(g)$, $\sigma \in G_K$, et des propositions 4.1, 4.2 et 4.3 (où l'on remplace χ par son conjugué), le théorème suivant :

THÉORÈME 4.4 ([29] Th. 1). — Dans \mathbb{C}/\mathcal{C} on a l'égalité

$$\begin{aligned} & \sum_{k \geq 0} (-1)^k \sum_{u \in (\mathbb{Z}/n)^*} \chi(u) \log |P_u(H^k(X))| \\ &= \sum_{k \geq 0} (-1)^k \frac{L'(\chi, 0)}{L(\chi, 0)} \sum_{u \in (\mathbb{Z}/n)^*} \sum_{p+q=k} p \dim_{\mathbb{C}}(H^{pq}(X(\mathbb{C}))_u) \chi(u). \end{aligned}$$

5. PREUVE DES PROPOSITIONS 4.1, 4.2 ET 4.3

5.1. Pour montrer la proposition 4.1, ii) et iii), on note d'abord que, si $\bar{E}' \oplus \bar{E}''$ est la somme directe orthogonale de deux fibrés hermitiens G -invariants sur X , on a

$$(5) \quad \widehat{\text{ch}}_g(\lambda_{-1}(\bar{E}' \oplus \bar{E}'')) = \widehat{\text{ch}}_g(\lambda_{-1}(\bar{E}')) \widehat{\text{ch}}_g(\lambda_{-1}(\bar{E}'')).$$

Or le fibré $\bar{\Omega}$ est la somme directe orthogonale de \overline{TY} et du fibré conormal \bar{N}^\vee . Par conséquent

$$(6) \quad \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\Omega})) = \widehat{\text{ch}}_g(\lambda_{-1}(\bar{N}^\vee)) \widehat{\text{ch}}_g(\lambda_{-1}(\overline{TY})).$$

Pour tout fibré hermitien \bar{E} sur Y on a la formule

$$(7) \quad \widehat{\text{ch}}(\lambda_{-1}(\bar{E})) = \frac{\hat{c}^{\text{top}}(\bar{E}^\vee)}{\widehat{\text{Td}}(\bar{E}^\vee)}.$$

En effet la formule (5) et le principe de scindage (dans sa version arithmétique) ramènent la preuve de cette identité au cas où \bar{E} est de rang un, qui suit de la formule évidente

$$1 - e^x = \frac{-x}{\text{Td}(-x)}.$$

Il suit du paragraphe 3.3, de (6) et de (7) que

$$(8) \quad \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\Omega})) \widehat{\text{Td}}_g(X) = \hat{c}^{\text{top}}(\overline{TY}).$$

Sur une composante Y_0 de Y de dimension d , la classe $\hat{c}^{\text{top}}(\overline{TY})$ est dans $\widehat{\text{CH}}^d(Y_0)_{\mathbb{Q}}$. L'énoncé iii) de la proposition 4.1 suit donc du fait que l'intégration \int_{Y_0} est nulle en degré autres que $d+1$ (cf. 2.1 et 3.5, Remarque ii)). La proposition 4.1, ii) suit de (8).

5.2. La proposition 4.1 i) est due à Ray et Singer si $n = 1$ [31]. Posons

$$A^{pq} = A^{0q}(X(\mathbb{C}), \Lambda^p \Omega_{\mathbb{C}}).$$

La décomposition de Hodge

$$A^{pq} = \mathcal{H}^{pq} \oplus \partial(A^{p-1,q}) \oplus \partial^*(A^{p+1,q}),$$

où \mathcal{H}^{pq} désigne les formes harmoniques, implique que le cobord $\partial : A^{pq} \rightarrow A^{p+1,q}$ induit des isomorphismes

$$\partial^*(A^{p+1,q}) \rightarrow \partial(A^{p,q}),$$

qui commutent à l'action de G et à celle du laplacien. Les valeurs propres de celui-ci sur ces deux espaces interviennent avec des signes opposés dans la définition de $T_g(\lambda_{-1}(\bar{\Omega}_{\mathbb{C}}))$, d'où l'énoncé.

5.3. Pour démontrer la proposition 4.2, on utilise d'abord l'identité

$$(9) \quad \sum_{\sigma \in G_K} \sigma(\gamma^u) \chi(\sigma) = \tau(\chi) \overline{\chi(u)}$$

(valable pour tout $u \in \mathbb{Z}/n$), pour en déduire

$$(10) \quad \begin{aligned} & \sum_{\sigma \in G_K} \chi(\sigma) \sum_{p,q} (-1)^{p+q} \widehat{\text{deg}}_{\sigma(g)} H^q(X, \Lambda^p \Omega) \\ &= \sum_{k \geq 0} (-1)^k \sum_{u \in (\mathbb{Z}/n)^*} \widehat{\text{deg}}(H_{dR}^k(X)_u) \left(\sum_{\sigma \in G_K} \sigma(\gamma)^u \chi(\sigma) \right) \\ &= \tau(\chi) \sum_{k \geq 0} (-1)^k \sum_{u \in (\mathbb{Z}/n)^*} \widehat{\text{deg}}(H_{dR}^k(X)_u) \overline{\chi(u)}. \end{aligned}$$

On compare ensuite $\widehat{\text{deg}}(H_{dR}^k(X)_u)$ et $\log |P_u(H^k(X))|$, pour un choix convenable de h_X (rappelons que, d'après (3), la somme (10) ne dépend pas de h_X). Soit $[H] \in H_B^2(X(\mathbb{C}), K(1))$ la classe d'une section hyperplane de $X(\mathbb{C})$ (où $K(1)$ est le twist à la Tate). On supposera que $[H]$ est invariante par G et que $(2\pi i)^{-1}[H]$ est la classe $[\omega_0]$ de la forme ω_0 (cf. 2.3) dans $H_B^2(X(\mathbb{C}), \mathbb{C})$. Le cup-produit par $[H]$ vérifiant le théorème de Lefschetz vache, on peut définir un opérateur $*$: $H_B^k(X(\mathbb{C}), K) \rightarrow H_B^{2d-k}(X(\mathbb{C}), K(d))$ par les formules habituelles, avec $d = \dim_F(X)$ (cf. par exemple [21]). Le cup-produit par $(2\pi i)\omega_0$ sur les formes différentielles C^∞ complexes sur $X(\mathbb{C})$ vérifie aussi le théorème de Lefschetz vache, et l'opérateur $*$ associé à ce cup-produit vérifie (aux signes près) les mêmes formules que son analogue algébrique [37]. Si $\text{tr}_B : H_B^{2d}(X(\mathbb{C}), K(d)) \rightarrow K$ est le morphisme trace en cohomologie de Betti [14], et si a, b sont les classes dans $H_B^k(X(\mathbb{C}), \mathbb{C})$ de deux formes différentielles α et β , on a donc

$$\text{tr}_B(a \cup *b) = \int_{X(\mathbb{C})} \alpha \wedge *b = \pm h_{L^2}(\alpha, \beta) (2\pi)^d.$$

Si v_{dR} et v_B sont des générateurs de $\det H_{dR}^k(X)_u$ et $\det H_B^k(X)_u$ respectivement on a donc, d'après (4),

$$\widehat{\text{deg}}(H_{dR}^k(X), h_{L^2}) = -\log \|v_{dR}\|_{L^2} = -\log \|v_B\|_B - \log |\lambda| - \frac{db_u}{2} \log(2\pi)$$

où $\|\cdot\|_B$ est la norme sur $\det H_B^k(X(\mathbb{C}), \mathbb{C})$ définie par $\text{tr}_B(a \cup *b)$ (métrique de Hodge) et $b_u = \dim_F H_{dR}^k(X)_u$. Comme $\|v_B\|_B$ est dans F^* et $|\lambda| = |P_u(H^k(X))|$, on trouve que

$$(11) \quad \widehat{\text{deg}}(H_{dR}^k(X), h_{L^2}) = -\log |P_u(H^k(X))| - \frac{db_u}{2} \log(2\pi)$$

dans $\mathbb{R}/\log|F^*|$. Le nombre b_u ne dépend pas de u , comme on le voit en faisant agir $\text{Gal}(\mathbb{C}/\mathbb{Q})$ sur les coefficients de $H_B^k(X(\mathbb{C}), \mathbb{C})$. La transformée de Fourier de la fonction constante $\frac{db_u}{2} \log(2\pi)$ est donc nulle en χ et la proposition 4.2 résulte de (10) et de (11).

5.4. Pour montrer la proposition 4.3 on remarque d’abord que, par définition (cf. 3.4),

$$(12) \quad R_g^{(0)}(TX(\mathbb{C})) = \sum_{u \in \mathbb{Z}/n} r_u (\zeta'(\gamma^u, 0) - \zeta'(\gamma^{-u}, 0)),$$

où r_u est le rang du fibré $TX(\mathbb{C})_u$.

Comme χ est impair, on déduit de (6) que, si $\text{Re}(s) > 1$,

$$(13) \quad \begin{aligned} & \frac{1}{2} \sum_{\sigma \in G_K} (\zeta(\sigma(\gamma), s) - \zeta(\sigma(\gamma), s)) \chi(\sigma) \\ &= \sum_{\sigma \in G_K} \zeta(\sigma(\gamma), s) \chi(\sigma) \\ &= \sum_{m \geq 1} \sum_{\sigma \in G_K} \frac{\sigma(\gamma)^m}{m^s} \chi(\sigma) \\ &= \tau(\chi) \sum_{m \geq 1} \frac{\bar{\chi}^{(m)}}{m^s} \\ &= \tau(\chi) L(\bar{\chi}, s). \end{aligned}$$

Les égalités (12) et (13) impliquent

$$(14) \quad \sum_{\sigma \in G_F} \chi(\sigma) R_{\sigma(g)}^{(0)}(TX(\mathbb{C})) = 2 \tau(\chi) L'(\bar{\chi}, 0) \left(\sum_{u \in \mathbb{Z}/n} \bar{\chi}(u) r_u \right).$$

LEMME — Dans \mathbb{C}/\mathcal{C} on a l’égalité

$$\begin{aligned} & \int_{Y(\mathbb{C})} \left(\frac{r_0}{2} + \sum_{\substack{u \in \mathbb{Z}/n \\ u \neq 0}} \frac{\gamma^u}{1 - \gamma^u} r_u \right) c^{\text{top}}(TY(\mathbb{C})) \\ &= \int_{Y(\mathbb{C})} \left(\sum_{p \geq 0} (-1)^p p \text{ch}_g(\Lambda^p \Omega_{\mathbb{C}}) \right) \text{Td}_g(TX(\mathbb{C})). \end{aligned}$$

PREUVE DU LEMME — Pour tout fibré G -invariant $E_{\mathbb{C}}$ sur $Y(\mathbb{C})$, considérons le polynôme

$$\phi(E_{\mathbb{C}}, t) = \sum_{p \geq 0} (-1)^p \text{ch}_g(\Lambda^p E_{\mathbb{C}}) t^p.$$

On a

$$(15) \quad \phi(E'_{\mathbb{C}} \oplus E''_{\mathbb{C}}, t) = \phi(E'_{\mathbb{C}}, t) \phi(E''_{\mathbb{C}}, t).$$

Si $E_{\mathbb{C}} = \bigoplus_j L_j$ est une somme directe de fibrés inversibles telle que g agit sur L_j par multiplication par $\gamma_j \in \mathbb{C}^*$, on a donc, en notant $\phi'(E_{\mathbb{C}}, t)$ la dérivée de $\phi(E_{\mathbb{C}}, t)$ par rapport à t ,

$$(16) \quad \sum_{p \geq 0} (-1)^p p \operatorname{ch}_g(\Lambda^p E_{\mathbb{C}}) = \phi'(E_{\mathbb{C}}, 1) = \phi(E_{\mathbb{C}}, 1) \left(\sum_j \frac{\phi'(L_j, 1)}{\phi(L_j, 1)} \right).$$

Si g agit sur un fibré inversible L par multiplication par γ on calcule, en posant $x = c_1(L)$,

$$\phi(L, t) = 1 - \gamma e^x t.$$

Par conséquent le terme de degré 0 de $\frac{\phi'(L, 1)}{\phi(L, 1)}$ est $-\frac{\gamma}{1-\gamma}$ si $\gamma \neq 1$, et $-\frac{1}{2}$ si $\gamma = 1$. On déduit donc de (16) que

$$\sum_p (-1)^p p \operatorname{ch}_g(\Lambda^p \Omega_{\mathbb{C}}) = -\phi(\Omega_{\mathbb{C}}, 1) \left(\frac{r_0}{2} + \sum_{u \neq 0} \frac{\gamma^u}{1 - \gamma^u} r_u \right).$$

D'après la proposition 4.1, ii), on sait que

$$\phi(\Omega_{\mathbb{C}}, 1) = \operatorname{Td}_g(TX(\mathbb{C}))^{-1} c^{\operatorname{top}}(TY(\mathbb{C})),$$

d'où le résultat.

FIN DE LA PREUVE DE LA PROPOSITION 4.3 — Si $z \neq 1$ on a $\zeta(z, 0) = \frac{z}{1-z}$. On déduit donc de (13) que

$$(17) \quad \sum_{\sigma \in G_K} \chi(\sigma) \frac{\sigma(\gamma)}{1 - \sigma(\gamma)} = \sum_{\sigma \in G_K} \chi(\sigma) \zeta(\sigma(\gamma), 0) = \tau(\chi) L(\bar{\chi}, 0).$$

D'après un théorème de Dirichlet et l'équation fonctionnelle, ce nombre est non nul puisque χ est primitif et non trivial. Par ailleurs, la formule de Lefschetz pour les fibrés holomorphes implique

$$\begin{aligned} & \int_{Y(\mathbb{C})} \left(\sum_{p \geq 0} (-1)^p p \operatorname{ch}_g(\Lambda^p \Omega_{\mathbb{C}}) \right) \operatorname{Td}_g(TX(\mathbb{C})) \\ &= \sum_{u \in \mathbb{Z}/n} \sum_{p, q} (-1)^{p+q} p \dim_{\mathbb{C}} H^{pq}(X(\mathbb{C}))_u \gamma^u. \end{aligned}$$

La transformée de Fourier de cette identité et de celle du lemme montre donc, compte tenu de (17), de (13) et du fait que χ est nul en dehors de $(\mathbb{Z}/n)^*$,

$$(18) \quad \begin{aligned} & \int_{Y(\mathbb{C})} \sum_{u \in (\mathbb{Z}/n)^*} \bar{\chi}(u) r_u c^{\operatorname{top}}(TY(\mathbb{C})) \\ &= \frac{1}{L(\bar{\chi}, 0)} \sum_{u, p, q} (-1)^{p+q} p \dim_{\mathbb{C}}(H^{pq}(X(\mathbb{C}))_u) \bar{\chi}(u). \end{aligned}$$

La proposition 4.3 suit de (14) et de (18).

6. GÉNÉRALISATIONS DE LA FORMULE DE CHOWLA-SELBERG

6.1. Le théorème 4.4 montre que deux sommes alternées sur $k \geq 0$ sont égales. Maillot et Roessler conjecturent que cette identité est vraie terme à terme ([29], Conjecture A). Ils le démontrent dans certains cas :

THÉORÈME 6.1 ([29], Th. 1, Th. 2, Cor. 4.2 et Cor. 4.3). — *Si $k = 1$, si $k = 2$ et $\dim(X) = 2$, ou si X est une hypersurface de dimension $k \geq 1$, l'égalité suivante est vraie dans \mathbb{C}/\mathbb{C} :*

$$\begin{aligned} & \sum_{u \in (\mathbb{Z}/n)^*} \chi(u) \log |P_u(H^k(X))| \\ &= \frac{L'(\chi, 0)}{L(\chi, 0)} \sum_{u \in (\mathbb{Z}/n)^*} \sum_{p+q=k} (-1)^k p \dim_{\mathbb{C}}(H^{pq}(X(\mathbb{C}))_u) \chi(u). \end{aligned}$$

6.2. L'égalité du théorème 6.1 étant compatible à la dualité de Poincaré, et triviale pour $k = 0$, il suffit de la montrer pour $k = 1$. On pourra alors supposer que X est une variété abélienne A sur F sur laquelle G agit par multiplication complexe (remplacer X par sa jacobienne). On sait alors que le cup-produit

$$\Lambda^k H_{dR}^1(A) \rightarrow H_{dR}^k(A)$$

est un isomorphisme. Si l'on choisit sur $A(\mathbb{C})$ la métrique de Kähler invariante par translation et de volume un, ces isomorphismes sont compatibles à la métrique L^2 . Sachant que, si \bar{M} et \bar{N} sont des F -espaces vectoriels hermitiens, et si r est le rang de M ,

$$\begin{aligned} \widehat{\deg}_g(\Lambda^k \bar{M}) &= \frac{(r-1)!}{(r-k)!(k-1)!} \widehat{\deg}_g(\bar{M}), \\ \widehat{\deg}_g(\bar{M} \oplus \bar{N}) &= \widehat{\deg}_g(\bar{M}) + \widehat{\deg}_g(\bar{N}) \end{aligned}$$

et

$$\widehat{\deg}_g(\bar{M}^\vee) = -\widehat{\deg}_{-g}(\bar{M}),$$

on peut calculer le terme gauche de l'égalité (3) quand $X = A$. On trouve ([29] Lemma 2.10 et p. 749)

$$(19) \quad \hat{\chi}_g(\lambda_{-1}(\bar{\Omega})) = -N \sum_{u \in \mathbb{Z}/n} \left(\frac{\gamma^u}{1 - \gamma^u} \right) \widehat{\deg}(H_{dR}^1(A)_u),$$

où

$$N = \prod_{u \in \mathbb{Z}/n} (1 - \gamma^u)^{b_u}.$$

Les points fixes de l'action de G sur $A(\mathbb{C})$ sont en nombre fini, égal à N par le théorème de Lefschetz pour les fibrés holomorphes. Compte tenu de (12) on a donc

$$(20) \quad \int_{Y(\mathbb{C})} R_g(TX(\mathbb{C})) c^{\text{top}}(TY(\mathbb{C})) = N \sum_{u \in \mathbb{Z}/n} r_u(\zeta'(\gamma^u, 0) - \zeta'(\gamma^{-u}, 0)).$$

Après transformation de Fourier sur $(\mathbb{Z}/n)^*$, il résulte de (3), (14), (19) et (20) que, si χ est un caractère primitif impair de $(\mathbb{Z}/n)^*$,

$$-L(\bar{\chi}, 0) \sum_{u \in (\mathbb{Z}/n)^*} \bar{\chi}(u) \widehat{\deg}(H_{dR}^1(A)_u) = 2 L'(\bar{\chi}, 0) \left(\sum_{u \in (\mathbb{Z}/n)^*} \bar{\chi}(u) r_u \right).$$

Comme $r_u = \dim_{\mathbb{C}} H^{10}(X(\mathbb{C}))_u$, le théorème 6.1 résulte alors de (11).

6.3. Supposons que $n = p$ est un nombre premier ; soit $\Gamma(s)$ la fonction Gamma. Si $u \in (\mathbb{Z}/n)^*$ et si $k \geq 0$, on note $(p(u), q(u))$ le type de Hodge de la droite $\det(H^k(X(\mathbb{C}), \mathbb{C})_u)$. On peut trouver des nombres rationnels $\varepsilon(a)$, $a \in (\mathbb{Z}/n)^*$, tels que

$$p(u) = \sum_{a \in \mathbb{Z}/p} \varepsilon(a) \left[\frac{au}{p} \right],$$

où $[\cdot]$ désigne la partie entière ([13, 35], Lemme 6.12).

COROLLAIRE ([29]) — Si $k = 1$, si $k = 2 = \dim(X)$, ou si X est une hypersurface de dimension k , pour tout $u \in (\mathbb{Z}/n)^*$ on a l'égalité dans $\mathbb{R}/(\mathbb{R} \cap \mathcal{C})$:

$$\log |P_u(H^k(X))| = \log \left| \prod_{a \in \mathbb{Z}/p} \Gamma \left(1 - \frac{a}{p} \right)^{\varepsilon(a/u)} \right|.$$

Remarques. —

- i) Pour démontrer ce corollaire, Maillot et Roessler utilisent le théorème 6.1, une formule classique ([11], III.1.2.2) exprimant la dérivée logarithmique de $L(\chi, s)$ en $s = 0$ en termes de valeurs de $\Gamma(s)$ et la relation

$$P_u(H^k(X)) \cdot P_{-u}(H^k(X)) = (2\pi i)^k$$

dans \mathbb{C}^*/F^* , qui provient de la dualité de Poincaré et du théorème de Lefschetz vache.

- ii) Une conjecture de Gross et Deligne ([19], p. 205) affirme que, pour tout entier $k \geq 0$, et tout $u \in (\mathbb{Z}/n)^*$, la période $P_u(H^k(X))$ est le produit de $\prod_{a \in \mathbb{Z}/p} \Gamma \left(1 - \frac{a}{p} \right)^{\varepsilon(a/u)}$ par un élément de $\bar{\mathbb{Q}}^*$. En dehors du cas des variétés abéliennes de type CM (i.e. $k = 1$), rien n'était connu avant ce corollaire.
- iii) Quand $k = 1$, cette conjecture de Gross et Deligne est une généralisation démontrée par Gross [19] et Anderson [1] de la formule de Chowla-Selberg [12] concernant les périodes des courbes elliptiques de type CM. Colmez a rendu plus précis ce résultat en calculant la hauteur de Faltings des variétés abéliennes de type CM [11]. Dans [24], Th. 1.3, Köhler et Roessler utilisent leur théorème de Lefschetz arithmétique sur la base $S = \text{Spec}(\mathcal{O}_F[1/n])$ pour retrouver la formule de Colmez, à l'addition près d'un élément du K -espace vectoriel engendré dans \mathbb{C} par les éléments $\log(p)$, où p est un diviseur premier de n .

iv) La preuve de [24] utilise un fibré ample hermitien inversible équivariant sur la variété A . Comme me l’ont fait remarquer Maillot et Roessler, les arguments de [29] présentés ci-dessus permettent aussi de retrouver la proposition 5.1 de [24] et d’éviter ainsi tout calcul de torsion analytique (comparer loc. cit. avec (3), (19) et (17)). On donnera en appendice un exemple où cette méthode conduit exactement à la formule de Chowla-Selberg.

7. COMPLÉMENTS

7.1. Comme $K \subset \mathbb{C}$, on peut identifier les éléments $u \in (\mathbb{Z}/n)^* = \text{Gal}(K/\mathbb{Q})$ aux plongements complexes $\sigma : K \hookrightarrow \mathbb{C}$ et réécrire l’égalité du théorème 6.1 sous la forme

$$\sum_{\sigma:K \hookrightarrow \mathbb{C}} \chi(\sigma) \log |P_\sigma(H^k(X))| = \frac{L'(\chi, 0)}{L(\chi, 0)} \sum_{\sigma} \sum_{p+q=k} (-1)^k p \dim_{\mathbb{C}}(H^{pq}(X(\mathbb{C}))_{\sigma}) \chi(\sigma)$$

dans \mathbb{C}/\mathbb{C} . Maillot et Roessler conjecturent que cette identité reste vraie quand on remplace $H^k(X)$ par n’importe quel motif défini sur F , à coefficients dans un corps K pas nécessairement abélien sur \mathbb{Q} , dont tous les plongements complexes se factorisent par F ([29], Conjecture A).

7.2. On peut étendre les théorèmes 2.1 et 3.1 à l’image directe par un morphisme $f : X \rightarrow Y$ entre deux variétés arithmétiques tel que l’application $X(\mathbb{C}) \rightarrow Y(\mathbb{C})$ induite par f soit une submersion ([18], [17], [38] ; pour la partie analytique, voir [2] et [6]). Maillot et Roessler en déduisent dans [30] des formules pour le caractère de Chern arithmétique sur Y de la cohomologie de De Rham relative. Celles-ci font intervenir la dérivée logarithmique de $L(\chi, s)$ en un entier s tel que $-\dim(Y/S) \leq s \leq 0$. Dans [22] Köhler étudie aussi cette situation et démontre un analogue arithmétique du principe de proportionalité d’Hirzebruch.

7.3. D’autres travaux portent sur la famille universelle de variétés abéliennes sur une variété de Shimura Y . Si Y n’est pas complète, il faut étendre la théorie d’intersection arithmétique en imposant à la forme ω de la formule [19] d’avoir des singularités logarithmiques à l’infini [8] [28] [10]. Dans ce contexte, Bost [9] et Kühn [28] ont montré que l’auto-intersection du dualisant relatif de la courbe elliptique universelle sur $X_0(N)$ est le produit de $2\zeta'(-1) + \zeta(-1)$ par $[SL_2(\mathbb{Z}) : \Gamma_0(N)]/2$. On ne sait pas pour l’instant interpréter leur formule par un théorème de Riemann-Roch arithmétique.

7.4. Des formules telles que celles évoquées en 7.2 et 7.3 résultent aussi des travaux de Kudla et de ses collaborateurs sur la généralisation en rang supérieur de la formule de Gross-Zagier [20] et la valeur au centre de symétrie des dérivées de fonctions L automorphes (voir [26], [27] et l’exposé [25] de ce séminaire). Gross m’avait signalé il y a longtemps qu’un multiple de $\zeta'(2)$ (i.e. $\zeta'(-1)$ par l’équation fonctionnelle) est présent dans ses calculs avec Zagier [20] ; on peut espérer qu’une variante du théorème 2.1 permettra d’aborder ces calculs différemment.

APPENDICE : UN EXEMPLE

Considérons la courbe elliptique E sur \mathbb{Q} complétée de la courbe affine plane d'équation

$$y^2 = x^3 + 6.$$

Nous allons calculer la hauteur de Faltings géométrique de E . Soit $X \subset \mathbb{P}_{\mathbb{Z}}^2$ le schéma projectif d'équation homogène

$$y^2 z = x^3 + 6 z^3.$$

C'est un modèle de E sur $S = \text{Spec}(\mathbb{Z})$. Le critère jacobien montre qu'il est lisse sur S en dehors des points fermés A et B de coordonnées $(0, 0, 1)$ dans la fibre de X sur \mathbb{F}_2 et \mathbb{F}_3 respectivement. On vérifie que X est un schéma régulier, et que x et y fournissent des paramètres locaux aux points A et B . L'ouvert $U = X - \{A, B\}$ est le modèle de Néron de E sur \mathbb{Z} ([33], Cor. 9.1). Notons $i : S \rightarrow U$ sa section nulle et $i^* \Omega_U$ la restriction à l'origine du fibré des différentielles de U sur S . Le complexifié de $i^* \Omega_U$ est canoniquement isomorphe à l'espace $H^0(E(\mathbb{C}), \Omega_{\mathbb{C}})$ des différentielles holomorphes sur $E(\mathbb{C})$. On le munit du produit scalaire hermitien tel que la norme $\|\alpha\|$ d'une forme α soit donnée par la formule

$$(A1) \quad \|\alpha\|^2 = \frac{1}{2\pi} \int_{E(\mathbb{C})} |\alpha \wedge \bar{\alpha}|.$$

Par définition ([17], [14] 1.2), la hauteur de Faltings de E sur \mathbb{Q} est le degré du \mathbb{Z} -module inversible hermitien $\overline{i^*(\Omega_U)}$:

$$(A2) \quad h(E/\mathbb{Q}) = \widehat{\text{deg}}(\overline{i^*(\Omega_U)}).$$

Si Δ est le discriminant minimal de E sur \mathbb{Q} , la hauteur géométrique de E ([14], loc. cit.) est donnée par la formule

$$(A3) \quad h_{\text{geom}}(E) = h(E/\mathbb{Q}) - \frac{1}{12} \log(|\Delta|)$$

(on utilise ici [34] et le fait que E a potentiellement bonne réduction).

Soit ω le fibré dualisant relatif de X sur S , c'est-à-dire l'unique fibré inversible prolongeant Ω_U . Si $f : X \rightarrow S$ est le morphisme de définition de X on a

$$\omega = f^* i^* \Omega_U,$$

puisque c'est vrai sur U , et il suit de (A1), de (A2) et de la formule de projection que

$$(A4) \quad h(E/\mathbb{Q}) = \widehat{\text{deg}}(H^0(X, \omega), h_{L^2})$$

pour tout choix d'une métrique h_X sur $TX(\mathbb{C})$.

Le schéma en groupes $G = \text{Spec}(\mathbb{Z}[T]/(T^6 - 1))$ agit sur X . On suppose que h_X est F_{∞} et G -invariante, on choisit la racine 6-ième primitive $\gamma = \exp(2\pi i/6)$ dans \mathbb{C}^* et on note $g \in G(\mathbb{C})$ l'élément correspondant. Il agit sur le point de coordonnées homogènes (x, y, z) par la formule

$$g(x, y, z) = (\gamma^2 x, -y, z).$$

Le schéma Y des points fixes de G sur X a trois composantes : A, B et l'adhérence Y_0 dans X du point $(0, 1, 0)$ de E . On va appliquer le théorème 3.1 au fibré hermitien virtuel $\lambda_{-1}(\bar{\omega})$. Puisque la torsion analytique équivariante est nulle (Proposition 4.1, i)), le terme gauche de l'égalité du théorème 3.1 est

$$(A5) \quad \hat{\chi}_g(\lambda_{-1}(\bar{\omega})) - \frac{1}{2} T_g(\lambda_{-1}(\bar{\omega}_{\mathbb{C}})) = \hat{\chi}_g(\lambda_{-1}(\bar{\omega})) = \hat{\chi}_g(\bar{\mathcal{O}}_X) - \hat{\chi}_g(\bar{\omega}_X).$$

La dualité de Grothendieck montre que $H^1(X, \omega_X) = H^0(X, \mathcal{O}_X) = \mathbb{Z}$ et que $H^1(X, \mathcal{O}_X) = H^0(X, \omega_X)^\vee$. Si l'on choisit h_X pour que $E(\mathbb{C})$ ait volume 1, ces isomorphismes respectent les métriques. De plus, comme $\frac{dx}{y}$ est une différentielle invariante sur $E(\mathbb{C})$, l'action de g sur $H^0(X, \omega)_{\mathbb{C}}$ est la multiplication par γ^{-1} . Par conséquent

$$(A6) \quad \hat{\chi}_g(\bar{\mathcal{O}}_X) - \hat{\chi}_g(\bar{\omega}_X) = (\gamma - \gamma^{-1}) \widehat{\text{deg}}(H^0(X, \omega), h_{L^2}).$$

Calculons maintenant le côté droit de l'identité du théorème 3.1. Puisque $Y_0 \rightarrow S$ est l'identité on a

$$(A7) \quad \int_{Y_0} \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\omega})) \widehat{\text{Td}}_g(X) = 0.$$

Par ailleurs, le fibré tangent à $A = \text{Spec}(\mathbb{F}_2)$ est trivial, et, puisque x et y sont des paramètres de son anneau local dans X , la restriction à A du fibré conormal N^\vee est un espace vectoriel de dimension 2 sur \mathbb{F}_2 , où g agit avec les valeurs propres γ^2 et γ^3 . On en déduit

$$(A8) \quad \int_A \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\omega})) \widehat{\text{Td}}_g(X) = \frac{1 - \gamma^{-1}}{(1 - \gamma^2)(1 - \gamma^3)} \log(2)$$

et, de même,

$$(A9) \quad \int_B \widehat{\text{ch}}_g(\lambda_{-1}(\bar{\omega})) \widehat{\text{Td}}_g(X) = \frac{1 - \gamma^{-1}}{(1 - \gamma^2)(1 - \gamma^3)} \log(3).$$

Soit χ le caractère quadratique de $\mathbb{Q}(\mu_6)$. D'après les propositions 4.1 ii) et 4.3, et la formule de [11], III.1.2.2, on a

$$(A10) \quad -\frac{1}{2} \int_{Y(\mathbb{C})} \text{ch}_g(\lambda_{-1}(\omega_{\mathbb{C}})) \text{Td}_g(TX(\mathbb{C})) R_g(TX(\mathbb{C})) \\ = -i \left(\frac{3\sqrt{3}}{2} \log(\Gamma(1/3)/\Gamma(2/3)) - \frac{\sqrt{3}}{2} \log(3) - \frac{\sqrt{3}}{3} \log(2) \right).$$

Sachant que $y^2 = x^3 + 6$ est une équation de Weierstrass minimale globale de E ([32] p. 172), on calcule

$$(A11) \quad \Delta = -2^6 3^5.$$

Il résulte du théorème 3.1 et de (A4) ... (A11) que

$$h_{\text{geom}}(E) = -\frac{3}{2} \log(\Gamma(1/3)/\Gamma(2/3)) + \frac{1}{4} \log(3),$$

conformément à la formule de Chowla-Selberg ([15], 1.5).

RÉFÉRENCES

- [1] G. ANDERSON – Logarithmic derivatives of Dirichlet L -functions and the periods of abelian varieties. *Compositio Math.* **45**, No. 3, 315-332 (1982).
- [2] J.-M. BISMUT – Equivariant immersions and Quillen metrics. *J. Differ. Geom.* **41**, No. 1, 53-157 (1995).
- [3] J.-M. BISMUT – Holomorphic families of immersions and higher analytic torsion forms. *Astérisque* **244**, Paris: Société Mathématique de France, vii, 275 p. (1997).
- [4] J.-M. BISMUT – Local index theory and higher analytic torsion. *Doc. Math.*, J. DMV Extra Vol. ICM Berlin 1998, Vol. I, 143-162 (1998).
- [5] J.-M. BISMUT, G. LEBEAU – Complex immersions and Quillen metrics. *Pub. Math. IHÉS* **74**, 1-297 (1991).
- [6] J.-M. BISMUT, X. MA – Holomorphic immersions and equivariant torsion forms. *J. reine angew. Math.* **575**, 189-235 (2004).
- [7] J.-B. BOST – Théorie de l’intersection et théorème de Riemann-Roch arithmétiques. *Sém. Bourbaki*, Vol. 1990/91, Exp. 731, *Astérisque* **201-203**, 43-88 (1991).
- [8] J.-B. BOST – Potential theory and Lefschetz theorems for arithmetic surfaces. *Annales Scientifiques de l’École Normale Supérieure*, Sér. 4, **32** no. 2, 241-312 (1999).
- [9] J.-B. BOST – Intersection theory on arithmetic surfaces and L_1^2 metrics, lettre du 6/03/1998, non publiée.
- [10] J.I. BURGOS, J. KRAMER, U. KÜHN – Cohomological Arithmetic Chow Rings. arXiv. math.AG/0404122.
- [11] P. COLMEZ – Périodes des variétés abéliennes à multiplication complexe. *Ann. Math.* (2) **138**, No. 3, 625-683 (1993).
- [12] S. CHOWLA, A. SELBERG – On Epstein’s zeta-function. *J. reine angew. Math.* **227**, 86-110 (1967).
- [13] P. DELIGNE – Applications de la formule des traces aux sommes trigonométriques. Sémin. géom. algèbr. Bois-Marie, SGA 4 1/2, *Lect. Notes Math.* **569**, 168-232 (1977).
- [14] P. DELIGNE – Hodge cycles on abelian varieties. (Notes by J.S. Milne) (English). Hodge cycles, motives, and Shimura varieties, *Lect. Notes Math.* **900**, 9-100 (1982).
- [15] P. DELIGNE – Preuve des conjectures de Tate et de Shafarevitch [D’après G. Faltings]. *Sém. Bourbaki*, Vol. 1983/84, Exp. 616, *Astérisque* **121/122**, 25-41 (1985).
- [16] M. DEMAZURE – Structures algébriques. Cohomologie des groupes. Dans *Séminaire de géométrie algébrique 3, 1962/64*, dirigé par M. Demazure et A. Grothendieck. Schémas en groupes. I : Propriétés générales des schémas en groupes. Exposé 1. M. Demazure et A. Grothendieck eds., *Lect. Notes Math.* **151**, Berlin-Heidelberg-New York : Springer- Verlag. XV, 1-42, (1970).

- [17] G. FALTINGS – Lectures on the arithmetic Riemann-Roch theorem: Princeton University Press (1992).
- [18] H. GILLET, C. SOULÉ – An arithmetic Riemann-Roch theorem. *Invent. Math.* **110**, No. 3, 473-543 (1992).
- [19] B.H. GROSS – On the periods of Abelian integrals and a formula of Chowla and Selberg. With an appendix by David E. Rohrlich. *Invent. Math.* **45**, 193-211 (1978).
- [20] B.H. GROSS, D.B. ZAGIER – Heegner points and derivatives of L -series. *Invent. Math.* **84**, 225-320 (1986).
- [21] S.L. KLEIMAN – Algebraic cycles and the Weil conjectures. Dix Exposés Cohomologie Schémas, *Advanced Studies Pure Math.* **3**, 359-386 (1968).
- [22] K. KÖHLER – A Hirzebruch Proportional Principle in Arakelov Geometry. Dans *Number Fields and Function Fields - Two Parallel Worlds*, G. Van der Geer, B. Moonen, R. Schoof eds., *Progress in Mathematics* **239**, Boston-Basel-Berlin: Birkhäuser, 237-268 (2005).
- [23] K. KÖHLER, D. ROESSLER – A fixed point formula of Lefschetz type in Arakelov geometry. I: Statement and proof. *Invent. Math.* **145**, No. 2, 333-396 (2001).
- [24] K. KÖHLER, D. ROESSLER – A fixed point formula of Lefschetz type in Arakelov geometry. IV: The modular height of C. M. Abelian varieties. *J. reine angew. Math.* **556**, 127-148.
- [25] S. KUDLA – Derivatives of Eisenstein series and generating functions for arithmetic cycles. *Sém. Bourbaki*. Vol. 1999/2000, Exp. 876, *Astérisque* **276**, 341-368 (2002).
- [26] S. KUDLA, M. RAPOPORT – Arithmetic Hirzebruch Zagier cycles. *J. reine angew. Math.* **515**, 155-244 (1999).
- [27] S. KUDLA, M. RAPOPORT, T. YANG – Modular forms and special cycles on Shimura curves, *Annals of Maths. Studies*, Princeton University Press, (2005).
- [28] U. KÜHN – Generalized arithmetic intersection numbers. *J. reine angew. Math.* **534**, 209-236 (2001).
- [29] V. MAILLOT, D. ROESSLER – Conjectures sur les dérivées logarithmiques des fonctions L d'Artin aux entiers négatifs. *Math. Res. Lett.* **9**, No. 5-6, 715-724 (2002).
- [30] V. MAILLOT, D. ROESSLER – On the periods of motives with complex multiplication and a conjecture of Gross-Deligne. *Ann. Math.* **160**, 727-754 (2004).
- [31] D.B. RAY, I.M. SINGER – Analytic torsion for complex manifolds. *Ann. Math.* (2) **98**, 154-177 (1973).
- [32] J. SILVERMAN – The arithmetic of elliptic curves. Corrected reprint of the 1986 original. *Graduate Texts in Math.*, **106**, New York-Heidelberg-Berlin: Springer-Verlag.
- [33] J. SILVERMAN – Advanced topics in the arithmetic of elliptic curves. *Graduate Texts in Math.*, **151**, New York-Heidelberg-Berlin: Springer-Verlag, (1994).
- [34] J. SILVERMAN – Heights and elliptic curves. Dans *Arithmetic geometry (Storr, Conn. 1984)*. New York-Heidelberg-Berlin: Springer-Verlag, 253-265 (1986).

- [35] C. SOULÉ – Hermitian vector bundles on arithmetic varieties. Kollár, János (ed.) et al., Algebraic geometry. Proceedings of the Summer Research Institute, Santa Cruz, CA, USA, July 9-29, 1995. Providence, RI: American Mathematical Society. *Proc. Symp. Pure Math.* **62** (pt.1), 383-419 (1997).
- [36] C. SOULÉ, D. ABRAMOVICH, J.-F. BURNOL, J. KRAMER – Lectures on Arakelov geometry. *Cambridge Studies in Advanced Mathematics* **33**, Cambridge: Cambridge University Press. 177 p. (1992).
- [37] R.O. WELLS – Differential analysis on complex manifolds. 2nd ed. *Graduate Texts in Math.* **65**. New York-Heidelberg-Berlin: Springer-Verlag., 260 p. (1980).
- [38] Y. ZHA – A General Arithmetic Riemann-Roch Theorem. Thèse, Chicago University (1997), non publiée.

Christophe SOULÉ

I.H.É.S.

35 route de Chartres

F-91440 BURES-sur-YVETTE

E-mail : soule@ihes.fr

**LA CONJECTURE DE MODULARITÉ DE SERRE :
LE CAS DE CONDUCTEUR 1
[d'après C. Khare]**

par **Jean-Pierre WINTENBERGER**

INTRODUCTION

Soit $\overline{\mathbb{Q}}$ la clôture algébrique de \mathbb{Q} dans \mathbb{C} . Notons $G_{\mathbb{Q}}$ le groupe de Galois de $\overline{\mathbb{Q}}/\mathbb{Q}$. Soit p un nombre premier et soit $\overline{\mathbb{F}_p}$ une clôture algébrique du corps à p élément. Soit $\overline{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_p})$ une représentation continue (donc à image finie), irréductible et impaire ($\det(\overline{\rho}(c)) = -1$ où c est la conjugaison complexe). Nous appelons une telle représentation galoisienne une représentation de type S .

Soit $\overline{\mathbb{Q}_p}$ une clôture algébrique du corps des nombres p -adiques \mathbb{Q}_p . Eichler, Shimura, Deligne, Deligne et Serre ont associé aux formes modulaires (propres) pour les sous-groupes de congruence de $\mathrm{SL}_2(\mathbb{Z})$ des représentations galoisiennes $G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$, dont les réductions modulo p , lorsqu'elles sont irréductibles, sont de type S ([18], [20] ; pour le poids 2, on pourra voir l'appendice de Conrad dans [44]). La conjecture de Serre, qui apparaît pour la première fois en 1972 (p. 9 de [56]), dit que toute représentation de $G_{\mathbb{Q}}$ de type S est *modulaire*, *i.e.* provient comme ceci d'une forme modulaire. Elle est énoncée dans [50] pour les représentations $\overline{\rho}$ non ramifiées en dehors de p (cas de niveau $N = 1$). Dans [52], Serre énonce pour tout niveau N ce que nous appelons la *forme forte* de la conjecture par opposition à sa *forme qualitative*. La forme forte précise le poids k , le niveau N et le caractère ϵ d'une forme primitive pour $\Gamma_1(N)$ dont provient $\overline{\rho}$.

Tate, en réponse à une lettre de Serre, prouve qu'il n'y a pas de représentation $G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_2})$ qui soit irréductible et non ramifiée hors de 2, prouvant ainsi la conjecture pour $p = 2$ et $N = 1$ (1973, [57]). La méthode de Tate, qui repose sur une minoration de discriminant, a été étendue par Serre au cas $p = 3$ et $N = 1$ (p. 710 de [51]), et, sous l'hypothèse de Riemann généralisée, par Brueggeman pour $p = 5$ ([11]).

Grâce aux travaux de Ribet, Mazur, Carayol, Gross, Coleman-Voloch, Edixhoven, Diamond..., on sait, pour $p \neq 2$, que la forme qualitative entraîne la forme forte. C'est un grand théorème sur lequel on trouvera d'excellents rapports dans [43], [26] et [44] : un cas de ce théorème a permis de déduire Fermat de la conjecture de Taniyama-Weil !

La démonstration, à la suite de Wiles, que toute courbe elliptique sur \mathbb{Q} est modulaire, donne la conjecture de Serre pour les représentations galoisiennes $\overline{\rho}$ provenant des points d'ordre p des courbes elliptiques sur \mathbb{Q} . Pour ce faire, Wiles prouve des énoncés du type suivant («*MR*» : modularité des relèvements ; § 3) : si $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$

est *géométrique*, *i.e.* vérifie des propriétés de ramification convenables, et si la réduction de ρ est du type S et modulaire, alors ρ est modulaire ([67], [64]). De plus, Wiles utilise un argument de « changement de nombre premier » : pour prouver que la représentation galoisienne $\overline{\rho}_5 : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\mathbb{F}_5)$ sur les points d'ordre 5 de la courbe elliptique E est modulaire, Wiles prouve l'existence d'une courbe elliptique E' sur \mathbb{Q} dont la représentation sur les points d'ordre 5 est isomorphe à $\overline{\rho}_5$, et la représentation $\overline{\rho}_3$ sur les points d'ordre 3 est irréductible. La représentation $\overline{\rho}_3$, dont l'image est résoluble, est modulaire d'après Langlands-Tunnell ([66]). Un théorème « MR » entraîne alors que la représentation 3-adique associée à E' est modulaire, donc aussi E' et $\overline{\rho}_5$.

À l'aide de théorèmes « MR » et d'un argument de « changement de nombre premier », Taylor prouve une version potentielle de la conjecture de Serre : une représentation de type S provient d'une forme modulaire de Hilbert après restriction à un corps de nombres totalement réel F ([61], [60] ; §4). On est alors confronté à un problème de changement de base de F à \mathbb{Q} . Dans certains cas, on peut s'assurer que F/\mathbb{Q} est résoluble, et, alors le théorème de Langlands et Tunnell permet de prouver la conjecture ([38], [27]). Les théorèmes de Taylor, le théorème de changement de base résoluble d'Arthur-Clozel ([3]), et des arguments de Taylor ([63]) permettent à Dieulefait de prouver que, étant donnée une représentation $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ possédant les propriétés d'une représentation galoisienne associée à une forme modulaire, il existe un système compatible (ρ_λ) de représentations ℓ -adiques de $G_{\mathbb{Q}}$ dont fait partie ρ ([24], §6).

La conjecture de Serre entraîne qu'une représentation $\overline{\rho}$ de type S admet des relèvements $G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$, qui de plus sont géométriques. Ramakrishna, utilisant des techniques de déformations de représentations galoisiennes, le prouve dans de nombreux cas sans supposer $\overline{\rho}$ modulaire ([41]). Khare réalise que ces techniques de déformation et en particulier des résultats de Böckle ([7]), la version potentielle de la conjecture de Serre due à Taylor et les théorèmes de type « MR » pour les corps totalement réels, permettent d'obtenir des relèvements de représentations de type S qui ont les propriétés de ramification plus précises que celles obtenues par Ramakrishna, et qui sont prédites par la forme forte de la conjecture de Serre.

L'existence de relèvements avec ces propriétés de ramification précises et l'existence de systèmes compatibles, permettent d'utiliser la technique de changement de nombre premier de Wiles. Dans [34], la conjecture de Serre pour $N = 1$ est prouvée pour $p = 5, 7$, et pour les poids $k \leq 14, k \neq 10$. Des stratégies sont données pour la ramener dans le cas général à des énoncés « MR ». Utilisant des relèvements de poids 2 avec Nebentypus, Khare parvient à déduire des théorèmes « MR » connus le cas général de niveau (conducteur) 1 ([35] ; pour un « survey » : [36]).

Enfin signalons que ce cercle d'idées a permis de grands progrès dans la preuve de la conjecture d'Artin pour les représentations impaires $G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\mathbb{C})$ ([13], [62]).

Je remercie Böckle de m'avoir communiqué son preprint [8] qui m'a beaucoup aidé pour la rédaction du § 5. Je remercie Khare pour ses remarques sur une première version du texte.

1. CONJECTURES DE MODULARITÉ DE SERRE ET DE FONTAINE-MAZUR

1.1. Représentations galoisiennes associées aux formes modulaires

Pour N entier ≥ 1 , soit $\Gamma_1(N)$ le groupe des matrices :

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}), \quad c \equiv 0 \pmod{N}, \quad a \equiv d \equiv 1 \pmod{N}.$$

Soit k un entier ≥ 1 . Soit $S_k(\Gamma_1(N))$ le \mathbb{C} -espace vectoriel des formes modulaires paraboliques de poids k et de niveau N . Une forme $f \in S_k(\Gamma_1(N))$ a un développement de Fourier à la pointe $i\infty$:

$$f(z) = \sum_{n=1}^{\infty} a_n q^n, \quad q = e^{2\pi iz}.$$

Une forme primitive f est propre pour les opérateurs de Hecke T_n , n entier > 1 , et pour les opérateurs diamant $\langle \bar{d} \rangle$, $\bar{d} \in (\mathbb{Z}/N\mathbb{Z})^*$. Elle est normalisée : $a_1 = 1$. Pour $n > 1$, a_n est la valeur propre de T_n . Notons $\epsilon : (\mathbb{Z}/N\mathbb{Z})^* \rightarrow \mathbb{C}^*$ le caractère $\langle \bar{d} \rangle(f) = \epsilon(\bar{d})f$. On a :

$$(1) : f\left(\frac{az+b}{cz+d}\right) = \epsilon(\bar{d})(cz+d)^k f(z), \quad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad c \equiv 0 \pmod{N},$$

\bar{d} étant l'image de d dans $(\mathbb{Z}/N\mathbb{Z})^*$. Les a_n et $\epsilon(\bar{d})$ engendrent un ordre de l'anneau des entiers d'une extension finie E_f de \mathbb{Q} contenue dans \mathbb{C} ; E_f est le corps des coefficients de f .

Eichler, Shimura, Deligne, et Deligne et Serre ont associé à f et à un plongement ι de E_f dans $\overline{\mathbb{Q}_p}$ une représentation galoisienne : $\rho_{f,\iota} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ qui est caractérisée à conjugaison près par les propriétés suivantes :

- $\rho_{f,\iota}$ est non ramifiée en dehors de $\{p\} \cup S_N$, S_N étant l'ensemble des premiers qui divisent N ;
- pour $\ell \notin \{p\} \cup S_N$, si Frob_{ℓ} est un élément de $G_{\mathbb{Q}}$ qui relève le Frobenius, on a : $\mathrm{tr}(\mathrm{Frob}_{\ell}) = \iota(a_{\ell})$.

Soit $\chi_p : G_{\mathbb{Q}} \rightarrow \mathbb{Z}_p^*$ le caractère cyclotomique. Le déterminant de $\rho_{f,\iota}$ est $\epsilon \chi_p^{k-1}$, où nous avons identifié de la manière naturelle $(\mathbb{Z}/N\mathbb{Z})^*$ avec le groupe de Galois de l'extension cyclotomique $\mathbb{Q}(\mu_N)/\mathbb{Q}$.

La représentation $\rho_{f,\iota}$ est impaire. En effet, il n'existe pas de forme parabolique non nulle de poids k et de caractère ϵ si l'on n'a pas $\epsilon(-1)(-1)^{k-1} = -1$, comme on le voit en considérant la matrice $-\mathrm{id}$ dans (1).

La représentation $\rho_{f,\iota}$ est irréductible ([20], [42]).

Elle est *géométrique* : elle est non ramifiée en dehors d'un ensemble fini de premiers et sa restriction au groupe de décomposition D_p est potentiellement semi-stable, au sens de la théorie de Fontaine ([1]). Ceci résulte, si $k \neq 1$, de ce que $\rho_{f,\iota}$ apparaît dans la

cohomologie étale d'une variété algébrique et des théorèmes de comparaison p -adiques ([65]). Pour $k = 1$, $\bar{\rho}$ a une image finie et est donc aussi géométrique.

1.2. La conjecture de Fontaine et Mazur ([31])

CONJECTURE 1.1. — *Soit $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ une représentation p -adique impaire, irréductible et géométrique. Alors, il existe f, ι comme ci-dessus et un entier j tels que ρ soit isomorphe à $\rho_{f,\iota}$ tordue par χ_p^j .*

La forme f , (N, k, ϵ) et j sont bien déterminés par ρ (f à conjugaison galoisienne près). En effet, les poids de Hodge-Tate de ρ sont $(j, j+k-1)$. Après torsion par χ_p^{-j} , on se ramène au cas $j = 0$. Soit pour nombre premier ℓ , r_ℓ la représentation F -semi-simple du groupe de Weil-Deligne WD_ℓ à valeurs dans $\mathrm{GL}_2(\overline{\mathbb{Q}_p})$ qui est associée à la restriction de ρ au groupe de décomposition D_ℓ . Pour $\ell \neq p$, r_ℓ a été définie par Grothendieck et Deligne ([19]). Pour $\ell = p$, elle a été définie par Fontaine à partir de l'action de D_p sur le module de Dieudonné filtré associé à la représentation potentiellement semi-stable $\rho|_{D_p}$ ([29]).

À r_ℓ est associée la partie ℓ primaire N_ℓ du conducteur. On a $N_\ell = 1$ si et seulement si soit $\ell \neq p$ et ρ est non ramifiée en ℓ , soit $\ell = p$ et $\rho|_{D_p}$ est cristalline. Le conducteur N est le produit des N_ℓ . Le caractère ϵ est défini par la formule $\det(\rho) = \epsilon \chi_p^{k-1}$.

Enfin f est déterminée par ρ par la formule $\mathrm{tr}(\mathrm{Frob}_\ell) = \iota(a_\ell)$ pour ℓ premier à N et p , puisque l'on a pris soin de choisir f primitive. On peut aussi dire que la correspondance de Langlands locale associée à r_ℓ une représentation π_ℓ de $\mathrm{GL}_2(\mathbb{Q}_\ell)$. La représentation automorphe associée à f a pour composantes locales les π_ℓ, π_∞ étant la représentation de $\mathrm{GL}_2(\mathbb{R})$ correspondant aux formes modulaires de poids k .

1.3. La conjecture de Serre

Soient f et ι comme au 1.1. Notons $\overline{\mathbb{Z}_p}$ l'anneau des entiers de $\overline{\mathbb{Q}_p}$ et fixons un isomorphisme du corps résiduel de $\overline{\mathbb{Z}_p}$ avec $\overline{\mathbb{F}_p}$.

On peut conjuguer $\rho_{f,\iota}$ de sorte que $\rho_{f,\iota}$ soit à valeurs dans $\mathrm{GL}_2(\overline{\mathbb{Z}_p})$. Par l'homomorphisme de réduction $\overline{\mathbb{Z}_p} \rightarrow \overline{\mathbb{F}_p}$, on en déduit une représentation $G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_p})$. La semi-simplifiée de cette représentation est bien déterminée à isomorphisme près. On la note $\bar{\rho}_{f,\iota}$. Elle est bien sûr impaire.

On dit qu'une représentation irréductible $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_p})$ est *modulaire* si elle est isomorphe à $\bar{\rho}_{f,\iota}$, pour f et ι comme ci-dessus. La forme qualitative de la conjecture de Serre est :

CONJECTURE 1.2. — *Soit $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}_p})$ une représentation de type S , i.e. irréductible et impaire. Alors $\bar{\rho}$ est modulaire.*

C'est un fait central dans le sujet qu'une représentation $\bar{\rho}$ peut provenir de formes f de niveaux et de poids différents. La forme forte de la conjecture définit $(k(\bar{\rho}), N(\bar{\rho}), \epsilon(\bar{\rho}))$ minimal, en un sens précisé, tel que $\bar{\rho}$ provienne de f de poids $k(\bar{\rho})$, de niveau $N(\bar{\rho})$. Si

l'on exclut certaines représentations diédrales pour $p = 2$ ou 3 , on peut de plus imposer que le caractère de f soit $\epsilon(\bar{\rho})$.

On peut formuler une version de la forme forte de la conjecture, en terme de formes modulaires modulo p de Katz, qui de plus prédit quand $\bar{\rho}$ provient d'une forme modulaire de Katz de poids 1. Si $p \neq 2$, la version qualitative de la conjecture entraîne la version forte (sous l'une ou l'autre de ses deux formes). On renvoie pour ceci au rapport d'Edixhoven ([26]).

Le niveau $N(\bar{\rho})$ est défini par la formule usuelle pour le conducteur d'une représentation, sauf que l'on ne tient compte que de la ramification en dehors de p . En particulier, on a $N(\bar{\rho}) = 1$ si et seulement si $\bar{\rho}$ est non ramifiée en dehors de p .

Le poids $k(\bar{\rho})$ ne dépend que de l'action de la ramification en p . Notons $I_p \subset G_{\mathbb{Q}}$ le sous-groupe d'inertie pour une valuation p -adique de $\overline{\mathbb{Q}}$. La définition de $k(\bar{\rho})$ repose sur les propriétés que l'on connaît de l'action de I_p dans les représentations p -adiques associées aux formes modulaires ([52],[25]). Rappelons-en quelques propriétés, pour $p \neq 2$.

Soit $\overline{\chi}_p : G_{\mathbb{Q}} \rightarrow \mathbb{F}_p^*$ la réduction modulo p du caractère cyclotomique ; notons encore $\overline{\chi}_p$ sa restriction à I_p . Soit Ψ un caractère fondamental de niveau 2 de I_p : on peut prendre pour Ψ la restriction à I_p du caractère de Kummer pour l'extension $K(p^{\frac{1}{p^2-1}})/K$, K étant l'extension quadratique non ramifiée de \mathbb{Q}_p dans $\overline{\mathbb{Q}_p}$.

On a : $2 \leq k(\bar{\rho}) \leq p^2 - 1$. Il existe j entier tel que l'on ait : $2 \leq k(\overline{\chi}_p^j \otimes \bar{\rho}) \leq p + 1$. On a $2 \leq k(\bar{\rho}) \leq p + 1$ si et seulement si l'on est dans l'un des deux cas suivants :

- il existe une droite D dans l'espace V de $\bar{\rho}$ telle que I_p opère trivialement sur V/D . L'action de I_p s'écrit alors :

$$\begin{pmatrix} \overline{\chi}_p^b & \eta \\ 0 & 1 \end{pmatrix},$$

avec $1 \leq b \leq p - 1$. Si $b \neq 1$, on a : $k(\bar{\rho}) = 1 + b$. Si $b = 1$, η est un 1-cocycle de $Z^1(I_p, \overline{\mathbb{F}_p}(\overline{\chi}_p))$. Sa classe de cohomologie $c(\eta)$ provient par la théorie de Kummer d'un élément de $\mathbb{Q}_{p,\text{nr}}^* \otimes \overline{\mathbb{F}_p}$, $\mathbb{Q}_{p,\text{nr}}$ désignant l'extension maximale non ramifiée de \mathbb{Q}_p . Soit $v : \mathbb{Q}_{p,\text{nr}}^* \otimes \overline{\mathbb{F}_p} \rightarrow \overline{\mathbb{F}_p}$ l'homomorphisme défini par la valuation. On a $k(\bar{\rho}) = 2$ si $v(c(\eta)) = 0$ (cas peu ramifié) et $k(\bar{\rho}) = p + 1$ sinon (cas très ramifié). Dans le cas peu ramifié, la restriction de $\bar{\rho}$ à D_p provient d'un schéma en groupes fini et plat sur \mathbb{Z}_p .

- on a un caractère fondamental Ψ de niveau 2 de I_p tel que l'action de I_p sur V soit semi-simple de caractères $\Psi^a \Psi^{pa}$ avec $1 \leq a \leq p - 1$. On a : $k(\bar{\rho}) = a + 1$.

Le caractère $\det(\bar{\rho})\overline{\chi}_p^{1-k(\bar{\rho})}$ se factorise à travers $\text{Gal}(\mathbb{Q}(\mu_N)/\mathbb{Q})$. Il s'identifie donc à un caractère de $(\mathbb{Z}/N\mathbb{Z})^*$, que l'on note $\epsilon(\bar{\rho})$.

La forme forte de la conjecture de Serre s'énonce :

CONJECTURE 1.3. — Soit $\bar{\rho}$ une représentation de type S . Alors, $\bar{\rho}$ provient d'une forme primitive de $S_{k(\bar{\rho})}(\Gamma_1(N(\bar{\rho})))$.

Supposons de plus que si $p = 2$ (resp. $p = 3$), $\bar{\rho}$ n'est pas induite de $\mathbb{Q}(\sqrt{-1})$ (resp. $\mathbb{Q}(\sqrt{-3})$) ; alors la conjecture dit que l'on peut supposer que le caractère de la forme soit $\epsilon(\bar{\rho})$.

2. LE THÉORÈME DE KHARE

Khare prouve la conjecture de Serre pour le niveau 1 :

THÉORÈME 2.1. — *Soit $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ une représentation irréductible impaire et non ramifiée en dehors de p . Alors, $\bar{\rho}$ provient d'une forme propre de $S_{k(\bar{\rho})}(\mathrm{SL}_2(\mathbb{Z}))$.*

Remarque 2.2. — La condition de parité entraîne que $S_k(\mathrm{SL}_2(\mathbb{Z}))$ est trivial si k n'est pas pair. Le théorème de Kronecker-Weber entraîne que, pour $\bar{\rho}$ non ramifiée hors de p , $\det(\bar{\rho})$ est une puissance de $\overline{\chi}_p$. La définition de $k(\bar{\rho})$ entraîne $\det(\bar{\rho}) = \overline{\chi}_p^{k(\bar{\rho})-1}$. On voit donc que la condition $\bar{\rho}$ impaire est équivalente à $k(\bar{\rho})$ pair.

Explicitons l'énoncé pour $k \leq 12$ (qui est prouvé dans [34] pour $k(\bar{\rho}) \neq 10$). Pour $0 < k < 12$, $S_k(\mathrm{SL}_2(\mathbb{Z}))$ est trivial et $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ est de dimension 1, engendré par $\Delta = q \prod_{n \geq 1} (1 - q^n)^{24}$. Swinnerton-Dyer a prouvé que pour $p \neq 2, 3, 5, 7, 691$, la représentation $\bar{\rho}(\Delta)$ est irréductible et qu'elle est réductible pour $p = 2, 3, 5, 7, 691$. Par torsion, on voit qu'il n'existe pas de $\bar{\rho}$ de type S non ramifiée hors de p si $p < 11$. Il n'existe pas de $\bar{\rho}$ de type S non ramifiée hors de p avec $k(\bar{\rho}) < 12$, ou si $k(\bar{\rho}) = 12$ et $p = 691$. Si $p \geq 11$ et $k(\bar{\rho}) = 12$, une telle $\bar{\rho}$ provient de Δ .

L'énoncé pour $k = 2$ entraîne qu'il n'existe pas de schéma en groupes sur \mathbb{Z} de hauteur 2 tel que la représentation galoisienne associée soit irréductible.

COROLLAIRE 1. — *On suppose $p \neq 2$. Soit $\bar{\rho}$ de type S , avec $k(\bar{\rho}) = 2$, $N(\bar{\rho}) = q$, q premier, alors $\bar{\rho}$ est modulaire.*

On a le cas particulier de la conjecture de Fontaine-Mazur :

COROLLAIRE 2. — *On suppose $p \neq 2$. Soit $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ une représentation continue irréductible géométrique et impaire de poids de Hodge-Tate $(0, k - 1)$ avec $k \geq 2$. On suppose ρ non ramifiée hors de p . On suppose de plus que la restriction $\rho|_{D_p}$ de ρ au groupe de décomposition D_p est :*

- soit cristalline de poids $2 \leq k \leq p + 1$;
- soit ordinaire.

Alors ρ provient d'une forme modulaire f .

3. MODULARITÉ DES RELÈVEMENTS DE REPRÉSENTATIONS MODULAIRES

La démonstration du théorème 2.1 nécessite des théorèmes «*MR*» pour des représentations du groupe de Galois des corps totalement réels.

Soit F un corps totalement réel de degré d sur \mathbb{Q} . Le groupe $\mathrm{GL}_2(F)$ agit sur le produit de d copies du demi-plan de Poincaré, ce qui permet de généraliser à $\mathrm{GL}_2(F)$ la théorie des formes modulaires pour $\mathrm{GL}_2(\mathbb{Q})$. On définit ainsi les formes modulaires de Hilbert. Le poids d'une telle forme est une collection de d entiers (k_1, \dots, k_d) indexés par les plongements de F dans \mathbb{R} . Nous considérons ici les formes de poids parallèle $k = k_1 = \dots = k_d$, $k \geq 2$.

Pour f forme propre, les valeurs propres des opérateurs de Hecke engendrent une extension finie E_f de \mathbb{Q} . Pour f parabolique primitive et ι un plongement du corps E_f dans $\overline{\mathbb{Q}_p}$, on sait associer une représentation $\rho_{f,\iota}$ du groupe de Galois G_F de $\overline{\mathbb{Q}}/F$ ([58]). Elle est *impaire*, i.e. pour tout $g \in G_{\mathbb{Q}}$, on a $\det(gcg^{-1}) = -1$, irréductible, non ramifiée en dehors d'un ensemble fini de premiers de F . Sous différentes hypothèses, on sait que sa restriction aux groupes de décompositions D_{\wp} pour \wp idéal premier au-dessus de p est potentiellement semi-stable de poids de Hodge-Tate $(0, k-1)$ ([6], [59], [9]).

Soit $\rho : G_F \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ une représentation irréductible impaire, géométrique de poids de Hodge-Tate $(0, k-1)$ pour les premiers de F au-dessus de p . On dit que ρ est *quasi-ordinaire* si pour tout premier \wp de F au-dessus de p , $\rho|_{D_{\wp}}$ est de la forme :

$$\begin{pmatrix} \psi_1 & * \\ 0 & \psi_2 \end{pmatrix}$$

avec $\psi_2(I_{\wp})$ finie, I_{\wp} désignant le sous-groupe d'inertie ; ρ est ordinaire si ψ_2 est non ramifié. On a la notion de forme f quasi-ordinaire (relativement à un plongement ι de son corps des coefficients dans $\overline{\mathbb{Q}_p}$), et pour une telle f , $\rho_{f,\iota}$ est quasi-ordinaire.

Si ρ est quasi-ordinaire, on dit que ρ est D_{\wp} -distinguée si les réductions $\overline{\psi}_1$ et $\overline{\psi}_2$ des caractères ψ_1 et ψ_2 sont distinctes pour tout \wp . Si ρ est D_{\wp} -distinguée et si ρ' est un relèvement quasi-ordinaire de $\overline{\rho}$, de caractères ψ'_1 et ψ'_2 , on dit que ρ' est un $\overline{\psi}_2$ -relèvement de $\overline{\rho}$ si les caractères $\overline{\psi}_2$ et $\overline{\psi}'_2$ coïncident pour tout \wp .

On a :

THÉORÈME 3.1. — *Supposons $p \neq 2$. Soit k un entier ≥ 2 . Soit F un corps de nombres totalement réel. Soit $\rho : G_F \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}_p})$ une représentation irréductible impaire, géométrique et dont les poids de Hodge-Tate sont $(0, k-1)$ pour tout premier \wp de F au-dessus de p . Soit $\overline{\rho}$ la réduction de ρ (1.3).*

On suppose que l'on a l'une des hypothèses suivantes :

- i) $F = \mathbb{Q}$, $\overline{\rho}$ est réductible, ρ est quasi-ordinaire et D_p -distinguée ;
- ii) $\overline{\rho}$ est irréductible, ρ est quasi-ordinaire et D_{\wp} -distinguée, $\overline{\rho}$ provient par réduction modulo p d'une forme f et d'un plongement ι de son corps des coefficients dans $\overline{\mathbb{Q}_p}$, f quasi-ordinaire relativement à ι , telle que $\rho_{f,\iota}$ soit un $\overline{\psi}_2$ relèvement de $\overline{\rho}$;

- iii) le commutant de $\bar{\rho}_{D_p}$ est réduit aux homothéties, $2 \leq k \leq 2p-1$, p est complètement décomposé dans F , pour tout premier \wp de F au-dessus de p , $\rho_{|D_\wp}$ est cristalline de poids de Hodge-Tate $(0, k-1)$, la restriction de $\bar{\rho}$ à $F(\mu_p)$ est irréductible, et $\bar{\rho}$ provient d'une forme de Hilbert f de poids parallèle k ;

- iv) $F = \mathbb{Q}$, $k = 2$, $\rho_{|D_\wp}$ est potentiellement cristalline, $\bar{\rho}$ provient d'une forme modulaire, et la restriction de $\bar{\rho}$ à $F(\mu_p)$ est irréductible.

Alors ρ provient d'une forme modulaire de poids parallèle k .

Les i) et ii) sont des théorèmes de Skinner-Wiles ([53], [55]). Les iii) et iv) sont dus à Kisin ([37]). Pour $F \neq \mathbb{Q}$ et $k > 2$, le iii) nécessite la généralisation de [54] au cas $k > 2$, qui n'est pas écrite (mais ce cas n'est pas nécessaire à la démonstration du théorème de Khare). Pour un cas particulier de iii), voir Taylor ([60]). Il y a une difficulté technique pour $p = 3$ ([35]). On trouvera une excellente introduction aux théorèmes «MR» dans [17].

4. VERSION POTENTIELLE DE LA CONJECTURE DE SERRE

Le théorème suivant de Taylor ([61], [60]) est fondamental :

THÉORÈME 4.1. — *On suppose $p \neq 2$. Soit $\bar{\rho}$ une représentation de type S à image non résoluble. Supposons $2 \leq k(\bar{\rho}) \leq p+1$ et $k(\bar{\rho}) \neq p$. Alors, il existe une extension finie galoisienne de \mathbb{Q} , totalement réelle, non ramifiée en p , de degré pair, telle que $\bar{\rho}_{|G_F}$ provienne d'une forme modulaire de Hilbert f sur F , le conducteur de f divisant p . On peut imposer l'une ou l'autre propriété suivante :*

- f est de poids parallèle $k(\bar{\rho})$ et de conducteur 1 ;
- f est de poids parallèle 2.

Remarque 4.2. — Soit H un sous-groupe fini de $\mathrm{PGL}_2(\overline{\mathbb{F}}_p)$. Rappelons le théorème de Dickson : H est conjugué :

- soit à un sous-groupe des matrices triangulaires supérieures ;
- soit à $\mathrm{PGL}_2(\mathbb{F}_{p^r})$ ou à $\mathrm{PSL}_2(\mathbb{F}_{p^r})$, pour un entier $r > 0$;
- soit à A_4 , S_4 (si $p \neq 2$), A_5 ou au groupe diédral d'ordre $2r$ pour un entier $r > 1$ non divisible par p .

Il en résulte facilement que, si $\bar{\rho}$ est de type S et d'image non résoluble, pour toute extension galoisienne totalement réelle de \mathbb{Q} , $\bar{\rho}(G_F)$ n'est pas résoluble.

Donnons une idée de la preuve du théorème. Taylor considère un problème de module classifiant les données (A, i, j, α) , où :

- A est une variété abélienne de dimension d ;
- i est un plongement de l'anneau des entiers O_M d'un corps de nombres M de degré d dans $\mathrm{End}(A)$;
- j est une donnée de polarisation de A ;

- α est une structure de niveau $\lambda\mathcal{L}_1$ ou $\lambda\mathcal{L}_1\mathcal{L}_2$ selon les cas. On a un premier λ de O_M au-dessus de p tel que α définisse un isomorphisme des points de λ -torsion $A[\lambda]$ avec $\bar{\rho}$. On a \mathcal{L}_1 de caractéristique résiduelle $\ell_1 \neq p$ tel que α induise un isomorphisme de $A[\mathcal{L}_1]$ avec un module induit irréductible $\text{ind}_{\mathbb{Q}}^L(\theta)$, L corps quadratique imaginaire.

Ce problème de module est représentable par un schéma X sur \mathbb{Q} , qui est lisse, géométriquement irréductible quasi-projectif. Les ensembles $X(\mathbb{R})$, $X(\mathbb{Q}_p)$ et $X(\mathbb{Q}_{\ell_i})$ sont non vides. Il résulte alors d'un théorème de Moret-Bailly que X a un point dans une extension F_0 de \mathbb{Q} qui est galoisienne finie et totalement décomposée aux places ∞ , p et ℓ_i ([40]).

On connaît par la théorie des caractères de Hecke que $\text{ind}_{\mathbb{Q}}^L(\theta)$ est modulaire. Ce module induit ind a été choisi de sorte qu'il reste irréductible après restriction au groupe de Galois de $F_0(\mu_{\ell_1})$ et qu'avec la représentation ℓ_1 -adique associée à A , il satisfasse aux hypothèses du ii) du théorème 3.1. Il en résulte que A_{F_0} provient d'une forme modulaire de Hilbert f_0 de poids 2.

Un théorème de Skinner-Wiles donne l'existence de $F_1 \supset F_0$ tel que $\bar{\rho}|_{G_{F_1}}$ provienne d'une forme f_1 non ramifiée hors de p ([54]). Taylor prouve par une analyse de la réduction de A modulo les premiers de F_1 au-dessus de p et par la théorie des congruences entre formes modulaires que l'on peut trouver $F \supset F_1$ et f vérifiant de plus les conditions en p du théorème (c'est ici qu'intervient la condition $k(\bar{\rho}) \neq p$).

5. RELÈVEMENTS AVEC CONDITIONS DE RAMIFICATION

On suppose $p \neq 2$. Soit $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\overline{\mathbb{F}_p})$ une représentation de type S à image non résoluble.

La conjecture de Serre entraîne l'existence d'un relèvement géométrique de $\bar{\rho}$. Ceci a été prouvé dans de nombreux cas par Ramakrishna sans supposer $\bar{\rho}$ modulaire ([41]). En fait, on s'attend à pouvoir être plus précis sur la ramification des relèvements. En effet, Diamond et Taylor prouvent dans beaucoup de cas le résultat suivant ([22], [23]). Supposons $\bar{\rho}$ modulaire. Soit S un ensemble fini de premiers $\neq p$, pour tout $\ell \in S$ une représentation $\tau_{\ell} : D_{\ell} \rightarrow \text{GL}_2(\overline{\mathbb{Q}_p})$ qui relève la restriction de $\bar{\rho}$ au groupe de décomposition D_{ℓ} . Alors, il existe un relèvement modulaire ρ de $\bar{\rho}$ tel que, pour tout $\ell \in S$, la restriction de ρ au groupe d'inertie I_{ℓ} soit isomorphe à la restriction de τ_{ℓ} à I_{ℓ} .

Les théorèmes suivants prouvent dans certains cas l'existence de relèvements avec conditions de ramification sans supposer que $\bar{\rho}$ est modulaire, et en incluant le cas $\ell = p$.

5.1. Relèvements minimaux

Soit ℓ un nombre premier $\neq p$. On sait que la restriction de $\bar{\rho}$ à I_{ℓ} est de l'un des types suivants :

$$\xi \otimes \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \xi_1 & 0 \\ 0 & \xi_2 \end{pmatrix}, \text{Ind}_{I_M}^{I_p}(\xi),$$

M étant une extension quadratique ramifiée de \mathbb{Q}_ℓ , soit $\ell = 2$ et l'image de I_2 dans la représentation projective $\bar{\rho}_{\text{proj}}$ associée à $\bar{\rho}$ est un sous-groupe de A_4 ([21]).

Soit ρ un relèvement de $\bar{\rho}$. Si $\ell \neq p$, on dit que ρ est *minimal* en ℓ si la restriction de ρ à I_ℓ est soit:

$$\tilde{\xi} \otimes \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \tilde{\xi}_1 & 0 \\ 0 & \tilde{\xi}_2 \end{pmatrix}, \text{Ind}_{I_M}^{I_p}(\tilde{\xi}),$$

$\tilde{*}$ désignant le représentant de Teichmüller de $*$, soit, si on est dans le cas $\ell = 2$ et $\bar{\rho}_{\text{proj}}(I_2) \subset A_4$, que $\bar{\rho}(I_2)$ et $\rho(I_2)$ soient isomorphes.

Si $2 \leq k(\bar{\rho}) < p + 1$, on dit que ρ est *minimal* en p si la restriction de ρ à I_p est cristalline de poids de Hodge-Tate $(0, k(\bar{\rho}) - 1)$. Si $k(\bar{\rho}) = p + 1$, on dit que ρ est en *minimal de type cristallin* si la restriction de ρ à I_p est cristalline de poids de Hodge-Tate $(0, p)$, et *minimal de type semi-stable* si la restriction de ρ à I_p est semi-stable de poids de Hodge Tate $(0, 1)$.

THÉORÈME 5.1 ([34]). — *Soit $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ une représentation de type S à image non résoluble. On suppose que $2 \leq k(\bar{\rho}) \leq p + 1$ et que $k(\bar{\rho}) \neq p$. Alors, $\bar{\rho}$ a un relèvement qui est minimal pour tout ℓ et si $k(\bar{\rho}) = p + 1$, on peut imposer en p soit que ρ soit minimal de type cristallin soit que ρ soit minimal de type semi-stable.*

5.2. Relèvements de poids 2

Le théorème suivant correspond au fait que toute forme modulaire de poids k , $2 \leq k \leq p + 1$, et de niveau N premier à p , est congruente modulo p à une forme de poids 2 pour $\Gamma_1(pN)$ (§ 3 de [49], prop 9.3 de [32]).

On suppose que $\bar{\rho}|_{I_p}$ est ordinaire, *i.e.* du type :

$$\begin{pmatrix} \overline{\chi}_p^{k-1} & * \\ 0 & 1 \end{pmatrix},$$

pour un entier k avec $2 \leq k \leq p$. Soit ω_p le relèvement de Teichmüller de $\overline{\chi}_p$. On dit qu'un relèvement ρ de $\bar{\rho}$ est *minimal en p de poids 2* si $\rho|_{I_p}$ est du type :

$$\begin{pmatrix} \omega_p^{k-2} \chi_p & \eta \\ 0 & 1 \end{pmatrix},$$

et, pour $k(\bar{\rho}) = 2$, si de plus $\rho|_{D_p}$ est finie, *i.e.* provient d'un schéma en groupes fini et plat sur \mathbb{Z}_p (le cocycle η provient via la théorie de Kummer d'une unité).

THÉORÈME 5.2 ([35]). — *Soit $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\overline{\mathbb{F}}_p)$ une représentation de type S à image non résoluble avec $k(\bar{\rho}) \neq p$ et $\bar{\rho}$ ordinaire. Alors $\bar{\rho}$ a un relèvement qui est minimal en tout $\ell \neq p$ et minimal de poids 2 en p .*

Remarque 5.3. — On devrait pouvoir étendre le théorème aux cas où $k(\bar{\rho}) = p$ et $\bar{\rho}|_{D_p}$ non ordinaire.

5.3. Relèvements avec Nebentypus

Le théorème suivant correspond à un théorème de Carayol qui dit que pour $p \geq 5$ ou pour $p = 3$ (resp. $p = 2$) et $\bar{\rho}$ n'est pas induite de $\mathbb{Q}(\sqrt{-3})$ (resp. $\mathbb{Q}(\sqrt{-1})$), si $\bar{\rho}$ provient par réduction d'une forme propre normalisée de $S_2(\Gamma_0(N), \chi)$, et si χ et χ' ont même réduction modulo p , $\bar{\rho}$ provient aussi d'une forme propre normalisée de $S_2(\Gamma_0(N), \chi')$ ([15]).

On suppose que l'on a un premier q tel que p divise $q - 1$, que $\bar{\rho}$ soit ramifiée en q , et que $\bar{\rho}|_{I_q}$ soit de la forme :

$$\begin{pmatrix} \gamma & * \\ 0 & 1 \end{pmatrix}$$

où γ est un caractère $\text{Gal}(\mathbb{Q}_q(\mu_q)/\mathbb{Q}_q) \rightarrow \overline{\mathbb{F}_p}^*$.

Soit $\omega_q : \text{Gal}(\mathbb{Q}_q(\mu_q)/\mathbb{Q}_q) \rightarrow \mu_{q-1}(\overline{\mathbb{Q}_q})$ le relèvement de Teichmüller du caractère cyclotomique. On choisit un isomorphisme de $\mu_{q-1}(\overline{\mathbb{Q}_q})$ sur $\mu_{q-1}(\overline{\mathbb{Q}_p})$, de sorte que l'on puisse voir ω_q comme un caractère de I_q à valeurs dans $\overline{\mathbb{Q}_p}^*$. Soit p^r l'exacte puissance de p qui divise $q - 1$; posons $\eta_q = \omega_q^{\frac{q-1}{p^r}}$, de sorte que η_q est un caractère d'ordre p^r et donc que sa réduction modulo p est triviale.

THÉORÈME 5.4 ([35]). — *Mêmes hypothèses que pour le théorème 5.1. Soit i un entier. Soit $\hat{\gamma}$ le relèvement de Teichmüller de γ . Alors $\bar{\rho}$ a un relèvement ρ qui est minimal en tout premier $\ell \neq q$ et est tel que $\rho|_{I_q}$ soit de la forme :*

$$\begin{pmatrix} \hat{\gamma}\eta_q^i & * \\ 0 & 1 \end{pmatrix}.$$

5.4. Esquisse de la preuve des théorèmes

Commençons par quelques rappels sur la théorie des déformations des représentations galoisiennes ([39]).

Soit \mathbb{F} un corps fini tel que l'image de $\bar{\rho}$ soit contenue dans $\text{GL}_2(\mathbb{F})$ et tel que $\bar{\rho}$ soit irréductible comme représentation à valeurs dans $\text{GL}_2(\mathbb{F})$. Notons W l'anneau des vecteurs de Witt à coefficients dans \mathbb{F} et soit $O = W$ pour les deux premiers théorèmes et $O = W(\mu_{p^r}(\overline{\mathbb{Q}_p}))$ pour le théorème 5.4. Notons \mathcal{C}_O la catégorie des O -algèbres R noethériennes locales séparées et complètes avec un isomorphisme i_R du corps résiduel R/\mathcal{M}_R avec \mathbb{F} . Les morphismes dans \mathcal{C}_O sont les morphismes $f : R \rightarrow R'$ de O -algèbres tels que $i_{R'} \circ f \simeq i_R$. Un relèvement de $\bar{\rho}$ est un morphisme : $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R)$ avec un isomorphisme de $\rho \bmod \mathcal{M}_R$ sur $\bar{\rho}$. Les déformations de $\bar{\rho}$ sont les classes d'isomorphisme de tels relèvements : deux relèvements ρ et ρ' sont donc isomorphes s'il existe $g \in \text{GL}_2(R)$, $g \equiv \text{id} \bmod \mathcal{M}_R$ tel que $\rho = \text{int}(g)(\rho')$.

Pour chacun des théorèmes, on a un ensemble fini S de places de \mathbb{Q} contenant p , ∞ et les premiers ramifiés dans $\bar{\rho}$ tel que les relèvements ρ recherchés soient ceux qui sont non ramifiés hors de S et qui sont tels que, pour tout $v \in S$, la restriction de ρ à I_v

vérifie une condition que l'on notera \mathcal{L}_v . On dira que les relèvements cherchés sont ceux qui vérifient \mathcal{L} .

Soit G_S le groupe de Galois de l'extension maximale de \mathbb{Q} qui est non ramifiée hors de S . Le groupe G_S vérifie la condition : la complétion p -adique de tout sous-groupe ouvert de G_S est engendrée topologiquement par un ensemble fini. Il en résulte que le foncteur des déformations de $\bar{\rho}$ qui sont non ramifiées hors de S a une enveloppe, et même est représentable puisque les endomorphismes de $\bar{\rho}$ sont réduits aux homothéties. On note $\rho_{\text{univ}} : G_S \rightarrow \text{GL}_2(R_{\text{univ}})$ la déformation universelle. Pour toute O -algèbre R de \mathcal{C}_O , on a donc une bijection de l'ensemble $\text{spec}(R_{\text{univ}})(R)$ sur l'ensemble des déformations de $\bar{\rho}$ à valeurs dans R .

De même, pour chaque $v \in S$, on a une déformation verselle $\rho_{v,\text{univ}} : D_v \rightarrow \text{GL}_2(R_{v,\text{univ}})$ de $\bar{\rho}|_{D_v}$. Il n'est pas difficile de voir que l'on a un quotient $R_{\mathcal{L}_v}$ de $R_{v,\text{univ}}$ tel que les déformations de $\bar{\rho}|_{D_v}$ qui satisfont à la condition \mathcal{L}_v sont exactement celles qui proviennent d'un point de $\text{spec}(R_{\mathcal{L}_v})$. On note $J_{\text{univ}/\mathcal{L}_v}$ le noyau de $R_{v,\text{univ}} \rightarrow R_{\mathcal{L}_v}$.

Pour $v \in S$, la restriction de ρ_{univ} au groupe de décomposition D_v induit un morphisme $r_v : R_{v,\text{univ}} \rightarrow R_{\text{univ}}$. On voit que le quotient $R_{\mathcal{L}}$ de R_{univ} par l'idéal engendré par les $r_v(J_{\text{univ}/\mathcal{L}_v})$ représente les déformations de $\bar{\rho}$ qui satisfont à la condition \mathcal{L} .

L'énoncé suivant et son corollaire rassemblent des résultats dus essentiellement à Böckle, Ramakrishna et Taylor ([7], [41], [62]).

THÉORÈME 5.5. — *On a une présentation $R_{\mathcal{L}} = O[[T_1, \dots, T_h]]/(f_1, \dots, f_r)$, avec moins de relations que de générateurs : $r \leq h$.*

COROLLAIRE 3. — *Si $R_{\mathcal{L}}$ est, en tant que O -module, de type fini, alors $R_{\mathcal{L}}$ est une O -algèbre plate d'intersection complète, en particulier, $\bar{\rho}$ a un relèvement ρ qui vérifie la condition \mathcal{L} cherchée.*

Esquisons la démonstration du théorème 5.5.

Il est pratique de remarquer tout d'abord que le déterminant des déformations qui satisfont à \mathcal{L} est fixé. En effet, la condition \mathcal{L}_v définit un morphisme $\tilde{\eta}_v : I_v \rightarrow O^*$ tel que, si ρ_v satisfait à \mathcal{L}_v , la restriction du déterminant de ρ_v à I_v a pour déterminant $\tilde{\eta}_v$. Pour $v \notin S$, on pose $\tilde{\eta}_v = 1$. Comme le groupe de Galois $\text{Gal}(\mathbb{Q}^{\text{ab}}/\mathbb{Q})$ de l'extension maximale abélienne de \mathbb{Q} est isomorphe au produit de ses sous-groupes d'inertie pour ℓ décrivant les nombres premiers, on voit que les η_ℓ définissent un caractère $\eta : \text{Gal}(\mathbb{Q}^{\text{ab}}/\mathbb{Q}) \rightarrow O^*$ tel que les déformations qui vérifient \mathcal{L} sont de déterminant η (la condition de parité $\det(\eta(c)) = -1$ est vérifiée car elle l'est pour $\bar{\rho}$ et que $p \neq 2$). On note η_v la restriction de η au groupe de décompositions D_v . La condition $\det(\rho) = \eta$ définit un sous-foncteur qui est une immersion fermée de celui des déformations qui sont non ramifiées hors de S . On note $\rho_\eta : G_{\mathbb{Q}} \rightarrow \text{GL}_2(R_\eta)$ la déformation universelle de $\bar{\rho}$ de déterminant η . De même, pour tout $v \in S$, on note $\rho_{\eta_v} : D_v \rightarrow \text{GL}_2(R_{\eta_v})$ la déformation verselle de déterminant η_v de $\bar{\rho}|_{D_v}$.

Pour R_η dans \mathcal{C}_O , on note t_η le \mathbb{F} -espace vectoriel tangent relatif et t_η^* son dual. On a donc : $t_\eta^* \simeq \mathcal{M}_{R_\eta}/(\mathcal{M}_{R_\eta})^2 + \mathcal{M}_O R_\eta$, \mathcal{M}_O désignant l'idéal maximal de O . Le choix de

relèvements des éléments d’une base de t_γ^* dans \mathcal{M}_{R_γ} permet d’écrire R_γ comme quotient d’un anneau de séries formelles S_γ sur O qui identifie les espaces tangents t_{R_γ} et t_{S_γ} . Il est facile de voir que deux telles présentations sont isomorphes. Si J est le noyau de $S_\gamma \rightarrow R_\gamma$, on pose $\text{rel}(R_\gamma) = \dim_{\mathbb{F}}(J/\mathcal{M}_{R_\gamma}J)$. Elle ne dépend pas de la présentation : c’est le nombre minimal de générateurs de J .

Les espaces cotangents s’identifient à des groupes de cohomologie galoisienne. En effet, t_η s’identifie à l’ensemble des déformations de $\bar{\rho}$ à valeurs dans $\mathbb{F}[\epsilon]/\epsilon^2$ qui sont non ramifiées hors de S et de déterminant égal à celui de $\bar{\rho}$, et $t_{\mathcal{L}}$ sont celles qui vérifient les conditions \mathcal{L}_v . On a un isomorphisme $t_\eta^* = H_0^1 := H^1(G_S, \text{ad}_0(\bar{\rho}))$, où $\text{ad}_0(\bar{\rho})$ sont les matrices de trace nulle, G_S agissant par la conjugaison.

Pour tout $v \in S$, les déformations de $\bar{\rho}|_{D_v}$ à valeurs dans $\mathbb{F}[\epsilon]/\epsilon^2$ qui sont de déterminant η_v forment un espace vectoriel qui s’identifie à $H^1(D_v, \text{ad}^0(\bar{\rho}))$. On le note $H_{v,0}^1$. Celles qui vérifient \mathcal{L}_v forment un sous-espace vectoriel de $H_{v,0}^1$ que l’on note L_v . Le sous-espace $t_{\mathcal{L}}^* \subset t_\eta^*$ s’identifie au sous-espace vectoriel des $c \in H^1(G_S, \text{ad}^0(\bar{\rho}))$ tels que, pour tout $v \in S$, l’image c_v de c dans $H^1(D_v, \text{ad}^0(\bar{\rho}))$ appartienne à L_v . On le note $H_{\mathcal{L}}^1$. On note $h_0^1, h_{\mathcal{L}}^1, h_{v,0}^1$, et l_v les dimensions des \mathbb{F} -espaces vectoriels correspondants.

Soit, pour tout $v \in S$, L_v^\perp le dual de L_v dans la dualité :

$$H^1(D_p, \text{ad}_0) \times H^1(D_p, \text{ad}_0(1)) \rightarrow \mathbb{F},$$

où (1) est la torsion par le caractère cyclotomique.

On note $H_{\mathcal{L}^\perp}^1$ le sous-espace vectoriel de $H^1(G_{\mathbb{Q}}, \text{ad}_0(1))$ formés des c dont les images c_v dans $H^1(D_p, \text{ad}_0(1))$ appartiennent à L_v^\perp .

Le théorème suivant est dû à Böckle ([7]).

THÉORÈME 5.6. — *On a :*

$$\text{rel}(R_{\mathcal{L}}) \leq h_{\mathcal{L}^\perp}^1 + \sum_{v \in S} \text{rel}(R_{\mathcal{L}_v}).$$

Donnons quelques indications sur la preuve. On peut écrire pour chaque v , la O -algèbre R_{η_v} comme un quotient d’une algèbre de séries formelles S_{η_v} à $h_{v,0}^1$ variables. Soit J_{η_v} le noyau. La théorie des déformations nous dit que l’on a un morphisme surjectif $(H^2(G_S, \text{ad}_0))^* \rightarrow J_\eta/\mathcal{M}_\eta J_\eta$ et de même pour les J_{η_v} . Il en résulte que le noyau J_η d’une présentation de R_η est engendré topologiquement par les J_{η_v} et par $\text{sh}^2 := \dim(\mathbf{III}_S^2(\text{ad}_0))$ éléments. On a donc :

$$\text{rel}(R_\eta) \leq \text{sh}^2 + \sum_{v \in S} \text{rel}(R_{\eta_v}).$$

Pour $s \in S$, notons \tilde{J}_v le noyau de $S_{\eta_v} \rightarrow R_{\eta_v} \rightarrow R_{\mathcal{L}_v}$. Alors, $R_{\mathcal{L}}$ est le quotient de S_η par l’idéal engendré par J_η et les \tilde{J}_v . Une chasse au diagramme permet d’en déduire que :

$$\text{rel}(R_{\mathcal{L}}) \leq \sum_{v \in S} \text{rel}(R_{\mathcal{L}_v}) + \text{sh}^2 + \sum_{v \in S} (h_{\eta_v}^1 - l_v) - h_{\eta}^1 + h_{\mathcal{L}}^1.$$

La suite exacte de Poitou-Tate donne alors le théorème 5.6.

L'énoncé suivant rassemble des calculs locaux faits dans ([7], [41], [62]).

THÉORÈME 5.7. — *Pour premier $\ell \in S$, la O -algèbre $R_{\mathcal{L}_\ell}$ est lisse de dimension relative $h^0(D_\ell, \text{ad}_0)$ si $\ell \neq p$, et $h^0(D_p, \text{ad}_0) + 1$ si $\ell = p$.*

Notons d_ℓ la dimension de $R_{\mathcal{L}_\ell}$ et $d_\infty = 0$. Le théorème 5.6 donne alors, puisque $R_{\mathcal{L}_v}$ est d'intersection complète :

$$\text{rel}(R_{\mathcal{L}}) \leq h_{\mathcal{L}^\perp}^1 + \sum_{v \in S} (l_v - d_v).$$

On a la formule de Wiles ([67]):

$$h_{\mathcal{L}}^1 - h_{\mathcal{L}^\perp}^1 = h^0(G_{\mathbb{Q}}, \text{ad}_0) - h^0(G_{\mathbb{Q}}, \text{ad}_0(1)) + \sum_{v \in S} (l_v - h^0(D_v, \text{ad}_0)).$$

Les h^0 globaux sont nuls. La contribution de ∞ est -1 . On en déduit :

$$h_{\mathcal{L}}^1 - \text{rel}(R_{\mathcal{L}}) \geq -1 + \sum_{\ell \in S} (d_\ell - h^0(D_\ell, \text{ad}_0)).$$

Le théorème 5.5 résulte alors du théorème 5.7.

Prouvons les théorèmes 5.1, 5.2 et 5.4.

D'après le corollaire 3, il suffit de prouver que $R_{\mathcal{L}}/p$ est finie. Comme $\bar{\rho}$ est absolument irréductible, $R_{\mathcal{L}}/p$ est engendrée par les traces des images des éléments de $G_{\mathbb{Q}}$ ([16]). Par un lemme simple d'algèbre commutative, on voit alors qu'il suffit de prouver que l'image de $G_{\mathbb{Q}}$ dans $\text{GL}_2(R_{\mathcal{L}}/p)$ est finie ([34]).

Soit F un corps totalement réel comme dans le théorème 4.1. On considère alors dans chacun des cas la déformation universelle $G_F \rightarrow \text{GL}_2(R_F)$ de la représentation irréductible $\bar{\rho}|_{G_F}$, de déterminant $\eta|_{G_F}$ et satisfaisant à certaines conditions locales. Rappelons que η est le déterminant des déformations satisfaisant \mathcal{L} . Ces conditions locales sont choisies de sorte qu'elles sont satisfaites par la restriction à G_F de tout relèvement $\rho : G_{\mathbb{Q}} \rightarrow \text{GL}(R)$ de $\bar{\rho}$, R étant une O -algèbre de \mathcal{C}_O telle que $pR = 0$. Par exemple, dans le cas des relèvements minimaux pour $k(\bar{\rho}) \neq p + 1$, ces conditions locales sont d'être non ramifiées pour tout premier de F qui n'est pas au-dessus de p , et pour \wp au-dessus de p , la condition est d'être un quotient d'une représentation cristalline de D_\wp de poids de Hodge-Tate $(0, k(\bar{\rho}) - 1)$, et donc de provenir d'un module de Fontaine-Laffaille de poids 0 et $k - 1$ ([30]).

Ainsi, on a un morphisme $R_F \rightarrow R_{\mathcal{L}}$ tel que $G_F \rightarrow \text{GL}_2(R_{\mathcal{L}}/p)$ se factorise à travers $G_F \rightarrow \text{GL}_2(R_F)$. Pour prouver que l'image de $G_{\mathbb{Q}}$ dans $\text{GL}_2(R_{\mathcal{L}}/p)$ est finie, il suffit donc de prouver que celle de G_F dans $\text{GL}_2(R_F/p)$ l'est. Pour ceci, il suffit de prouver que R_F est de type fini en tant que O -module. Pour ceci, on considère l'algèbre de

Hecke \mathbb{T} des formes modulaires de Hilbert relatives à F de poids et niveau convenables. Par exemple, dans le cas des relèvements minimaux, le poids est $k(\bar{\rho})$ et le niveau 1. La représentation $\bar{\rho}|_{G_F}$ étant modulaire, elle définit un idéal maximal de \mathbb{T} ; soit $\widehat{\mathbb{T}}$ le complété de \mathbb{T} en cet idéal. On a un morphisme $R_F \rightarrow \widehat{\mathbb{T}}$. Dans le cas où $\bar{\rho}|_{D_p}$ est irréductible, le iii) du théorème 3.1 entraîne que $R_F \simeq \widehat{\mathbb{T}}$, et donc que R_F est un O -module de type fini puisque $\widehat{\mathbb{T}}$ l'est. Dans le cas ordinaire, comme, dans le cas ii) de théorème 3.1, on utilise [55] qui n'implique pas que $R_F \simeq \widehat{\mathbb{T}}$, il faut vérifier que [55] entraîne bien que R_F est un O -module de type fini, ce qui est fait dans [35], utilisant la théorie de Hida ([33]).

6. SYSTÈMES COMPATIBLES

Soit, comme dans le paragraphe précédent, $\bar{\rho}$ une représentation de type S à image non résoluble. Soit ρ un relèvement de $\bar{\rho}$ comme dans l'un des théorèmes 5.1, 5.2, 5.4. Pour tout ℓ , on note r_ℓ la représentation du groupe de Weil-Deligne $\text{WD}_\ell \rightarrow \text{GL}_2(\overline{\mathbb{Q}}_p)$ associée à la représentation géométrique ρ (1.2). La conjecture de Fontaine-Mazur entraîne que ρ devrait être membre d'un système compatible de représentations ℓ -adiques (ρ_ℓ). Le théorème suivant, moins fort, suffit pour prouver le théorème 2.1. Il est essentiellement dû à Dieulefait ([24] ; voir aussi [68]) :

THÉORÈME 6.1. — *Il existe un corps de nombres E et, pour tout q et tout plongement $\iota_q : E \hookrightarrow \overline{\mathbb{Q}}_q$, une représentation $\rho_\iota : G_{\mathbb{Q}} \rightarrow \text{GL}_2(\overline{\mathbb{Q}}_q)$ qui est impaire, irréductible. De plus :*

- i) *il existe $\iota_p : E \hookrightarrow \overline{\mathbb{Q}}_p$ tel que ρ_{ι_p} soit isomorphe à ρ ;*
- ii) *pour ℓ et q premiers avec $\ell \neq p, q$, et pour ι_q plongement de E dans $\overline{\mathbb{Q}}_q$, les représentations du groupe de Weil-Deligne WD_ℓ associées à ρ et à ρ_{ι_q} sont isomorphes, via ι et ι_q ;*
- iii) *On suppose que dans le cas de relèvement avec Nebentypus (5.4) on a $k(\bar{\rho}) = 2$. Soit ι_q de caractéristique $q \neq 2, p$. La restriction à I_p de ρ_{ι_q} est celle définie par r_p . Si ρ n'est pas ramifiée en q , la restriction de ρ_{ι_q} à D_q est cristalline de même poids que ρ . Si $\rho|_{I_q}$ est comme dans 5.3, $\rho|_{D_q}$ est soit semi-stable de poids $(0, 1)$, soit cristalline après restriction à $\mathbb{Q}_q(\mu_q)$ et de poids de Hodge-Tate $(0, 1)$.*

La preuve utilise des idées de Taylor ([63]). La version potentielle de la conjecture de Serre (th. 4.1) donne qu'il existe F corps totalement réel galoisien sur \mathbb{Q} tel que $\rho|_{G_F}$ provienne d'une forme modulaire de Hilbert f pour F . Par [3], pour tout sous-corps F' de F tel que F/F' soit un groupe résoluble, $\rho|_{G_{F'}}$ provient d'une forme modulaire de Hilbert $f_{F'}$ pour F' . L'existence de E et de ι_p provient alors de l'existence des corps des coefficients pour les formes $f_{F'}$. Grâce au théorème de Brauer, on écrit ρ comme une somme virtuelle de tordues de représentations associées aux $f_{F'}$ et à ι_p . On définit ρ_ι comme la somme virtuelle correspondante, ι_p étant remplacé par ι . On prouve que

c'est une vraie représentation. Les propriétés de compatibilité résultent d'un théorème de Carayol et Taylor ([14], [58]), de Saito ([46], [45]), Breuil ([9]) et Berger ([4]).

7. PREUVE DU THÉORÈME

Pour p un premier, soit $S(p)$ l'énoncé : une représentation $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ de type S non ramifiée hors de p provient d'une forme modulaire. Pour p premier $\neq 2$ et k entier avec $2 \leq k \leq p+1$, soit $S(k, p)$ l'énoncé : une représentation $\bar{\rho} : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ de type S non ramifiée hors de p avec $k(\bar{\rho}) = k$ provient d'une forme modulaire. Par torsion par une puissance du caractère cyclotomique, on voit que $S(p)$ est la conjonction des énoncés $S(k, p)$. Comme, pour $p \neq 2$, $k(\bar{\rho})$ est pair pour toute représentation de type S qui est non ramifiée hors de p , on voit que l'on n'a à considérer $S(k, p)$ que pour k pair.

L'énoncé $S(2)$ est un théorème de Tate. La démonstration utilise la structure des groupes de décompositions et les minorations de discriminants. Serre a prouvé $S(3)$ par un argument analogue.

THÉORÈME 7.1 (Tate, Serre). — *Soit $\bar{\rho}$ une représentation continue de $G_{\mathbb{Q}}$ dans $\mathrm{GL}_2(\overline{\mathbb{F}}_2)$ (resp. $\mathrm{GL}_2(\overline{\mathbb{F}}_3)$) qui est non ramifiée hors de 2 (resp. 3). Alors $\bar{\rho}$ est réductible.*

Comme $S(2)$ est vrai, on suppose désormais p impair.

La stratégie pour prouver $S(p)$ pour tout p est une récurrence. Soit $\bar{\rho}$ une représentation de type S de poids $k(\bar{\rho}) = k$. Si $\bar{\rho}$ a une image résoluble, $\bar{\rho}$ est modulaire grâce au théorème de Langlands et Tunnell. Sinon, on relève $\bar{\rho}$ en une représentation $\rho : G_{\mathbb{Q}} \rightarrow \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ grâce à l'un des théorèmes du 5. On obtient grâce au théorème 6.1 un système compatible (ρ_{ι}) dont ρ fait partie. On choisit un ι_q de caractéristique q et on considère la réduction $\bar{\rho}_{\iota_q}$ modulo q de ρ_{ι_q} . Si elle est irréductible, l'hypothèse de récurrence entraîne qu'elle est modulaire. On vérifie que l'on peut appliquer le théorème 3.1 et ρ_{ι_q} est modulaire (si $\bar{\rho}_{\iota_q}$ est réductible, on doit vérifier les hypothèses du i) du théorème 3.1). Il en résulte que (ρ_{ι}) est modulaire, et donc $\bar{\rho}$.

PROPOSITION 7.2. — *i) $S(2, p)$ et $S(4, p)$ sont vrais.*

ii) On suppose $S(k, p)$ (donc $k \leq p+1$). Alors $S(k, q)$ est vrai pour tout premier q (tel que $k \leq q+1$).

En particulier, si $S(p)$ est vrai, $S(k, q)$ l'est pour $k \leq p+1$. Il en résulte que $S(q)$ l'est si $q < p$ et, pour prouver le théorème de Khare, il suffit de prouver $S(p)$ pour une infinité de premiers.

Preuve de la proposition

Prouvons $S(2, p)$. Soit $\bar{\rho}$ une représentation modulo p de type S et de poids 2. On relève en un système compatible (ρ_{ι}) de poids 2 non ramifié partout. On considère la

réduction modulo 3 de ρ_{ι_3} , pour ι_3 de caractéristique 3. On sait par la proposition précédente que cette réduction n'est pas irréductible. Comme son déterminant est $\overline{\chi}_3$, on voit qu'elle est isomorphe à $1 \oplus \overline{\chi}_3$. Alors comme ρ a un poids $2 \leq 3 - 1$, il résulte facilement de la théorie des schémas en groupes finis et plats que la restriction de ρ_{ι_3} à D_3 est ordinaire. Il résulte alors du i) du théorème 3.1 que ρ_{ι_3} provient d'une forme modulaire. Il en est de même de (ρ_λ) et par suite de $\overline{\rho}$.

Prouvons le ii). On relève $\overline{\rho}$ de caractéristique q en ρ minimal de type cristallin, puis on met ρ dans un système compatible (ρ_ι) . On choisit ι_p de caractéristique p et on considère la réduction $\overline{\rho_{\iota_p}}$ de ρ_{ι_p} . La restriction de ρ_{ι_p} à D_p est cristalline de poids de Hodge-Tate 0 et $k - 1$. Il résulte de [30] que si $k \leq p - 1$, on a $k(\overline{\rho_{\iota_p}}) = k$ et que, si la restriction de $\overline{\rho_{\iota_p}}$ à D_p est réductible, alors $\rho_{\iota_p|D_p}$ est ordinaire. Si $k = p + 1$, il résulte de [5] que soit $\rho_{\iota_p|D_p}$ est ordinaire, soit $k(\overline{\rho_{\iota_p}}) = 2$ et la restriction à D_p de $\overline{\rho_{\iota_p}}$ est irréductible.

Le i) entraîne que si $k(\overline{\rho_{\iota_p}}) = 2$, $\overline{\rho_{\iota_p}}$ est réductible, donc aussi sa restriction à D_p . On voit que l'on est dans l'un des cas :

- $\overline{\rho_{\iota_p}}$ est irréductible, et si $k = p + 1$, $\rho_{\iota_p|D_p}$ est ordinaire ;
- $\overline{\rho_{\iota_p}}$ est réductible et $\rho_{\iota_p|D_p}$ est ordinaire.

Dans le premier cas, $\overline{\rho_{\iota_p}}$ est modulaire par hypothèse. Un lemme facile prouve que si la restriction de $\overline{\rho_{\iota_p}}$ à $\mathbb{Q}(\sqrt{(-1)^{(p-1)/2}p})$ est réductible, $\rho_{\iota_p|D_p}$ est ordinaire. On voit que les hypothèses du ii) ou du iii) du théorème 3.1 sont vérifiées et ρ_{ι_p} est modulaire. Dans le cas $\overline{\rho_{\iota_p}}$ réductible, le i) de 3.1 donne la modularité de ρ_{ι_p} . On en déduit celle de $\overline{\rho}$. Le ii) est prouvé.

L'énoncé S(4, p) résulte du ii) et du fait que S(3) est vraie.

PROPOSITION 7.3. — S(6, p) est vrai pour tout $p \geq 5$; S(5) est vraie.

Preuve. La proposition précédente entraîne qu'il suffit de prouver S(6, 5), i.e. qu'il n'existe pas de représentations de type S de poids 6 et de caractéristique 5. Soit donc $\overline{\rho}_5$ une telle représentation. On la relève en une représentation ρ irréductible qui est minimale de type semi-stable. Elle est donc de poids 2 semi-stable en 5. La version potentielle de la conjecture de Serre (th. 4.1) et les i) et ii) du théorème 3.1 entraînent qu'il existe un corps totalement réel F galoisien sur \mathbb{Q} tel que la restriction de ρ à G_F provienne d'une forme modulaire de Hilbert pour F qui est de poids 2 et Steinberg en 5. La représentation ρ_{G_F} provient alors d'une variété abélienne. Un argument de descente donne que ρ provient d'une variété abélienne sur \mathbb{Q} qui a bonne réduction hors de 5 et est semi-stable en 5 ([34]). Mais Brumer et Kramer ont prouvé qu'une telle variété abélienne n'existe pas ([12], voir aussi Schoof [48]). Les résultats de Brumer-Kramer et Schoof utilisent des minorations de discriminants comme pour le théorème 7.1. L'étude de la ramification en p est beaucoup plus délicate ; ces résultats généralisent des théorèmes de Fontaine et Abrashkin ([28], [2]).

7.0.1. *Le cas général de conducteur 1.* — On a donc prouvé $S(p)$ pour $p \leq 5$. Avec la proposition 7.2, on voit qu’il suffit de prouver que, si $p \geq 5$ et si $S(p)$ est vrai, il existe $P > p$ tel que $S(P)$ soit vrai. On prend pour P le plus petit premier $> p$ qui n’est pas de Fermat : $P - 1$ est divisible par un premier ℓ impair, et on note ℓ^r la plus grande puissance de ℓ qui divise $P - 1$.

La théorie analytique des nombres premiers donne les inégalités :

LEMME 1. — *Posons $\ell^r = 2m + 1$. On a :*

$$\frac{P}{p} \leq \frac{2m + 1}{m + 1} - \frac{m}{p(m + 1)}, \quad p + 1 \geq \frac{m + 1}{2m + 1}(P - 1) + 2 = (P + 1) - \frac{m}{2m + 1}(P - 1).$$

Soit donc $\bar{\rho}$ une représentation de type S de caractéristique P , de poids k , $2 \leq k \leq P + 1$. Par la proposition 7.1, on connaît $S(k, P)$ pour $k \leq p + 1$.

Si la restriction au groupe de décomposition D_P est irréductible, la représentation obtenue en tordant $\bar{\rho}$ par une puissance du caractère cyclotomique convenable est de poids $P + 3 - k$. Il résulte de la première inégalité du lemme 1 que soit k , soit $P + 3 - k$ est $\leq p + 1$, et on conclut.

Supposons donc $\bar{\rho}$ ordinaire. On la relève en un système compatible (ρ_ι) de poids 2 grâce aux théorèmes 5.2 et 6.1. Soit ι_ℓ un plongement de caractéristique ℓ et soit $\overline{\rho_{\iota_\ell}}$ la réduction de ρ_{ι_ℓ} . Si la restriction de $\overline{\rho_{\iota_\ell}}$ à $\mathbb{Q}(\mu_\ell)$ est réductible, ou si $\overline{\rho_{\iota_\ell}}$ est irréductible à image résoluble donc modulaire par Langlands-Tunnell, on vérifie, de manière similaire à la proposition 7.2, que le théorème 3.1 s’applique, et ρ_{ι_ℓ} est modulaire, donc aussi $\bar{\rho}$. Sinon, on choisit un relèvement avec Nebentypus ρ' de $\overline{\rho_{\iota_\ell}}$ grâce au théorème 5.4. On choisit dans le théorème 5.4 l’entier i de sorte que le caractère $\hat{\gamma}\eta_q^i$ soit égal à ω_P^j avec :

$$j \in \left[\frac{m}{2m + 1}(P - 1), \frac{m + 1}{2m + 1}(P - 1) \right],$$

où ω_P est le représentant de Teichmüller du caractère cyclotomique $\text{Gal}(\mathbb{Q}_P/\mathbb{Q})(\mu_P) \rightarrow \mathbb{F}_P^*$ (rappelons que l’on choisit un isomorphisme de $\mu_{P-1}(\overline{\mathbb{Q}_\ell})$ sur $\mu_{P-1}(\overline{\mathbb{Q}_P})$). On prolonge ρ' en un système compatible $(\rho'_{\iota'})$ tel qu’il existe ι'_ℓ de caractéristique ℓ avec $\rho'_{\iota'_\ell}$ isomorphe à ρ' grâce à 6.1. On choisit ι'_P de caractéristique P de sorte que ι'_ℓ, ι'_P soient compatibles à l’identification $\mu_{P-1}(\overline{\mathbb{Q}_\ell}) \simeq \mu_{P-1}(\overline{\mathbb{Q}_P})$. Il résulte alors de [47] et [10] que la réduction $\overline{\rho'_{\iota'_P}}$ est de poids $j + 2$ ou $P + 1 - j$. Le choix de j et le lemme 1 font que soit $j + 2$, soit $P + 1 - j$ est $\leq p + 1$. On a donc que $\overline{\rho'_{\iota'_P}}$ est soit modulaire, soit réductible. On s’assure que les théorèmes «*MR*» permettent d’en déduire que $(\rho'_{\iota'})$, puis que (ρ_ι) sont modulaires, donc aussi $\bar{\rho}$. Ceci achève la preuve du théorème 2.1.

Le corollaire 1 résulte alors de la technique qui consiste à «*tuer la ramification*» ([34]). Si $q \neq 2$, on relève $\bar{\rho}$ de poids 2 et de niveau q en un système compatible (ρ_ι) qui est minimal de poids 2. On considère $\overline{\rho_{\iota_q}}$ pour ι_q de caractéristique q . On la relève en un système compatible qui est minimal, donc n’est pas ramifié en q (mais n’est plus en général de poids 2). On déduit du théorème 2.1 qu’il est modulaire.

Le corollaire 2 résulte du théorème et des théorèmes «*MR*» ([36]).

RÉFÉRENCES

- [1] *Périodes p -adiques*. Société Mathématique de France, Paris, 1994. Papers from the seminar held in Bures-sur-Yvette, 1988, Astérisque No. 223 (1994).
- [2] V.A. Abrashkin. Ramification in étale cohomology. *Invent. Math.*, 101(3):631–640, 1990.
- [3] J. Arthur, L. Clozel. *Simple algebras, base change, and the advanced theory of the trace formula*, volume 120 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1989.
- [4] L. Berger. Limites de représentations cristallines. *Compos. Math.*, 140(6):1473–1498, 2004.
- [5] L. Berger, H. Li, H.J. Zhu. Construction of some families of 2-dimensional crystalline representations. *Math. Ann.*, 329(2):365–377, 2004.
- [6] D. Blasius, J.D. Rogawski. Motives for Hilbert modular forms. *Invent. Math.*, 114(1):55–87, 1993.
- [7] G. Böckle. A local-to-global principle for deformations of Galois representations. *J. reine angew. Math.*, 509:199–236, 1999.
- [8] G. Böckle. Presentations of universal deformation rings. *preprint*, pages 1–27, 2005.
- [9] C. Breuil. Une remarque sur les représentations locales p -adiques et les congruences entre formes modulaires de Hilbert. *Bull. Soc. Math. France*, 127(3):459–472, 1999.
- [10] C. Breuil, A. Mézard. Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbf{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ en $l = p$. *Duke Math. J.*, 115(2):205–310, 2002. With an appendix by Guy Henniart.
- [11] S. Brueggeman. The nonexistence of certain Galois extensions unramified outside 5. *J. Number Theory*, 75(1):47–52, 1999.
- [12] A. Brumer, K. Kramer. Non-existence of certain semistable abelian varieties. *Manuscripta Math.*, 106(3):291–304, 2001.
- [13] K. Buzzard, R. Taylor. Companion forms and weight one forms. *Ann. of Math. (2)*, 149(3):905–919, 1999.
- [14] H. Carayol. Sur les représentations l -adiques associées aux formes modulaires de Hilbert. *Ann. Sci. École Norm. Sup. (4)*, 19(3):409–468, 1986.
- [15] H. Carayol. Sur les représentations galoisiennes modulo l attachées aux formes modulaires. *Duke Math. J.*, 59(3):785–801, 1989.
- [16] H. Carayol. Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet. In *p -adic monodromy and the Birch and Swinnerton-Dyer conjecture (Boston, MA, 1991)*, volume 165 of *Contemp. Math.*, pages 213–237. Amer. Math. Soc., Providence, RI, 1994.

- [17] H. Darmon, F. Diamond, R. Taylor. Fermat’s last theorem. In *Current developments in mathematics, 1995 (Cambridge, MA)*, pages 1–154. Internat. Press, Cambridge, MA, 1994.
- [18] P. Deligne. Formes modulaires et représentations ℓ -adiques. In *Séminaire Bourbaki 355*, pages 139–172. Lecture Notes in Math., Vol. 179. Springer, Berlin, 1971.
- [19] P. Deligne. Les constantes des équations fonctionnelles des fonctions L . In *Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*, pages 501–597. Lecture Notes in Math., Vol. 349. Springer, Berlin, 1973.
- [20] P. Deligne, J.-P. Serre. Formes modulaires de poids 1. *Ann. Sci. École Norm. Sup. (4)*, 7:507–530 (1975), 1974.
- [21] F. Diamond. An extension of Wiles’ results. In *Modular forms and Fermat’s last theorem (Boston, MA, 1995)*, pages 475–489. Springer, New York, 1997.
- [22] F. Diamond, R. Taylor. Lifting modular mod l representations. *Duke Math. J.*, 74(2):253–269, 1994.
- [23] F. Diamond, R. Taylor. Nonoptimal levels of mod l modular representations. *Invent. Math.*, 115(3):435–462, 1994.
- [24] L.V. Dieulefait. Existence of families of Galois representations and new cases of the Fontaine-Mazur conjecture. *J. reine angew. Math.*, 577:147–151, 2004.
- [25] B. Edixhoven. The weight in Serre’s conjectures on modular forms. *Invent. Math.*, 109(3):563–594, 1992.
- [26] B. Edixhoven. Serre’s conjecture. In *Modular forms and Fermat’s last theorem (Boston, MA, 1995)*, pages 209–242. Springer, New York, 1997.
- [27] J. Ellenberg. Serre’s conjecture over \mathbb{F}_9 . *Annals of Mathematics*, 161(3):1111–1142, 2005.
- [28] J.-M. Fontaine. Il n’y a pas de variété abélienne sur \mathbf{Z} . *Invent. Math.*, 81(3):515–538, 1985.
- [29] J.-M. Fontaine. Représentations l -adiques potentiellement semi-stables. *Astérisque*, (223):321–347, 1994. Périodes p -adiques (Bures-sur-Yvette, 1988).
- [30] J.-M. Fontaine, G. Laffaille. Construction de représentations p -adiques. *Ann. Sci. École Norm. Sup. (4)*, 15(4):547–608 (1983), 1982.
- [31] J.-M. Fontaine, B. Mazur. Geometric Galois representations. In *Elliptic curves, modular forms, & Fermat’s last theorem (Hong Kong, 1993)*, Ser. Number Theory, I, pages 41–78. Internat. Press, Cambridge, MA, 1995.
- [32] B.H. Gross. A tameness criterion for Galois representations associated to modular forms (mod p). *Duke Math. J.*, 61(2):445–517, 1990.

- [33] H. Hida. On p -adic Hecke algebras for GL_2 over totally real fields. *Ann. of Math. (2)*, 128(2):295–384, 1988.
- [34] C. Khare, J.-P. Wintenberger. On Serre’s reciprocity conjecture for 2-dimensional mod p representations of the Galois group $g_{\mathbb{Q}}$. *arXiv math.NT/0412076*, 2004.
- [35] C. Khare. On Serre’s modularity conjecture for 2-dimensional mod p representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ unramified outside p , preprint.
- [36] C. Khare. Serre’s modularity conjecture, preprint.
- [37] M. Kisin. Modularity for some geometric Galois representations. *preprint*, pages 1–21, 2005.
- [38] J. Manoharmayum. Serre’s conjecture for mod 7 Galois representations. In *Modular curves and abelian varieties*, volume 224 of *Progr. Math.*, pages 141–149. Birkhäuser, Basel, 2004.
- [39] B. Mazur. An introduction to the deformation theory of Galois representations. In *Modular forms and Fermat’s last theorem (Boston, MA, 1995)*, pages 243–311. Springer, New York, 1997.
- [40] L. Moret-Bailly. Groupes de Picard et problèmes de Skolem. I, II. *Ann. Sci. École Norm. Sup. (4)*, 22(2):161–179, 181–194, 1989.
- [41] R. Ramakrishna. Deforming Galois representations and the conjectures of Serre and Fontaine-Mazur. *Ann. of Math. (2)*, 156(1):115–154, 2002.
- [42] K.A. Ribet. Galois representations attached to eigenforms with Nebentypus. In *Modular functions of one variable, V (Proc. Second Internat. Conf., Univ. Bonn, Bonn, 1976)*, pages 17–51. Lecture Notes in Math., Vol. 601. Springer, Berlin, 1977.
- [43] K.A. Ribet. Report on mod l representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. In *Motives (Seattle, WA, 1991)*, volume 55 of *Proc. Sympos. Pure Math.*, pages 639–676. Amer. Math. Soc., Providence, RI, 1994.
- [44] K.A. Ribet, W.A. Stein. Lectures on Serre’s conjectures. In *Arithmetic algebraic geometry (Park City, UT, 1999)*, volume 9 of *IAS/Park City Math. Ser.*, pages 143–232. Amer. Math. Soc., Providence, RI, 2001.
- [45] T. Saito. Hilbert modular forms and p -adic Hodge theory. *preprint*, pages 1–40.
- [46] T. Saito. Modular forms and p -adic Hodge theory. *Invent. Math.*, 129(3):607–620, 1997.
- [47] D. Savitt. On a conjecture of Conrad, Diamond, and Taylor. *Duke Math. J.*, 128(1):141–197, 2005.
- [48] R. Schoof. Abelian varieties over \mathbb{Q} with bad reduction in one prime only. *Compos. Math.*, 141(4):847–868, 2005.

- [49] J.-P. Serre. Formes modulaires et fonctions zêta p -adiques. In *Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972)*, pages 191–268. Lecture Notes in Math., Vol. 350. Springer, Berlin, 1973.
- [50] J.-P. Serre. Valeurs propres des opérateurs de Hecke modulo l . In *Journées Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, 1974)*, pages 109–117. Astérisque, Nos. 24–25. Soc. Math. France, Paris, 1975.
- [51] J.-P. Serre. *Œuvres. Vol. III*. Springer-Verlag, Berlin, 1986. 1972–1984.
- [52] J.-P. Serre. Sur les représentations modulaires de degré 2 de $\text{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. *Duke Math. J.*, 54(1):179–230, 1987.
- [53] C.M. Skinner, A.J. Wiles. Residually reducible representations and modular forms. *Inst. Hautes Études Sci. Publ. Math.*, (89):5–126 (2000), 1999.
- [54] C.M. Skinner, A.J. Wiles. Base change and a problem of Serre. *Duke Math. J.*, 107(1):15–25, 2001.
- [55] C.M. Skinner, A.J. Wiles. Nearly ordinary deformations of irreducible residual representations. *Ann. Fac. Sci. Toulouse Math. (6)*, 10(1):185–215, 2001.
- [56] H.P.F. Swinnerton-Dyer. On l -adic representations and congruences for coefficients of modular forms. In *Modular functions of one variable, III (Proc. Internat. Summer School, Univ. Antwerp, 1972)*, pages 1–55. Lecture Notes in Math., Vol. 350. Springer, Berlin, 1973.
- [57] J. Tate. The non-existence of certain Galois extensions of \mathbf{Q} unramified outside 2. In *Arithmetic geometry (Tempe, AZ, 1993)*, volume 174 of *Contemp. Math.*, pages 153–156. Amer. Math. Soc., Providence, RI, 1994.
- [58] R. Taylor. On Galois representations associated to Hilbert modular forms. *Invent. Math.*, 98(2):265–280, 1989.
- [59] R. Taylor. On Galois representations associated to Hilbert modular forms. II. In *Elliptic curves, modular forms, & Fermat’s last theorem (Hong Kong, 1993)*, Ser. Number Theory, I, pages 185–191. Internat. Press, Cambridge, MA, 1995.
- [60] R. Taylor. On the meromorphic continuation of degree two L-functions. *Preprint*, pages 1–53, 2001.
- [61] R. Taylor. Remarks on a conjecture of Fontaine and Mazur. *J. Inst. Math. Jussieu*, 1(1):125–143, 2002.
- [62] R. Taylor. On icosahedral Artin representations. II. *Amer. J. Math.*, 125(3):549–566, 2003.
- [63] R. Taylor. Galois representations. *Ann. Fac. Sci. Toulouse Math. (6)*, 13(1):73–119, 2004.
- [64] R. Taylor, A. Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995.

- [65] T. Tsuji. Semi-stable conjecture of Fontaine-Jannsen: a survey. *Astérisque*, (279):323–370, 2002. Cohomologies p -adiques et applications arithmétiques, II.
- [66] J. Tunnell. Artin’s conjecture for representations of octahedral type. *Bull. Amer. Math. Soc. (N.S.)*, 5(2):173–175, 1981.
- [67] A. Wiles. Modular elliptic curves and Fermat’s last theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995.
- [68] J.-P. Wintenberger. On p -adic geometric representations of $G_{\mathbb{Q}}$. *preprint*, pages 1–9, 2005.

Jean-Pierre WINTENBERGER

Université Louis Pasteur et CNRS
I.R.M.A.

7, rue René Descartes

F–67084 STRASBOURG Cedex

E-mail : wintenb@math.u-psud.fr

ÉCARTS ENTRE NOMBRES PREMIERS SUCCESSIFS [d'après Goldston, Pintz, Yıldırım, ...]

par Emmanuel KOWALSKI

1. INTRODUCTION

Rien n'est plus commun que de remarquer l'importance de la suite infinie

$$2 = p_1 < 3 = p_2 < \cdots < p_n < p_{n+1} < \cdots$$

des nombres premiers dans l'imaginaire des mathématiciens et des arithméticiens tout particulièrement. Parmi les belles questions qui restent complètement ouvertes concernant les nombres premiers, celles qui ont trait aux *écarts* entre nombres premiers consécutifs

$$\gamma(n) = p_{n+1} - p_n \text{ pour } n \geq 1$$

ont exercé depuis longtemps une fascination particulière.

Le point de départ d'une analyse rigoureuse de la répartition des valeurs de $\gamma(n)$ est forcément le théorème des nombres premiers. Notons $\pi(X)$, comme d'habitude, le nombre de nombres premiers $p \leq X$; on a alors

$$\pi(X) \sim \frac{X}{\ln X}, \text{ quand } X \rightarrow +\infty.$$

De manière équivalente, le n -ième nombre premier p_n vérifie

$$p_n \sim n \ln n, \text{ quand } n \rightarrow +\infty,$$

et, «en moyenne», la distance $\gamma(n)$ entre nombres premiers consécutifs est de l'ordre de $\ln n$: on a

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{2 \leq n \leq N} \frac{\gamma(n)}{\ln n} = 1,$$

d'où en particulier :

$$\liminf_{n \rightarrow +\infty} \frac{\gamma(n)}{\ln n} \leq 1.$$

L'autre source de conjectures – qui n'a pas attendu la preuve du théorème des nombres premiers – consiste à «regarder» ce qui se passe dans des tables de nombres premiers. Il avait été observé⁽¹⁾ qu'un nombre relativement grand de paires (p_n, p_{n+1})

⁽¹⁾Dickson attribue cela à Polignac.

vérifie $p_{n+1} - p_n = 2$. En conséquence il a été conjecturé qu'une infinité de telles paires existe. C'est la conjecture des nombres premiers jumeaux : avec la notation ci-dessus, elle affirme que

$$(1) \quad \liminf_{n \rightarrow +\infty} \gamma(n) = 2.$$

Le thème particulier de ce rapport est celui de l'existence de petits écarts entre nombres premiers (par opposition, par exemple, avec la question de borner $\gamma(n)$ pour tout n). Après des progrès sporadiques, dus à Erdős, Rankin, Bombieri-Davenport [2], Huxley, Maier, en particulier, l'étude des petits écarts entre nombres premiers a été bouleversée durant l'année 2005. Goldston, Pintz et Yıldırım [11], [13], [14] ont en effet démontré :

THÉORÈME 1.1 (Goldston, Pintz, Yıldırım). — *On a*

$$\liminf_{n \rightarrow +\infty} \frac{\gamma(n)}{\ln n} = 0.$$

Le meilleur résultat précédemment connu était dû à Maier :

$$\liminf_{n \rightarrow +\infty} \frac{\gamma(n)}{\ln n} \leq 0,2484\dots$$

De plus, en admettant certaines hypothèses usuelles concernant la répartition des nombres premiers dans les progressions arithmétiques, on a des résultats vraiment spectaculaires :

THÉORÈME 1.2 (Goldston, Pintz, Yıldırım). — (1) *Si l'exposant de répartition θ au sens fort de la suite des nombres premiers vérifie $\theta > 1/2$, on a*

$$\liminf_{n \rightarrow +\infty} \gamma(n) < +\infty.$$

(2) *Si $\theta = 1$, alors*

$$\liminf_{n \rightarrow +\infty} \gamma(n) \leq 16.$$

L'exposant de répartition θ est défini ci-dessous (Définition 2.2). On sait que $\theta \geq 1/2$: c'est le célèbre théorème de Bombieri-Vinogradov.

Dans la section 7, nous présenterons des énoncés plus forts et d'autres résultats annoncés, concernant par exemple les écarts entre nombres de la forme $p_1 p_2$ avec $p_1 \neq p_2$ (c'est-à-dire, ayant exactement deux facteurs premiers).

Mais, pour commencer, nous allons expliquer la démonstration des théorèmes 1.1 et 1.2, après avoir présenté la base de la méthode telle qu'elle apparaît à l'heure actuelle.

Il faut signaler que le sujet est en pleine ébullition : tous les résultats sont encore sous forme de prépublications, et différentes variantes de la méthode apparaissent dans chacun d'entre eux. Malgré la malencontreuse aventure de 2003, aucun doute ne subsiste cependant quant à la validité des deux énoncés ci-dessus.

Notations. Rappelons quelques notations de théorie analytique des nombres.

– p désignera toujours un nombre premier, et n, m des entiers ≥ 1 .

– $\Lambda(n)$ désigne la fonction de von Mangoldt ; $\Lambda(n) = 0$ si n n'est pas une puissance d'un nombre premier p et $\Lambda(p^k) = \ln p$. De plus,

$$\psi(x; q, a) = \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \Lambda(n).$$

– $\mu(n)$ désigne la fonction de Möbius, $\tau(n)$ la fonction « nombre de diviseurs » et $\varphi(n)$ la fonction d'Euler. Pour $n = p^k$, $k \geq 1$, on a $\mu(p^k) = -1$ si $k = 1$, 0 sinon, $\tau(p^k) = k + 1$ et $\varphi(p^k) = p^k - p^{k-1}$.

– $f \star g$ désigne la convolution arithmétique de f et g :

$$f \star g(n) = \sum_{d|n} f(d)g(n/d) = \sum_{ab=n} f(a)g(b).$$

– \sum^b désigne une somme restreinte aux entiers n sans facteurs carrés.

– Les notations de Landau $f = O(g)$ et de Vinogradov $f \ll g$ sont considérées comme synonymes : $f(x) = O(g(x))$ pour tout $x \in D$ signifie qu'il existe une constante « implicite » $C \geq 0$ (le plus souvent, fonction d'autres paramètres explicitement mentionnés) telle que $|f(x)| \leq Cg(x)$ pour tout $x \in D$. Cette définition *diffère* de celle, topologique, de Bourbaki [4, Chap. V]. Par contre, nous utilisons les notations $f(x) \sim g(x)$ et $f = o(g)$ dans le sens asymptotique de loc. cit.

Remerciements. Je remercie D. Goldston, J. Pintz, C. Yıldırım, J. Sivak, R. de la Bretèche et É. Fouvry pour leurs remarques et corrections concernant les différents brouillons de ce texte.

2. RAPPELS DE QUELQUES NOTIONS DE CRIBLE

Cette section rappelle quelques définitions relatives aux méthodes de crible (voir par exemple [15] pour un traitement classique très complet, [18, Ch. 6], [19, Ch. 4] pour des exposés assez courts comportant les preuves des assertions de base, et [20], [5] pour deux survols dans ce séminaire).

Soit $\mathcal{A} = (a_n)$, $n \geq 1$, une suite de nombres réels positifs. On note

$$\mathcal{A}(x) = \sum_{n \leq x} a_n$$

sa fonction sommatoire.

Soit P un entier, qui sera le plus souvent le produit de certains nombres premiers bien choisis, par exemple le produit des nombres premiers $p < z$. Les *sommes criblées* correspondantes sont

$$\mathcal{A}(x, P) = \sum_{\substack{n \leq x \\ (n, P) = 1}} a_n.$$

L'objectif général du crible est d'estimer $\mathcal{A}(x, P)$ avec la plus grande généralité possible.

Pour aborder l'étude des sommes criblées $\mathcal{A}(x, P)$, les méthodes de « petit » crible sont inspirées par la formule de Legendre⁽²⁾ : on a

$$\mathcal{A}(x, P) = \sum_{d|P} \mu(d) \mathcal{A}_d(x), \quad \text{avec} \quad \mathcal{A}_d(x) = \sum_{\substack{n \leq x \\ n \equiv 0 \pmod{d}}} a_n.$$

Bien qu'elle ne soit pas exploitable directement en général car le nombre de termes (i.e., le nombre de diviseurs de P) est trop grand, cette formule suggère d'étudier les suites $\mathcal{A}_d = (a_{nd})$. Dans beaucoup de cas intéressants du point de vue de la théorie multiplicative des nombres, il se dégage naturellement une approximation de $\mathcal{A}_d(x)$ de la forme

$$(2) \quad \mathcal{A}_d(x) = g(d)X + r_d(\mathcal{A}; x)$$

où $X \geq 0$ est une fonction de x , $d \mapsto g(d)$ est une fonction arithmétique que l'on suppose multiplicative⁽³⁾ et $r_d(\mathcal{A}; x)$ est un « reste ».

En pratique, cette approximation est assez précise pour des valeurs de d assez grandes, disons $d < D$, c'est-à-dire que le reste total

$$\sum_{d < D} |r_d(\mathcal{A}; x)|$$

peut être estimé « suffisamment bien ». Plus précisément, la notion *d'exposant de répartition* est classiquement définie :

DÉFINITION 2.1. — *La suite \mathcal{A} a un exposant de répartition au sens faible $\geq \theta$ si, pour tout $A > 0$ et tout $\varepsilon > 0$, on a*

$$(3) \quad \sum_{d < D} |r_d(\mathcal{A}; x)| \ll \frac{X}{(\ln X)^A}$$

pour $x \geq 2$ et $D \leq x^{\theta - \varepsilon}$, la constante implicite dépendant au plus de A , ε et \mathcal{A} .

La méthode de Goldston, Pintz et Yıldırım requiert cruciallement des informations, non seulement pour une suite $\mathcal{A} = (a_n)$, mais pour les suites « décalées » (a_{n+t}) , où $t \geq 0$ est un entier quelconque. Cela mène à un renforcement naturel de la notion d'exposant de répartition.

Pour la définir, on généralise donc (2) en écrivant

$$(4) \quad \sum_{\substack{n \leq x \\ n \equiv t \pmod{d}}} a_n = g_t(d)X + r_d(\mathcal{A}; x, t)$$

où X est encore une fonction de x et $d \mapsto g_t(d)$ est pour tout t une fonction multiplicative dont la valeur ne dépend que de t modulo d .

⁽²⁾Que l'on peut voir, au choix, comme l'expression du principe d'inclusion–exclusion, ou de la formule $\sum_{d|n} \mu(d) = 0$ si $n \neq 1$.

⁽³⁾Au sens où $g(nm) = g(n)g(m)$ si $(n, m) = 1$.

DÉFINITION 2.2. — Une suite $\mathcal{A} = (a_n)$ de réels positifs a un exposant de répartition au sens fort $\geq \theta$ si pour tout $A > 0$ et tout $\varepsilon > 0$, on a

$$(5) \quad \sum_{d \leq D} \max_{t \pmod{d}} \max_{y \leq x} |r_d(\mathcal{A}; y, t)| \ll \frac{X}{(\ln X)^A},$$

pour $x \geq 2$ et $D \leq x^{\theta - \varepsilon}$, la constante implicite pouvant dépendre de A , ε et \mathcal{A} .

Pour le théorème 1.2, la suite concernée est la suite $(\Lambda(n))$, ou bien la fonction caractéristique des nombres premiers.

On va voir dans la preuve du théorème 1.2 qu’une notion plus faible se dégage. Comme cela pourrait fournir une voie plus accessible à la preuve inconditionnelle de ce résultat, il semble utile de l’isoler.

DÉFINITION 2.3. — Une suite $\mathcal{A} = (a_n)$ de réels positifs a un P -exposant de répartition $\geq \theta$ si pour tout $A > 0$, tout $\varepsilon > 0$ et tout polynôme primitif $F \in \mathbf{Z}[X]$ de degré ≥ 1 on a

$$\sum_{d \leq D}^b \sum_{\substack{\nu \pmod{d} \\ F(\nu)=0}} \max_{y \leq x} |r_d(\mathcal{A}; y, \nu)| \ll \frac{X}{(\ln X)^A},$$

pour $x \geq 2$ et $D \leq x^{\theta - \varepsilon}$, la constante implicite pouvant dépendre de A , ε , \mathcal{A} et F .

Remarque 2.4. — Une prépublication de Pintz et Motohashi [21], parue après la première version de ce survol, introduit explicitement une variante de cette dernière notion dans le cas de la suite des nombres premiers, avec une plus grande flexibilité potentiellement intéressante.

Le cas le plus important pour notre propos est celui de la suite $a_n = \Lambda(n)$. Dans ce cas on a pour $t \geq 0$

$$\sum_{\substack{n \leq x \\ n \equiv t \pmod{d}}} a_n = \sum_{\substack{n \leq x \\ n \equiv t \pmod{d}}} \Lambda(n) = \psi(x; d, t),$$

et si $(d, t) = 1$ on a l’approximation

$$(6) \quad \sum_{\substack{n \leq x \\ n \equiv t \pmod{d}}} a_n = \frac{x}{\varphi(d)} + r_d(\mathcal{A}; x, t)$$

avec la borne

$$r_d(\mathcal{A}; x, t) \ll x(\ln x)^{-A}$$

valide pour tout $x \geq 2$, $d \geq 1$, t modulo d (tel que $(d, t) = 1$) et tout $A > 0$, la constante implicite (ineffective) dépendant seulement de A ; c’est le théorème de la progression arithmétique sous la forme donnée par Siegel-Walfisz ([18, Cor. 5.29] par exemple).

Tenant compte des cas $(d, t) \neq 1$, on peut définir

$$X = x, \quad g_t(d) = \begin{cases} 0 & \text{si } (d, t) \neq 1, \\ \frac{1}{\varphi(d)} & \text{si } (d, t) = 1, \end{cases}$$

pour avoir (4) avec

$$r_d(\mathcal{A}; x, t) \ll x(\ln x)^{-A}$$

dans tous les cas.

Tel quel, cela ne donne qu'un exposant de répartition nul. L'ingrédient crucial pour cribler efficacement une telle suite est le théorème de Bombieri-Vinogradov (voir par exemple [18, Ch. 17]) :

THÉORÈME 2.5. — *Pour tout $A > 0$, il existe $B > 0$ tel que*

$$\sum_{d \leq D} \max_{\substack{(a,d)=1 \\ a \pmod{d}}} \max_{y \leq x} \left| \psi(y; d, a) - \frac{y}{\varphi(d)} \right| \ll \frac{x}{(\ln x)^A}$$

pour $x \geq 2$ et $D \leq x^{1/2}(\ln x)^{-B}$, la constante implicite dépendant uniquement de A .

La restriction à $(a, d) = 1$ étant inoffensive, cela implique que la suite $(\Lambda(n))$ a un exposant de répartition $\geq 1/2$. La limite à $1/2$ est bien comprise : c'est ce que donnerait trivialement l'Hypothèse de Riemann Généralisée pour les fonctions L de Dirichlet. Aller au-delà revient donc, d'une manière ou d'une autre, à obtenir des informations sur la répartition des zéros de $L(\chi, s)$ par rapport à χ variable (voir la section 8). Dit d'une autre manière, analytiquement, il faut aller au-delà de l'inégalité générale du Grand Crible qui est l'un des ingrédients cruciaux de la preuve (un autre étant (6)).

Remarque 2.6. — (1) On appellera, suivant l'usage, *Conjecture d'Elliott-Halberstam* la conjecture qui énonce⁽⁴⁾ que $(\Lambda(n))$ a un exposant de répartition (au sens fort⁽⁵⁾) égal à 1.

(2) Les généralisations du théorème de Bombieri - Voir la dernière section de ce rapport pour quelques remarques concernant des suites pour lesquelles on connaît un exposant de répartition $> 1/2$ (faible ou fort).

(3) Le P -exposant de répartition semble être une notion nouvelle. Bien entendu, si une suite a un exposant de répartition $\geq \theta$ au sens fort, il en est de même pour le P -exposant. Trouver des suites non triviales pour lesquelles le P -exposant de répartition dépasse la limite (universelle) du grand crible semble un problème intéressant.

L'exemple de crible qui est sous-jacent à la méthode de Goldston, Pintz et Yıldırım est lié à la «conjecture des k -uplets» de Hardy et Littlewood. Soit $k \geq 2$ et $\mathbf{h} = (h_1, \dots, h_k)$ un k -uplet d'entiers ≥ 0 distincts. Le problème est de trouver les entiers n tels que $n + h_1, \dots, n + h_k$ soient tous premiers, autrement dit de trouver les entiers $n \geq 1$ tels que

$$F_{\mathbf{h}}(n) = (n + h_1) \cdots (n + h_k)$$

⁽⁴⁾Ce n'est pas la version originale de Elliott et Halberstam, qui est fausse, comme l'ont démontré Friedlander et Granville.

⁽⁵⁾Dans la suite, on omettra parfois de préciser «au sens fort», qui sera sous-entendu.

possède k facteurs premiers exactement. On notera $\pi(X, \mathbf{h})$ le nombre des $n \leq X$ vérifiant cette condition.

Puisque tout entier de ce type assez grand (en fonction de $\varepsilon > 0$) a tous ses facteurs premiers $> n > F_{\mathbf{h}}(n)^{1/k-\varepsilon}$, les méthodes de crible fournissent aisément une borne supérieure (voir par exemple [18, Th. 6.7], [15, Th. 5.7]) :

THÉORÈME 2.7. — *On a*

$$\pi(X, \mathbf{h}) \leq 2^k k! \mathfrak{S}(\mathbf{h}) X (\ln X)^{-k} \left(1 + O\left(\frac{\ln \ln X}{\ln X}\right) \right)$$

pour $X \geq 3$, où

$$(7) \quad \mathfrak{S}(\mathbf{h}) = \prod_p \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p} \right) \left(1 - \frac{1}{p} \right)^{-k},$$

avec $\nu_{\mathbf{h}}(p)$ la cardinalité de l'ensemble $\{h_1 \pmod{p}, \dots, h_k \pmod{p}\}$ des réductions modulo p des composantes de \mathbf{h} , ce produit étant absolument convergent.

Comme dans [11], [12], [13], on dira que \mathbf{h} est *admissible* si les composantes de \mathbf{h} sont distinctes et si $\mathfrak{S}(\mathbf{h}) \neq 0$. Cela signifie qu'il n'existe pas de nombre premier p avec $\nu_{\mathbf{h}}(p) = p$, c'est-à-dire qu'il n'existe pas de nombre premier p qui divise toujours l'un au moins des entiers $n + h_1, \dots, n + h_k$. Des raisonnements heuristiques (effectués, dans ce cas, par Hardy et Littlewood, voir [18, 13.1] ou [20, 2], par exemple, pour une présentation moderne) suggèrent que la borne supérieure obtenue est du bon ordre de grandeur⁽⁶⁾ si \mathbf{h} est admissible, et plus précisément :

CONJECTURE 2.8 (Hardy–Littlewood). — *Si \mathbf{h} est admissible, on a*

$$\pi(X, \mathbf{h}) \sim \mathfrak{S}(\mathbf{h}) X (\ln X)^{-k} \text{ quand } X \rightarrow +\infty.$$

Bien entendu, le cas $\mathbf{h} = (0, 2)$ correspond à la conjecture des nombres premiers jumeaux.

Le facteur⁽⁷⁾ $\mathfrak{S}(\mathbf{h})$ joue un rôle important dans la partie inconditionnelle des travaux de Goldston, Pintz et Yıldırım (le théorème 1.1). Plus précisément, il est nécessaire de connaître sa valeur moyenne lorsque \mathbf{h} parcourt les k -uplets admissibles $\mathbf{h} = (h_i)$ avec $1 \leq h_i \leq h$ pour k fixé. Celle-ci a été calculée par Gallagher [9].

PROPOSITION 2.9. — *Pour $h \geq 1$ et $k \geq 1$ des entiers, soit*

$$(8) \quad \mathfrak{S}(h, k) = \frac{1}{h^k} \sum_{\substack{\mathbf{h}=(h_1, \dots, h_k) \\ 1 \leq h_i \leq h}}^* \mathfrak{S}(\mathbf{h})$$

⁽⁶⁾ Cela peut se « deviner » en disant que chaque $n + h_i$ a une « probabilité » $1/(\ln n)$ d'être premier et en supposant les différentes conditions indépendantes.

⁽⁷⁾ Appelé parfois la « série singulière » par référence à la terminologie employée dans la méthode du cercle de Hardy et Littlewood.

où \sum^* signifie que la somme est restreinte aux k -uplets admissibles⁽⁸⁾.

Pour tout k fixé, on a

$$\lim_{h \rightarrow +\infty} S(h, k) = 1.$$

Cela veut dire que la valeur moyenne de $\mathfrak{S}(\mathbf{h})$ est 1, ce qui est cohérent avec ce qu'on attend d'un produit eulérien. L'estimation uniforme suivante est également utile : il existe une constante $b_k \geq 0$ telle que

$$(9) \quad \mathfrak{S}(\mathbf{h}) \ll (\ln \ln 10h)^{b_k},$$

pour tout k -uplet \mathbf{h} tel que $1 \leq h_i \leq h$ pour tout i , la constante implicite ne dépendant que de k (voir par exemple [13, Lemma 6]).

Remarque 2.10. — Ce résultat de Gallagher est lié à une conjecture concernant la répartition de $\gamma(n)$, à savoir que la suite $(\gamma(n)/\ln n)$ est, pour $n \rightarrow +\infty$, équirépartie pour la mesure $e^{-t} dt$ sur $[0, +\infty[$ («distribution de Poisson», correspondant à des nombres purement aléatoires et sans corrélations). On s'attend donc à avoir

$$(10) \quad \lim_{N \rightarrow +\infty} \frac{1}{N} |\{n \leq N \mid \alpha \leq \frac{\gamma(n)}{\ln n} \leq \beta\}| = \int_{\alpha}^{\beta} e^{-t} dt$$

pour tout $0 \leq \alpha < \beta$. Le lien avec la proposition 2.9 est que celle-ci permet de démontrer qu'une forme forte et uniforme de la conjecture des k -uplets implique (10), en calculant les moments nécessaires.

La loi de Poisson est aussi la répartition que l'on attend pour la suite des écarts normalisés entre valeurs propres du laplacien sur la courbe modulaire $SL(2, \mathbf{Z}) \backslash \mathbf{H}$; rappelons [20] que la répartition des écarts normalisés des zéros de $\zeta(s)$ est (conjecturalement) très différente⁽⁹⁾.

3. PRINCIPE DE LA MÉTHODE

Dans cette section, nous allons présenter – comme l'auteur de ce rapport est parvenu à le comprendre – le principe de base de la démonstration des théorèmes 1.1 et 1.2. La description tente de présenter le raisonnement le plus naturellement du monde, mais n'en respecte pas nécessairement la chronologie exacte. Le preprint [12] fournit plus de détails à ce sujet.

La première chose à faire est de considérer que la conjecture des k -uplets est «plus fondamentale» que l'énoncé du théorème 1.1 par exemple. On va donc chercher à aborder des formes plus faibles de cette conjecture. Il est classique d'essayer de le faire

⁽⁸⁾En particulier, dont les composantes h_i sont distinctes ; par contre, les composantes ne sont pas ordonnées.

⁽⁹⁾L'analogie du théorème 1.1 pour l'espacement moyen des parties imaginaires des zéros de $\zeta(s)$ n'est pas connu ; d'après un résultat – en fait beaucoup plus fort – de Conrey et Iwaniec, cela impliquerait qu'il n'existe pas de zéros de Landau-Siegel pour les fonctions L de Dirichlet.

en cherchant des k -uplets «presque premiers». La première idée nouvelle est *d'affaiblir différemment la conjecture*. Soit donc \mathbf{h} un k -uplet admissible. On cherchera des n tels que le k -uplet $(n + h_1, \dots, n + h_k)$ contienne *au moins* deux nombres premiers. Lorsque c'est le cas, on a évidemment «trouvé» deux nombres premiers tels que $p < p'$ et $p' - p \leq \max(h_i - h_j)$. Le théorème 1.2 en sera une conséquence ; quant au théorème 1.1, un seul k -uplet ne suffisant pas, on combinera à l'aide de la proposition 2.9 l'effet de tous ceux tels que $h_i \leq h$, avec $h = \delta \ln n \dots$

Un intérêt de cet affaiblissement (qui n'en est un que si $k \geq 3$, d'ailleurs) est qu'un «détecteur» manipulable de la condition voulue peut se décrire assez facilement. Notons $n \triangleleft \mathbf{h}$ la relation

«Il y a au moins deux (puissances de) nombres premiers parmi $n + h_1, \dots, n + h_k$ », et posons

$$\vartheta_{\mathbf{h}}(n) = \sum_{1 \leq i \leq k} \Lambda(n + h_i).$$

On a alors, si $h_i \leq n$:

$$(11) \quad \vartheta_{\mathbf{h}}(n) > \ln 2n \text{ implique } n \triangleleft \mathbf{h}.$$

On va tenter de comparer⁽¹⁰⁾ deux sommes, à savoir

$$Q_g = \sum_{N < n \leq 2N} \vartheta_{\mathbf{h}}(n) E(n) \quad \text{et} \quad Q_p = (\ln 3N) \sum_{N < n \leq 2N} E(n)$$

où $N \geq 1$ et les poids $E(n) \geq 0$ sont à notre disposition. Le détecteur ci-dessus fournit la minoration :

$$(k - 1)(\ln 3N) \sum_{\substack{N < n \leq 2N \\ n \triangleleft \mathbf{h}}} E(n) \geq (Q_g - Q_p),$$

en particulier, si $Q_g > Q_p$, il existe au moins une valeur de n qui convient dans l'intervalle $N < n \leq 2N$.

C'est dans le choix des coefficients $E(n)$ que se trouve la clé du succès des travaux de Goldston, Pintz et Yıldırım. Il n'y a rien d'évident à ce qu'il existe un choix impliquant un résultat aussi fort que le théorème 1.2, et cette méthode de détection, suivant le problème, est plus ou moins efficace⁽¹¹⁾. *A posteriori*, l'approche peut sembler «évidente» (si le rédacteur de ce rapport a bien fait son travail...), mais comme souvent en théorie analytique des nombres, il faut combiner deux ou trois idées significatives pour obtenir un nouveau résultat vraiment convaincant.

⁽¹⁰⁾Sous la forme plus précise qui sera abordée ci-dessous, Goldston, Pintz et Yıldırım attribuent cette formulation de leur méthode à Granville et Soundararajan.

⁽¹¹⁾Par exemple, l'analogie du théorème 1.1 pour les écarts $p_{n+2} - p_n$, que l'on est tenté d'aborder en considérant $\vartheta_{\mathbf{h}}(n) - 2 \ln 3N$, n'est pas connu inconditionnellement ; il l'est cependant sous la conjecture d'Elliott-Halberstam, voir théorème 7.3.

On cherche donc alors à minorer Q_g et majorer Q_p . On ne peut réellement espérer qu'une chose : que Q_g et Q_p soient du même ordre de grandeur, mais avec

$$\rho = \limsup_{N \rightarrow +\infty} \frac{Q_g}{Q_p} > 1.$$

Autrement dit : *la réussite dépendra d'une inégalité numérique*⁽¹²⁾. Même si, avec le type de coefficients $E(n)$ discutés ci-dessous, il est assez aisé de deviner la valeur de ρ (sous la conjecture 2.8, et éventuellement d'autres) et donc de justifier de manière plausible le choix effectué, la difficulté est de parvenir à démontrer rigoureusement que $\rho > 1$.

Remarque 3.1. — La méthode de Bombieri et Davenport [2] (réinterprétée par Goldston [10]) peut essentiellement se ramener à ce cadre, en changeant le détecteur et avec $E(n) = \Lambda(n)$: il s'agit grosso modo de démontrer que

$$Q_1 = \sum_{N < n \leq 2N} \left(\sum_{k=1}^h \alpha(k) \Lambda(n+2k) \right) \Lambda(n),$$

pour certains coefficients $\alpha(k)$ bien choisis, est suffisamment proche de sa valeur attendue

$$Q_2 = \sum_{N < n \leq 2n} \left(\sum_{k=1}^h \alpha(k) \mathfrak{S}(0, 2k) \right) \Lambda(n) \sim \left(\sum_{k=1}^h \alpha(k) \mathfrak{S}(0, 2k) \right) N$$

(donnée par la conjecture de Hardy-Littlewood) pour impliquer l'existence de nombres premiers à distance $\leq 2h$, lorsque h est de l'ordre de $\frac{1}{4} \log N$. Il faut noter qu'ici les deux sommes Q_1 et Q_2 sont asymptotiquement égales (conjecturalement), c'est-à-dire que $\rho = 1$.

Le coefficient $1/4$ est lié à l'exposant de répartition (au sens fort) des suites de nombres premiers décalés, et est donné par le théorème de Bombieri-Vinogradov. En particulier, comme l'observe Goldston [10], cette méthode démontrait déjà qu'une certaine forme de la conjecture d'Elliott-Halberstam implique le théorème 1.1 (*pas* le théorème 1.2).

La somme Q_g porte sur un support limité aux entiers n tels que $\{n+h_1, \dots, n+h_k\}$ contienne un nombre premier (ou une puissance...) au moins, alors que Q_p n'a pas cette restriction a priori. Il est donc naturel de demander à $E(n)$ d'imposer une restriction similaire, afin d'ajuster la « densité » relative espérée pour Q_p à celle de Q_g .

Bien entendu, prendre

$$E'(n) = \Lambda(n+h_1) \cdots \Lambda(n+h_k)$$

est le meilleur choix en principe : conjecturalement, la somme Q_g construite avec $E'(n)$ compte exactement (avec multiplicité k) les k -uplets de nombres premiers. Mais – c'est

⁽¹²⁾Cela peut être contrasté avec la méthode de « mollification » utilisée, par exemple, par Selberg pour minorer le nombre de zéros critiques de $\zeta(s)$. Le succès ne dépend alors que d'avoir le bon ordre de grandeur du second moment mollifié.

toute la question que l'on cherche à résoudre – on ne saurait pas démontrer rigoureusement les estimations nécessaires. On va donc introduire des approximations de $E'(n)$ qui seront plus manipulables⁽¹³⁾.

La recherche de la bonne approximation a procédé par tâtonnements. Il est classique d'approcher la fonction Λ en écrivant

$$\Lambda(n) = \sum_{d|n} \mu(d) \left(\ln \frac{n}{d} \right)$$

et en utilisant une convolution tronquée

$$\Lambda(n, x) = \sum_{\substack{d|n \\ d \leq x}} \mu(d) \left(\ln \frac{x}{d} \right),$$

où $x \geq 1$ est un paramètre, que l'on souhaite pouvoir prendre le plus grand possible.

La tentation est grande de choisir comme coefficients

$$E_0(n) = \Lambda(n + h_1, x) \cdots \Lambda(n + h_k, x).$$

Un premier obstacle apparaît : il n'est plus vrai que $E_0(n) \geq 0$; cela rend impossible de déduire d'une inégalité $Q_g > Q_p$ qu'il existe n tel que $n \triangleleft \mathbf{h}$. On peut vouloir contourner le problème en prenant $E_0(n)^2$; c'est en effet la précédente approche de Goldston et Yıldırım. Elle s'avère insuffisante pour démontrer le théorème 1.1⁽¹⁴⁾.

Il s'avère que trouver une bonne approximation nécessite de considérer le problème plus globalement en rapport avec l'objectif recherché par la méthode considérée : puisqu'on se contente (maintenant) d'avoir au moins deux nombres premiers dans les k -uplets considérés, il peut être (et il est) préférable de faire appel à un poids qui tienne compte de ce relâchement. Plutôt que $E'(n)$, qui « compte » des k -uplets d'entiers tous premiers, on utilise donc comme référence un poids qui sélectionne les k -uplets $(n + h_1, \dots, n + h_k)$ d'entiers dont le produit n'a que peu de facteurs premiers, mais éventuellement plus que k .

Précisément, l'innovation critique de Goldston, Pintz et Yıldırım est donc d'utiliser un poids (de type « crible ») que l'on utiliserait plutôt pour sélectionner les entiers n tels que $F_{\mathbf{h}}(n)$ ait au plus $k + \ell$ facteurs premiers, où

$$F_{\mathbf{h}}(n) = (n + h_1) \cdots (n + h_k)$$

et où le paramètre $\ell \geq 1$ est également disponible pour être optimisé. Si ℓ reste suffisamment petit par rapport à k , une grande proportion des k -uplets ainsi sélectionnés vérifieront la relation $n \triangleleft \mathbf{h}$ (même s'ils ont bien entendu davantage que $k + \ell$ facteurs premiers, puisque le crible n'est pas parfait). L'interprétation exacte de ce paramètre reste mystérieuse ; peut-être peut-on y voir une forme de « lissage » des k -uplets presque premiers.

⁽¹³⁾Là aussi, on peut penser à la méthode de mollification pour l'étude des zéros de fonctions L .

⁽¹⁴⁾Mais c'est un des outils arithmétiques de la preuve du théorème de Green-Tao [17].

Pour concrétiser cela, [11] utilise le poids

$$E(n) = \Lambda_{k+\ell}(F_{\mathbf{h}}(n), x)$$

où $\Lambda_m(n, x)$, pour tout entier $m \geq 1$, est une troncation de la m -ième fonction de von Mangoldt Λ_m , par analogie avec le cas $\Lambda = \Lambda_1$:

$$\Lambda_m(n) = (\mu \star \ln^m)(n) = \sum_{d|n} \mu(d) \left(\ln \frac{n}{d}\right)^m \quad \text{et} \quad \Lambda_m(n, x) = \sum_{\substack{d|n \\ d \leq x}} \mu(d) \left(\ln \frac{x}{d}\right)^m.$$

Le choix de $\Lambda_m(n)$ dans l'étude de la répartition des nombres «presque premiers» dans une suite est largement justifié par la théorie du crible, en particulier par le «crible asymptotique» de Bombieri où ces fonctions ont fait leur apparition. Il est facile de voir que Λ_m est supportée sur les entiers ayant au plus m facteurs premiers (sans multiplicité)⁽¹⁵⁾, et Bombieri est parvenu à démontrer (entre autres) que la conjecture d'Elliott-Halberstam implique la formule asymptotique

$$\sum_{n \leq X} \Lambda(n) \Lambda_m(n+2) \sim m \mathfrak{S}(0, 2) X (\log X)^{m-1} \quad \text{quand } X \rightarrow +\infty,$$

pour tout $m \geq 2$ ⁽¹⁶⁾.

Remarque 3.2. — La remarque qui suit est assez gratuite, mais puisque l'arithmétique est une science expérimentale, il serait dommage de s'en passer. Considérons le 6-uplet admissible $\mathbf{h} = (7, 11, 13, 17, 19, 23)$. Pour $X = 10,000,000$, et $1 \leq k \leq 6$, voici le nombre N_k d'entiers $n \leq X$ tels que $\{n+7, \dots, n+23\}$ contienne au moins k nombres premiers, et pour $6 \leq \omega \leq 11$, le nombre M_ω d'entiers $n \leq X$ tels que $F_{\mathbf{h}}(n)$ ait ω facteurs premiers (sans multiplicité) :

k	N_k	k	N_k	ω	M_ω	ω	M_ω
1	2,917,127	2	890,130	6	19	7	381
3	160,592	4	18,479	8	3281	9	16955
5	1,092	6	17	10	58947	11	150090

Remarque 3.3. — Concernant l'introduction cruciale du paramètre ℓ , un travail de Heath-Brown [16], lui-même inspiré par des idées de Selberg [22, §18, §23], est cité comme influence dans [11], [12].

Peut-être une raison de l'efficacité de ce paramètre, même petit – sous la conjecture d'Elliott-Halberstam, on prendra $\ell = 1$ –, est-elle à chercher dans le fait que demander que $F_{\mathbf{h}}(n)$ ait $k + \ell$ facteurs premiers avec $\ell \neq 0$ n'impose plus, en particulier, que la parité du nombre de facteurs premiers soit constante.

⁽¹⁵⁾Cela découle par récurrence de $\Lambda_1 = \Lambda$ et $\Lambda_{m+1} = (\ln)\Lambda_m + \Lambda \star \Lambda_m$.

⁽¹⁶⁾Friedlander et Iwaniec ont considérablement développé ce crible asymptotique, en fournissant des variantes qui peuvent traiter le cas $m = 1$ en ajoutant des hypothèses bilinéaires aux conditions classiques, voir [20].

Dans [11], Goldston, Pintz et Yıldırım évaluent donc

$$Q_g = \sum_{N < n \leq 2N} \vartheta_{\mathbf{h}}(n) \Lambda_{k+\ell}(F_{\mathbf{h}}(n), x)^2$$

et

$$Q_p = (\ln 3N) \sum_{N < n \leq 2N} \Lambda_{k+\ell}(F_{\mathbf{h}}(n), x)^2.$$

Plutôt que de présenter ces calculs⁽¹⁷⁾, nous allons utiliser un poids différent qu'ils emploient avec Graham dans [13]. Son avantage est de permettre une évaluation *élémentaire* de Q_g et Q_p alors que, dans [11], il est nécessaire de faire appel à des intégrales multiples de fractions rationnelles en la fonction zéta de Riemann et conséquemment à la zone sans zéro de celle-ci, ce qui est relativement délicat.

L'idée dans [13] est de voir le problème de maximiser Q_g/Q_p comme proche du type de problème d'optimisation caractéristique du crible de Selberg, et c'est d'ailleurs – pour un problème relativement similaire – ce qu'avait fait Selberg lui-même. On cherche alors

$$E(n) = \lambda(F_{\mathbf{h}}(n))^2 \quad \text{où} \quad \lambda(n) = \sum_{d|n} \lambda_d,$$

les coefficients λ_d étant à notre disposition, avec la seule contrainte que $\lambda_d = 0$ pour $d > x$, où x (le «niveau», dans le langage du crible) est un paramètre jouant évidemment le même rôle que dans $\Lambda_m(n, x)$ – qui est de la forme postulée ici.

On voit que Q_g et Q_p deviennent des formes quadratiques :

$$(12) \quad Q_g = \sum_{N < n \leq 2N} \vartheta_{\mathbf{h}}(n) \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_d \right)^2,$$

$$(13) \quad Q_p = (\ln 3N) \sum_{N < n \leq 2N} \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_d \right)^2.$$

La situation typique du crible de Selberg est de minimiser une forme quadratique avec une contrainte linéaire. Ici, le problème de maximiser une forme quadratique par rapport à une autre est différent et n'admet pas de solution «formelle» aussi générale.

Remarque 3.4. — Il est amusant de noter que cette méthode de comparaison de formes quadratiques se trouve dans les travaux de Selberg et Heath-Brown déjà mentionnés, et qu'une variante a aussi été introduite récemment par Soundararajan pour le problème de trouver des grandes valeurs de $|\zeta(\frac{1}{2} + it)|$ ou de $|L(f, \frac{1}{2})|$ lorsque f parcourt une famille de fonctions L .

Il y a aussi un goût de la méthode d'amplification de Duke, Friedlander et Iwaniec. Puisqu'il est important de donner un nom aux choses, on peut suggérer «méthode de

⁽¹⁷⁾Les résultats sont asymptotiquement identiques à ceux qui sont obtenus ci-dessous.

résonance»⁽¹⁸⁾, ou – plus poétiquement – «méthode d'écémage»⁽¹⁹⁾ pour ce procédé de détection de valeurs extrêmes.

Dans son cas, Selberg avait procédé en minimisant la «petite» forme quadratique sous la contrainte $\lambda_1 = 1$, puis avait évalué la «grande» forme quadratique pour ce choix de valeurs. La stratégie dans [13] est légèrement différente ; précisément, on obtient un $\lambda_{d,0}$ de cette manière, mais ensuite il est modifié «brutalement» pour incorporer la souplesse du paramètre ℓ . (Le choix initial correspondant en effet à $\ell = 0$.) Il est utile de préciser que ce paramètre est spécifique à la situation des écarts entre nombres premiers, et n'a pas d'analogue dans les autres cas mentionnés.

4. CHOIX ET ESTIMATION DES DEUX FORMES QUADRATIQUES

Considérons d'abord de manière assez générale une suite $\mathcal{A} = (a_n)$ de réels positifs tels que la série $\sum a_n$ converge, et un k -uplet admissible \mathbf{h} . On va étudier l'expression

$$(14) \quad Q = \sum_n a_n \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_d \right)^2,$$

vue comme forme quadratique en les variables λ_d , $d < x$.

On fait l'hypothèse que λ_d soit supportée sur les entiers $d < x$ sans facteurs carrés. Cette restriction est justifiée par l'expérience du crible de Selberg, et simplifie techniquement les manipulations qui vont suivre⁽²⁰⁾.

Dans cette section, on suppose qu'un analogue de (15) est valide, sous la forme suivante :

$$(15) \quad \sum_{n \equiv t \pmod{d}} a_n = g_t(d)X + r_d(\mathcal{A}, t),$$

pour $d \geq 1$ et t modulo d , où les fonctions arithmétiques $d \mapsto g_t(d)$ sont multiplicatives et de plus $g_t(d)$ ne dépend que de t modulo d . Pour commencer, cette décomposition est formelle, mais on verra plus bas quelles hypothèses sont faites sur les restes $r_d(\mathcal{A}, t)$ pour les contrôler.

⁽¹⁸⁾Où les coefficients formeraient un résonateur ; Soundararajan utilise ce nom.

⁽¹⁹⁾Où les coefficients formeraient un écémair.

⁽²⁰⁾En particulier, une fonction arithmétique f multiplicative supportée sur les entiers sans facteurs carrés est caractérisée par $f(p)$ pour p premier, avec

$$f(d) = \mu^2(d) \prod_{p|d} f(p)$$

et l'inverse de f pour la convolution arithmétique vérifie $f^{*(-1)}(d) = \mu(d)f(d)$ pour d sans facteur carré.

Exemple 4.1. — Pour les applications en vue, on considérera essentiellement deux cas de suites (a_n) , supportées sur $N < n \leq 2N$ où $N \geq 1$ est un entier :

$$a_n = 1, \quad X = N, \quad g_t(d) = \frac{1}{d},$$

$$a_n = \Lambda(n + j), \quad X = N, \quad g_t(d) = \begin{cases} \frac{1}{\varphi(d)} & \text{si } (d, t + j) = 1, \\ 0 & \text{sinon} \end{cases}$$

(avec j fixé).

Remarque 4.2. — Dans ce qui suit, on peut considérer le k -uplet \mathbf{h} fixé pour le besoin du théorème 1.2 ; autrement dit, les « constantes implicites » peuvent alors dépendre de \mathbf{h} (et le théorème A.1 de l’appendice peut remplacer le théorème A.2). Pour le théorème 1.1, inconditionnel, les termes de reste doivent être *uniformes* par rapport à \mathbf{h} , pour k et ℓ fixés ; on supposera que \mathbf{h} vérifie

$$(16) \quad 0 \leq h_i \leq h$$

et ces estimations uniformes seront exprimées en fonction du paramètre h , que l’on supposera être ≥ 10 (pour que $\log \log h > 0$).

On commence par une transformation formelle de la forme quadratique.

LEMME 4.3. — *Avec les notations et hypothèses ci-dessus, on a*

$$Q = XH + R$$

où

$$H = \sum_{d,e < x} g_{\mathbf{h}}([d, e]) \lambda_d \lambda_e, \quad R = \sum_{d,e < x} \lambda_d \lambda_e R_{[d,e]}(\mathcal{A}, \mathbf{h}),$$

avec

$$g_{\mathbf{h}}(d) = \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu)=0}} g_{\nu}(d), \quad R_d(\mathcal{A}, \mathbf{h}) = \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu)=0}} r_d(\mathcal{A}, \nu)$$

pour tout d sans facteurs carrés.

Démonstration. — Il suffit de développer le carré et d’inverser l’ordre de sommation, ce qui fait apparaître la somme intérieure

$$\sum_{\substack{n \\ F_{\mathbf{h}}(n) \equiv 0}} a_n = \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu)=0}} \sum_{n \equiv \nu \pmod{[d,e]}} a_n,$$

et il ne reste plus qu’à appliquer (15) à la somme sur n . □

Puisque $g_{\nu}(d)$ ne dépend que de ν modulo d , le théorème chinois implique que $d \mapsto g_{\mathbf{h}}(d)$ est encore multiplicative. La forme quadratique H est typique dans le crible de Selberg. Le reste R , pour sa part, fait intervenir les quantités $r_d(\mathcal{A}, \nu)$ et pas seulement $r_d(\mathcal{A})$; pour le contrôler il est donc nécessaire de comprendre la répartition de \mathcal{A}

dans des progressions arithmétiques de module $[d, e] < x^2$ pour des classes de congruence non fixées, et c'est de là que vient l'insuffisance de l'exposant de répartition au sens faible.

Quoi qu'il en soit, suivant le synopsis du crible de Selberg, on procède à la diagonalisation de H .

LEMME 4.4. — *Pour toute fonction multiplicative $d \mapsto g(d)$, on a*

$$\sum_{d,e < x} g([d, e]) \lambda_d \lambda_e = \sum_r^b t(r) y_r^2$$

où on a, pour tout r sans facteurs carrés,

$$t(r) = \sum_{ab=r} g(a) \mu(b) g(b)^2 = g(r) \prod_{p|r} (1 - g(p)), \quad \text{et} \quad y_r = \sum_d g(d) \lambda_{rd}.$$

De plus, on a la formule d'inversion

$$(17) \quad \lambda_d = \sum_r \mu(r) g(r) y_{rd}.$$

Démonstration. — On pose $a = (d, e)$, donc $[d, e] = a \frac{d}{a} \frac{e}{a}$. Pour d, e sans facteurs carrés on a $(a, d/a) = (a, e/a) = 1$. On somme sur a d'abord, puis sur d et e divisibles par a et premiers entre eux. Cette condition $(d, e) = 1$ est détectée à l'aide de la fonction de Möbius

$$\sum_{b|(d,e)} \mu(b) = \begin{cases} 1 & \text{si } (d, e) = 1, \\ 0 & \text{sinon,} \end{cases}$$

et il vient

$$\begin{aligned} \sum_{d,e < x} g([d, e]) \lambda_d \lambda_e &= \sum_a g(a) \sum_{\substack{d,e < x/a \\ (d,e)=1}} g(d) g(e) \lambda_{ad} \lambda_{ae} \\ &= \sum_a g(a) \sum_b \mu(b) g(b)^2 \sum_{d,e < x/ab} g(d) g(e) \lambda_{abd} \lambda_{abe} \\ &= \sum_r^b t(r) y_r^2. \end{aligned}$$

Par inversion de Möbius on retrouve λ_d en fonction des variables y_r à l'aide de l'inverse de convolution j de g , à savoir

$$\lambda_d = \sum_r^b j(r) y_{rd} = \sum_r \mu(r) g(r) y_{rd}$$

puisque $j(r) = \mu(r) g(r)$ pour r sans facteurs carrés. □

Il est facile de minimiser H sous une contrainte linéaire. Ici, nous le faisons pour le cas $a_n = 1$ pour $N < n \leq 2N$. Dans ce cas, $X = N$, $g(d) = 1/d$ et $g_{\mathbf{h}}(d) = \nu_{\mathbf{h}}(d)/d$ où $\nu_{\mathbf{h}}(d)$ est le nombre de racines de $F_{\mathbf{h}}$ modulo d , ce qui coïncide avec la description de $\nu_{\mathbf{h}}(p)$

dans le théorème 2.7. On note $t_{\mathbf{h}}(r)$ la fonction apparaissant dans le changement de variable $(\lambda_d) \mapsto (y_r)$:

$$(18) \quad t_{\mathbf{h}}(r) = g_{\mathbf{h}}(r) \prod_{p|r} (1 - g_{\mathbf{h}}(p)) = \frac{\nu_{\mathbf{h}}(r)}{r} \prod_{p|r} \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p}\right).$$

La contrainte linéaire la plus simple est $\lambda_1 = 1$. On déduit alors de l'inégalité de Cauchy-Schwarz que si $\sum \mu(r)g_{\mathbf{h}}(r)y_r = \lambda_1 = 1$, on a

$$\begin{aligned} 1 &= \left(\sum_r \mu(r)g_{\mathbf{h}}(r)y_r\right)^2 \leq \left(\sum_r \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)}\right) \left(\sum_r t_{\mathbf{h}}(r)y_r^2\right) \\ &= \tilde{H}H(\lambda), \quad \text{avec } \tilde{H} = \sum_r \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)}, \end{aligned}$$

l'égalité étant obtenue, et donc H minimale, lorsque

$$y_r = \frac{1}{\tilde{H}} \frac{\mu(r)g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)}, \quad \text{pour } 1 \leq r < x.$$

Comme expliqué dans la précédente section, on modifie ce choix en insérant un facteur logarithme visant à « mimer » la $(k + \ell)$ -ième fonction de von Mangoldt. On pose

$$(19) \quad y_{r,\ell} = \frac{\mu(r)g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)} \left(\ln \frac{x}{r}\right)^\ell = \mu(r) \prod_{p|r} \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p}\right)^{-1} \left(\ln \frac{x}{r}\right)^\ell$$

pour $r < x$, $\ell \geq 1$; on a pu omettre le facteur $1/\tilde{H}$ pour simplifier, en raison de l'homogénéité des formules considérées. Les deux formes quadratiques Q_g et Q_p vont maintenant être évaluées asymptotiquement pour ce choix des variables y_r (et le choix correspondant de λ_d).

Commençons par le terme « principal » H_p correspondant à Q_p .

THÉORÈME 4.5. — *Soit $h \geq 3$ un entier, \mathbf{h} un k -uplet admissible vérifiant (16), $\ell \geq 0$ un entier. Soit*

$$H_p = \sum_{d,e < x} g_{\mathbf{h}}([d, e])\lambda_d\lambda_e = \sum_{d,e < x} \frac{\nu_{\mathbf{h}}([d, e])}{[d, e]}\lambda_d\lambda_e = \sum_r \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)}$$

la forme quadratique ci-dessus. Lorsque (λ_d) est donné par (19), on a

$$H_p(y_{r,\ell}) = \frac{(2\ell)!}{(k + 2\ell)! \mathfrak{S}(\mathbf{h})} \frac{1}{(\log x)^{k+2\ell}} \left(1 + O\left(\frac{\mathfrak{S}(\mathbf{h})(\ln \ln h)}{\ln x}\right)\right),$$

pour $x \geq 2$, la constante implicite ne dépendant que de k et ℓ .

Démonstration. — Il s'agit d'évaluer

$$(20) \quad H_p(y_{r,\ell}) = \sum_{r < x} t_{\mathbf{h}}(r)y_{r,\ell}^2 = \sum_{r < x} \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)} \left(\ln \frac{x}{r}\right)^{2\ell}.$$

Pour $\ell = 0$, il s'agit de la moyenne d'une fonction multiplicative sur $r < x$, qui est un sujet bien balisé, et le facteur lisse $(\ln x/r)^{2\ell}$ peut être traité par sommation par parties.

Plus précisément, on a pour r sans facteurs carrés

$$(21) \quad \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)} = \frac{\nu_{\mathbf{h}}(r)}{r} \prod_{p|r} \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p}\right)^{-1} = \prod_{p|r} \frac{\nu_{\mathbf{h}}(p)}{p - \nu_{\mathbf{h}}(p)}.$$

On a

$$\sum_{p < x} \frac{\nu_{\mathbf{h}}(p)}{p} \ln p = k \ln x + O(1)$$

et comme $\nu_{\mathbf{h}}(p) < p$, on en déduit

$$\sum_{p < x} \frac{\nu_{\mathbf{h}}(p)}{p - \nu_{\mathbf{h}}(p)} \ln p = k \ln x + O(1),$$

pour $x \geq 2$, la constante implicite dépendant de k . Cela permet d'appliquer le théorème A.2 de l'appendice ; après avoir vérifié que l'on peut prendre $L \ll \ln \ln h$ dans (31), on termine la preuve en vérifiant que la constante c de loc. cit. est bien égale à $1/\mathfrak{S}(\mathbf{h})$. \square

Passons à l'évaluation du terme principal de Q_g . Il s'agit donc de calculer

$$\sum_{i=1}^k \sum_{N < n \leq 2N} \Lambda(n + h_i) \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_d \right)^2,$$

et on le fait pour chaque i séparément. Pour la suite $a_n = \Lambda(n + h_i)$, i fixé, $N < n \leq 2N$, on a (15) avec $X = N$; dans le lemme 4.3, on a

$$\sum_{F_{\mathbf{h}}(\nu) \equiv 0 \pmod{p}} g_{\nu}(p) = \frac{\nu_{\mathbf{h}}(p) - 1}{p - 1},$$

pour p premier (voir exemple 4.1), et donc le rôle de $g_{\mathbf{h}}(d)$ est joué ici par

$$(22) \quad g_{\mathbf{h}}^{\sharp}(d) = \frac{1}{\varphi(d)} \prod_{p|d} (\nu_{\mathbf{h}}(p) - 1),$$

pour d sans facteurs carrés.

THÉORÈME 4.6. — *Avec les notations et hypothèses ci-dessus, et avec H_g la forme quadratique du lemme 4.3 pour la suite $a_n = \Lambda(n + h_i)$ pour $N < n \leq 2N$, on a*

$$H_g(y_{r,\ell}) = \frac{1}{(\ell + 1)^2} \frac{(2\ell + 2)!}{(k + 2\ell + 1)!} \frac{1}{\mathfrak{S}(\mathbf{h})} (\ln x)^{k+2\ell+1} \left(1 + O\left(\frac{\mathfrak{S}(\mathbf{h})(\ln \ln h)}{\ln x}\right) \right),$$

pour $x \geq 2$, où la constante implicite ne dépend que de k et ℓ .

Démonstration. — L'idée est d'appliquer le lemme 4.4 à H_g , et d'exprimer les nouvelles variables, disons z_r , qui y apparaissent, en fonction des variables y_r qui interviennent dans la diagonalisation de H_p . Cela permet d'abord d'évaluer z_r , puis ensuite H_g . Certaines simplifications viennent faciliter le travail, mais c'est essentiellement un exercice (tant que \mathbf{h} est fixé).

Tout d’abord, on a la diagonalisation

$$H_g = \sum_r^b t_{\mathbf{h}}^{\sharp}(r) z_r^2$$

avec

$$z_r = \sum_d g_{\mathbf{h}}^{\sharp}(d) \lambda_{rd} \quad \text{et} \quad t_{\mathbf{h}}^{\sharp}(r) = g_{\mathbf{h}}^{\sharp}(r) \prod_{p|r} (1 - g_{\mathbf{h}}^{\sharp}(p)).$$

D’après l’expression (17) qui fournit λ_d en fonction des y_r , on a donc

$$z_r = \sum_b \left(\sum_{de=b} g_{\mathbf{h}}^{\sharp}(d) \mu(e) g_{\mathbf{h}}(e) \right) y_{br}.$$

Pour $b = p$, la fonction multiplicative qui apparaît est égale à

$$g_{\mathbf{h}}^{\sharp}(p) - g_{\mathbf{h}}(p) = \frac{\nu_{\mathbf{h}}(p) - 1}{p - 1} - \frac{\nu_{\mathbf{h}}(p)}{p} = -\frac{p - \nu_{\mathbf{h}}(p)}{p(p - 1)}.$$

D’après (19) on trouve pour $(y_r) = (y_{r,\ell})$:

$$z_r = \sum_{b < x/r} \frac{\mu(b)}{\varphi(b)} \prod_{p|b} \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p} \right) \frac{\mu(br) g_{\mathbf{h}}(br)}{t_{\mathbf{h}}(br)} \left(\ln \frac{x}{br} \right)^{\ell}.$$

Les termes qui contribuent vérifient $(b, r) = 1$, ce qui permet de séparer le produit br . En utilisant (18) on trouve

$$z_r = \frac{\mu(r) g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)} \sum_{\substack{b < x/r \\ (r,b)=1}}^b \frac{1}{\varphi(b)} \left(\ln \frac{x}{br} \right)^{\ell}.$$

Une application très simple du théorème A.2 (avec $\kappa = 1$) donne

$$z_r = \frac{1}{\ell + 1} \frac{\mu(r) g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)} \frac{\varphi(r)}{r} \left(\ln \frac{x}{r} \right)^{\ell+1} + O\left((\ln \ln r) \left(\ln \frac{x}{r} \right)^{\ell} \right).$$

Le calcul de $H_g(y_{r,\ell})$ est de nouveau une application de ce même résultat. On a

$$t_{\mathbf{h}}^{\sharp}(r) z_r^2 = \frac{1}{(\ell + 1)^2} t_{\mathbf{h}}^{\sharp}(r) \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)^2} \left(\frac{\varphi(r)}{r} \right)^2 \left(\ln \frac{x}{r} \right)^{2\ell+2} + (\text{reste}).$$

Pour $r = p$ premier, il vient

$$\begin{aligned} t_{\mathbf{h}}^{\sharp}(r) \frac{g_{\mathbf{h}}(r)^2}{t_{\mathbf{h}}(r)^2} \left(\frac{\varphi(r)}{r} \right)^2 &= \frac{\nu_{\mathbf{h}}(p) - 1}{p - 1} \left(1 - \frac{\nu_{\mathbf{h}}(p) - 1}{p - 1} \right) \left(1 - \frac{\nu_{\mathbf{h}}(p)}{p} \right)^{-2} \left(\frac{p - 1}{p} \right)^2 \\ &= \frac{\nu_{\mathbf{h}}(p) - 1}{p - \nu_{\mathbf{h}}(p)}. \end{aligned}$$

Le théorème A.1 (avec $\kappa = k - 1$ maintenant) et de nouvelles estimations, longues mais élémentaires, de la contribution du reste donnent

$$H_g(y_{r,\ell}) = \frac{1}{(\ell + 1)^2} \frac{(2\ell + 2)!}{(k + 2\ell + 1)!} \frac{1}{\mathfrak{S}(\mathbf{h})} (\ln x)^{k+2\ell+1} \left(1 + O\left(\frac{\mathfrak{S}(\mathbf{h})(\ln \ln h)}{\ln x} \right) \right),$$

car la constante c de loc. cit. est (encore !)

$$(23) \quad c = \prod_p \left(1 - \frac{1}{p}\right)^{k-1} \left(1 + \frac{\nu_{\mathbf{h}}(p) - 1}{p - \nu_{\mathbf{h}}(p)}\right) = \prod_p \left(1 - \frac{1}{p}\right)^{k-1} \frac{p - 1}{p - \nu_{\mathbf{h}}(p)} = \frac{1}{\mathfrak{S}(\mathbf{h})}.$$

□

Pour être prêt à démontrer le théorème 1.2, il ne reste qu'à contrôler les termes de reste provenant du lemme 4.3. On suppose pour cela que \mathcal{A} s'obtient par «localisation dyadique» d'une suite infinie \mathbf{A} , c'est-à-dire que pour une suite $\mathbf{A} = (a_n)$ vérifiant (4), on pose $\mathcal{A} = (a_n)_{N < n \leq 2N}$ pour un certain $N \geq 1$. On écrit alors (15) avec les choix «évidents» des paramètres, provenant de (4), en particulier :

$$(24) \quad r_d(\mathcal{A}, t) = r_d(\mathcal{A}; 2N, t) - r_d(\mathcal{A}; N, t).$$

LEMME 4.7. — *Soit \mathbf{h} un k -uplet. Soit $\mathbf{A} = (a_n)$ une suite de réels positifs telle que \mathbf{A} ait un P -exposant de répartition $\geq \theta$, par exemple telle que \mathbf{A} ait un exposant de répartition au sens fort $\geq \theta$. Supposons de plus que (λ_d) vérifie*

$$(25) \quad |\lambda_d| \leq \tau(d)^{C_1} (\ln 2x)^{C_2}$$

pour tout $d < x$, et que

$$(26) \quad |r_d(\mathbf{A}; N, \nu)| \leq d^{-1} \tau(d)^{C_3} N (\ln 2N)^{C_4}$$

pour $N \geq 1$, d sans facteurs carrés et ν modulo d , C_1, \dots, C_4 étant des constantes ≥ 0 .

Alors pour $\mathcal{A} = (a_n)_{N < n \leq 2N}$, et avec $r_d(\mathcal{A}, t)$ défini comme ci-dessus, on a

$$R = \sum_{d, e < x} \lambda_d \lambda_e R_{[d, e]}(\mathcal{A}, \mathbf{h}) \ll N (\ln N)^{-A}$$

pour tout $x \geq 2$ tel que $x^2 < N^{\theta - \varepsilon}$ avec $\varepsilon > 0$, et tout $A \geq 1$. La constante implicite dépend de $(\mathbf{A}, \varepsilon, A, C_1, C_2, C_3, C_4)$.

Démonstration. — On a

$$|R| \leq \sum_{d < x^2}^b \tau_x(d) \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu) = 0}} |r_d(\mathcal{A}, \nu)|$$

où

$$(27) \quad \tau_x(d) = \sum_{[a, b] = d} |\lambda_a \lambda_b| \ll \tau(d)^{2+2C_1} (\ln 2x)^{2C_2}.$$

On applique alors l'inégalité de Cauchy-Schwarz pour obtenir

$$|R|^2 \leq \left(\sum_{d < x^2}^b \tau_x(d)^2 \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu) = 0}} |r_d(\mathcal{A}, \nu)| \right) \left(\sum_{d < x^2}^b \sum_{\substack{\nu \pmod{d} \\ F_{\mathbf{h}}(\nu) = 0}} |r_d(\mathcal{A}, \nu)| \right).$$

On borne $|r_d(\mathcal{A}, \nu)|$ à l'aide de (24) ; pour le premier facteur, on applique (26) et (27) ; la somme avec des fonctions diviseurs et des logarithmes fait «perdre» un facteur de type $(\ln N)^{C_5}$ avec C_5 une constante assez grande. L'hypothèse $x^2 < N^{\theta - \varepsilon}$ permet

d'appliquer au second facteur la borne (5) définissant le P -exposant de répartition, et cette puissance de logarithme égarée est «récupérée» avec toute la marge nécessaire. \square

Noter que (26) n'est pas une condition onéreuse ; par exemple si $g_\nu(p) \leq kp^{-1}$ et a_n est bornée par une puissance de logarithme, elle sera aussitôt vérifiée. De même pour l'estimation individuelle (25). Pour les λ_d discutés précédemment, on a

LEMME 4.8. — Soit \mathbf{h} un k -uplet admissible, $\ell \geq 0$ et (λ_d) donné par (19) et (17) pour $d < x$. On a alors

$$\lambda_d \ll \frac{1}{\mathfrak{S}(\mathbf{h})} (\ln x)^{k+\ell}$$

pour tout $x \geq 2$ et $d \geq 2$ sans facteurs carrés, la constante implicite ne dépendant que de k et ℓ .

Voir [13, Lemma 9] ; pour \mathbf{h} fixé, c'est à peu près évident.

Les lecteurs sont encouragés à passer à la section suivante où les résultats ci-dessus sont appliqués à la preuve du théorème 1.2. Mais pour celle (inconditionnelle) du théorème 1.1, un ingrédient supplémentaire similaire à ce qui précède est requis ; il s'agit de considérer la forme quadratique (14) avec $a_n = \Lambda(n + j)$ pour $N < n \leq 2N$ et $j \geq 0$, mais cette fois sous l'hypothèse que j n'est pas une composante de \mathbf{h} . On note Q_i cette somme. Comme il s'agit de la même suite a_n que pour Q_g on a la même forme de (15) que dans ce cas. On écrit donc

$$Q_i = NH_i + R_i$$

selon le lemme 4.3. Si l'on note \mathbf{j} le $(k + 1)$ -uplet (\mathbf{h}, j) , on s'aperçoit alors que H_i est donnée par

$$H_i = \sum_{d, e < x} g_{\mathbf{j}}^\sharp([d, e]) \lambda_d \lambda_e,$$

où la fonction $d \mapsto g_{\mathbf{j}}^\sharp(d)$ est définie par (22), pour \mathbf{j} au lieu de \mathbf{h} .

THÉORÈME 4.9. — Soient $h \geq 3$, \mathbf{h} un k -uplet admissible, $j \geq 0$ un entier qui n'est pas une composante de h tel que $\mathbf{j} = (\mathbf{h}, j)$ soit admissible et vérifie (16). Soient $\ell \geq 1$ un entier et (λ_d) donné par (19), appliqué à \mathbf{h} toujours. On a alors

$$H_i(y_{r,\ell}) = \frac{(2\ell)!}{(k + 2\ell)!} \frac{\mathfrak{S}(\mathbf{j})}{\mathfrak{S}(\mathbf{h})^2} (\ln x)^{k+2\ell} \left(1 + O\left(\frac{\mathfrak{S}(\mathbf{h})(\ln \ln h)}{\ln x} \right) \right),$$

pour $x \geq 2$, la constante implicite ne dépendant que de k et ℓ .

Démonstration. — Formellement, cela ressemble beaucoup à ce qui précède. Le lemme 4.4 permet de diagonaliser

$$H_i = \sum_r^b t_{\mathbf{j}}^\sharp(r) w_r^2$$

avec

$$w_r = \sum_d g_{\mathbf{j}}^\sharp(d) \lambda_{rd} = \sum_b \left(\sum_{de=b} g_{\mathbf{j}}^\sharp(d) \mu(e) g_{\mathbf{h}}(d) \right) y_{br}.$$

En insérant la valeur de $y_{r,\ell}$, on trouve

$$w_r = \frac{\mu(r)g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)} \sum_{\substack{b < x/r \\ (r,b)=1}} \frac{\mu(b)g_{\mathbf{h}}(b)}{t_{\mathbf{h}}(b)} \left(\sum_{de=b} g_{\mathbf{j}}^{\#}(d)\mu(e)g_{\mathbf{h}}(d) \right) \left(\ln \frac{x}{br} \right)^{\ell}.$$

Le point où le calcul diffère du précédent est que la série

$$\sum_{\substack{b \geq 1 \\ (r,b)=1}} \frac{\mu(b)g_{\mathbf{h}}(b)}{t_{\mathbf{h}}(b)} \left(\sum_{de=b} g_{\mathbf{j}}^{\#}(d)\mu(e)g_{\mathbf{h}}(d) \right)$$

converge ; en effet, en calculant formellement (puis en justifiant sans peine), elle est égale à

$$\prod_{p \nmid r} \left(1 - \frac{1}{1 - g_{\mathbf{h}}(p)} (g_{\mathbf{j}}^{\#}(p) - g_{\mathbf{h}}(p)) \right) = \prod_{p \nmid r} \frac{1 - g_{\mathbf{j}}^{\#}(p)}{1 - g_{\mathbf{h}}(p)},$$

dont la convergence provient du fait que $g_{\mathbf{j}}^{\#}(p)$ et $g_{\mathbf{h}}(p)$ sont toutes deux de l'ordre de k/p en moyenne.

Il n'est pas très difficile de déduire de cela que

$$\begin{aligned} w_r &= \frac{\mu(r)g_{\mathbf{h}}(r)}{t_{\mathbf{h}}(r)} \prod_{p \nmid r} \frac{1 - g_{\mathbf{j}}^{\#}(p)}{1 - g_{\mathbf{h}}(p)} \left(\ln \frac{x}{r} \right)^{\ell} + (\text{reste}), \\ &= \mu(r) \prod_{p \mid r} \frac{1}{1 - g_{\mathbf{h}}(p)} \prod_{p \nmid r} \frac{1 - g_{\mathbf{j}}^{\#}(p)}{1 - g_{\mathbf{h}}(p)} \left(\ln \frac{x}{r} \right)^{\ell} + (\text{reste}), \\ &= \mu(r) \prod_{p \mid r} (1 - g_{\mathbf{j}}^{\#}(p))^{-1} \frac{\mathfrak{S}(\mathbf{j})}{\mathfrak{S}(\mathbf{h})} \left(\ln \frac{x}{r} \right)^{\ell} + (\text{reste}) \\ &= \frac{\mu(r)g_{\mathbf{j}}^{\#}(r)}{t_{\mathbf{j}}^{\#}(r)} \frac{\mathfrak{S}(\mathbf{j})}{\mathfrak{S}(\mathbf{h})} \left(\ln \frac{x}{r} \right)^{\ell} + (\text{reste}). \end{aligned}$$

Finalement, on a

$$\sum_r^b t_{\mathbf{j}}^{\#}(r)w_r^2 = \frac{\mathfrak{S}(\mathbf{j})^2}{\mathfrak{S}(\mathbf{h})^2} \sum_{r < x}^b \frac{g_{\mathbf{j}}^{\#}(r)^2}{t_{\mathbf{j}}^{\#}(r)} \left(\ln \frac{x}{r} \right)^{2\ell} + (\text{reste}).$$

Comme dans le théorème 4.5 (voir (20) et (21)), on a d'après le théorème A.2

$$\sum_r^b t_{\mathbf{j}}^{\#}(r)w_r^2 = \frac{\mathfrak{S}(\mathbf{j})^2}{\mathfrak{S}(\mathbf{h})^2} \frac{1}{\mathfrak{S}(\mathbf{j})} \frac{(2\ell)!}{(k + 2\ell)!} \left(\ln \frac{x}{r} \right)^{2\ell+1} + (\text{reste})$$

(pour le calcul de la constante c , voir (23) appliqué à \mathbf{j}), ce qui donne le résultat, au terme de reste près dont il faut s'assurer qu'il est tel que décrit (c'est relativement long mais pas du tout difficile). \square

5. PREUVE DU THÉORÈME 1.2

Pour démontrer le théorème 1.2, il ne reste qu'à combiner les informations obtenues dans la section précédente. Fixons un k -uplet admissible \mathbf{h} , avec $k \geq 2$, et supposons que le P -exposant de répartition de la suite des nombres premiers (donc de ses translatées) soit $\geq \theta$. Prenons de plus $\ell \geq 1$.

Pour $x = N^\beta$ avec $2\beta < \theta - \varepsilon$, la combinaison des théorèmes 4.5 et 4.6 (appliqué pour chaque $\Lambda(n + h_i)$, $1 \leq i \leq k$) et du lemme 4.7 permet d'affirmer qu'il existe (λ_d) tels que les formes quadratiques Q_g et Q_p définies par (12) et (13) vérifient

$$Q_p = \frac{(2\ell)!}{(k + 2\ell)!} \frac{\beta^{k+2\ell}}{\mathfrak{S}(\mathbf{h})} N(\ln N)^{k+2\ell+1} \left(1 + O\left(\frac{1}{\ln N}\right)\right),$$

$$Q_g = \frac{k}{(\ell + 1)^2} \frac{(2\ell + 2)!}{(k + 2\ell + 1)!} \frac{\beta^{k+2\ell+1}}{\mathfrak{S}(\mathbf{h})} N(\ln N)^{k+2\ell+1} \left(1 + O\left(\frac{1}{\ln N}\right)\right)$$

pour tout $N \geq 2$, les constantes implicites dépendant ici de k et \mathbf{h} . Donc

$$(28) \quad Q_g - Q_p \geq \frac{1}{2} \left\{ \frac{2k\beta(2\ell + 1)}{(\ell + 1)(k + 2\ell + 1)} - 1 \right\} \frac{\beta^{k+2\ell}}{\mathfrak{S}(\mathbf{h})} \frac{(2\ell)!}{(k + 2\ell)!} N(\ln N)^{k+2\ell+1},$$

pour N assez grand.

Déjà lorsque $k = 7$, $\ell = 1$, le terme de droite est > 0 si $2\beta > 20/21$, ce qui démontre que la conjecture d'Elliott-Halberstam implique que pour tout 7-uplet admissible \mathbf{h} , il existe une infinité de n tels que $n \triangleleft \mathbf{h}$. Puisque $\mathbf{h} = (11, 13, 17, 19, 23, 29, 31)$ est admissible, on a⁽²¹⁾ $\liminf \gamma(n) \leq 20$. Pour obtenir la première partie du théorème 1.2 (où 16 remplace 20), une astuce supplémentaire est requise (on crée une combinaison des cas $\ell = 0$ et $\ell = 1$) ; voir [13, p. 44]. Noter aussi que $\ell = 0$ est toujours insuffisant.

De plus la fraction rationnelle entre accolades dans (28) tend vers $2\theta - 1$ si $k, \ell \rightarrow +\infty$ avec $\ell = o(k)$ (par exemple, $\ell = \lfloor \sqrt{k} \rfloor$). Si $\theta > 1/2$, on peut choisir k et ℓ tels que $Q_g > Q_p$ pour tout N assez grand, ce qui démontre la première partie du théorème 1.2⁽²²⁾.

6. PREUVE DU THÉORÈME 1.1

Inconditionnellement, on peut appliquer (28) pour tout $\beta < 1/4$, mais on échoue alors tout juste. Il faut une autre idée, pas du tout évidente non plus, pour passer la difficulté. Il s'agit de donner un petit coup de pouce à Q_g en rajoutant les sommes du type

$$Q_i = \sum_{N < n \leq 2N} \Lambda(n + j) \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_d \right)^2$$

⁽²¹⁾ On doit aussi remarquer que la contribution à $Q_g - Q_p$ des puissances de nombres premiers également détectées par $\vartheta_{\mathbf{h}}(n)$ est $\ll N^{1/2+\varepsilon}$ pour tout $\varepsilon > 0$.

⁽²²⁾ Même remarque que ci-dessus.

considérées à la fin de la section 4, où cette fois j n'apparaît pas dans \mathbf{h} . Il n'y a aucune corrélation alors à espérer entre le fait que $n + j$ soit premier et que $F_{\mathbf{h}}(n)$ soit presque premier ; cela signifie, comme on l'a constaté, que le terme principal dans Q_i est plus petit par un facteur logarithmique de celui dans Q_g . Mais toute somme de type Q_i apporte une contribution « infinitésimale » telle que la combinaison d'un grand nombre d'entre elles finit par faire déborder le vase...

Voici le fonctionnement précis de cette idée. On considère $h \geq 1$ et on pose

$$\vartheta'(n, h) = \sum_{1 \leq j \leq h} \Lambda(n + j),$$

puis on définit les quantités

$$Q'_g = \frac{1}{h^k} \sum_{\substack{\mathbf{h}=(h_1, \dots, h_k) \\ h_i \leq h}}^* \sum_{N < n \leq 2n} \vartheta'(n, h) \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_{d, \mathbf{h}} \right)^2,$$

$$Q'_p = \frac{(\ln 3N)}{h^k} \sum_{\substack{\mathbf{h}=(h_1, \dots, h_k) \\ h_i \leq h}}^* \sum_{N < n \leq 2N} \left(\sum_{d|F_{\mathbf{h}}(n)} \lambda_{d, \mathbf{h}} \right)^2,$$

le symbole \sum^* signifiant que la somme est restreinte aux k -uplets admissibles. Noter que l'on peut s'autoriser à choisir des coefficients dépendant de \mathbf{h} , ce qui est important puisque ceux définis dans la section précédente en dépendent effectivement. Précisément, pour tout \mathbf{h} on va prendre

$$(29) \quad \lambda_{d, \mathbf{h}} = \mathfrak{S}(\mathbf{h}) \lambda_d$$

où λ_d est donné par (17) et (19) ; l'insertion du facteur $\mathfrak{S}(\mathbf{h})$ permet par homogénéité de simplifier la moyenne sur \mathbf{h} en la ramenant à celle considérée par Gallagher (Proposition 2.9).

On note encore que si $Q'_g > Q'_p$, il existe n tel que $\vartheta'(n, h) > \ln 2N$, et alors (si $h \leq N$) il existe deux (puissances de) nombres premiers $p < p'$, $N < p \leq 2N$, tels que $p' - p \leq h$.

THÉORÈME 6.1. — Soient $h \geq 1$, $k \geq 1$ et $\ell \geq 1$ des entiers. Pour chaque \mathbf{h} , soit $\lambda_{d, \mathbf{h}}$ défini par (29). On a alors pour tout $\varepsilon > 0$

$$Q'_p = \frac{(2\ell)!}{(k + 2\ell)!} N (\ln x)^{k+2\ell} (\ln 3N) \mathfrak{S}(h, k) + O(N (\ln x)^{k+2\ell-1+\varepsilon})$$

et

$$Q'_g = \frac{(2\ell)!}{(k + 2\ell)!} N (\ln x)^{k+2\ell+1} \left\{ \frac{2k \mathfrak{S}(h, k)}{(\ell + 1)(k + 2\ell + 1)} + \frac{h \mathfrak{S}(h, k + 1)}{\ln x} \right\} + O(N (\ln x)^{k+2\ell+\varepsilon})$$

où $\mathfrak{S}(h, k)$ est défini par (8), pour tout $N \geq 2$, $x \geq 2$ tel que $h \leq \ln x \leq \ln N$, la constante implicite dépendant de k , ℓ et ε seulement.

Démonstration. — Le calcul de chaque terme dans Q'_p provient directement du théorème 4.5 ; la somme sur \mathbf{h} fait apparaître exactement le terme $S(h, k)$ d'après sa définition.

Si l'on développe $\vartheta'(n, h)$ dans Q'_g , la contribution de chacun des k -uplets \mathbf{h} dans lesquels j apparaît est donnée par le théorème 4.6, et il apparaît le coefficient

$$\sum_{j=1}^k \sum_{\substack{\mathbf{h}=(h_1, \dots, h_k) \\ 1 \leq h_i \leq h \\ h_i \neq j}}^* \mathfrak{S}(\mathbf{h}) = kS(h, k)$$

en inversant l'ordre de sommation.

La dernière contribution est celle des k -uplets \mathbf{h} et des j tels que j n'apparaît pas dans \mathbf{h} . Chaque terme relève du théorème 4.9, et il apparaît le facteur

$$\sum_{1 \leq j \leq k} \sum_{\substack{\mathbf{h}=(h_1, \dots, h_k) \\ j \neq h_i \leq h}}^* \mathfrak{S}(\mathbf{h}, j) = \sum_{\substack{j=(h_1, \dots, h_k, h_{k+1}) \\ h_i \leq h}}^* \mathfrak{S}(j) = S(h, k+1).$$

□

On conclut alors de la manière suivante : en faisant la différence de $Q'_g - Q'_p$ avec $x = N^\beta$ le terme principal est maintenant

$$\Xi = \frac{(2\ell)!}{(k+2\ell)!} \beta^{k+2\ell} N (\ln N)^{k+2\ell+1},$$

où

$$\Xi = \frac{2k\beta(2\ell+1)S(h, k)}{(\ell+1)(k+2\ell+1)} + \frac{hS(h, k+1)}{\ln N} - 1.$$

Puisque pour k fixé on a $S(h, k) \rightarrow 1$ et $S(h, k+1) \rightarrow 1$, d'après le résultat de Gallagher (Proposition 2.9), on a approximativement⁽²³⁾

$$\Xi \approx \Xi' = \frac{2k\beta(2\ell+1)}{(\ell+1)(k+2\ell+1)} + \frac{h}{\ln N} - 1,$$

et lorsque h est grand, cela montre qu'on a gagné un facteur $h/\ln N$ par rapport au cas précédent (voir (28)). Puisque ce dernier argument échouait tout juste, il suffit que $h/\ln N$ soit strictement positif pour tout N assez grand (mais éventuellement arbitrairement petit) pour pouvoir conclure.

Précisément, fixons $\delta > 0$ arbitrairement petit et posons $h = \delta \ln N$, pour $N \geq \exp(\delta^{-1})$. La limite de Ξ' quand $k, \ell \rightarrow +\infty$ avec $\ell^2 \leq k$ est alors égale à $4\beta - 1 + \delta$. On peut trouver $\beta < 1/4$ tel que $4\beta + \delta > 1$ et, cela fait, il existe k et ℓ (dépendant seulement de δ et β) tels que $\Xi' > 0$ pour tout N suffisamment grand (N est maintenant la seule variable). On en déduit par la proposition 2.9 que $\Xi > 0$ également pour tout N assez grand puisque $h \rightarrow +\infty$ lorsque $N \rightarrow +\infty$.

⁽²³⁾Le symbole \approx est heuristique seulement.

Faisant finalement appel au théorème de Bombieri-Vinogradov pour estimer comme précédemment les termes de reste, on peut donc conclure que, pour tout N assez grand, on a

$$Q'_g - Q'_p \geq \frac{1}{2} \Xi \frac{(2\ell)!}{(k+2\ell)!} \beta^{k+2\ell} N (\ln N)^{k+2\ell+1} > 0.$$

Par conséquent, il existe une infinité d'entiers tels que $\gamma(n) \leq h \leq \delta \ln N$, ce qui signifie que

$$\liminf_{n \rightarrow +\infty} \frac{\gamma(n)}{\ln n} \leq \delta.$$

Finalement, le théorème 1.1 est obtenu puisque δ est arbitrairement petit.

Remarque 6.2. — Il est à remarquer que dans cette preuve l'uniformité des estimations de Q'_p et Q'_g par rapport au k -uplet \mathbf{h} est absolument essentielle.

7. AUTRES RÉSULTATS

La méthode utilisée par Goldston, Pintz et Yıldırım est robuste et a un fort potentiel d'applications au-delà de la preuve des théorèmes 1.1 et 1.2. Nous allons ici simplement mentionner quelques-uns des résultats annoncés, référant aux prépublications disponibles et à venir pour plus de détails.

En premier lieu, dans [13], Goldston, Graham, Pintz et Yıldırım démontrent que l'exposant de répartition $1/2$ n'est pas nécessairement une barrière intrinsèque : concernant les écarts entre entiers $n = p_1 p_2$ avec p_i des nombres premiers distincts⁽²⁴⁾, ils démontrent :

THÉORÈME 7.1. — *Soient*

$$e_1 = 6 < e_2 = 10 < e_3 = 15 < \cdots < e_{278} = 7 \cdot 137 < \cdots < e_n < \cdots$$

la suite des entiers qui sont produits de deux facteurs premiers distincts. Alors on a

$$\liminf_{n \rightarrow +\infty} (e_{n+1} - e_n) \leq 26 < +\infty.$$

Pour mettre ce résultat en contexte, on peut remarquer que la suite (e_n) est plus dense que celle des nombres premiers : on a

$$e_n \sim \frac{n \ln n}{\ln \ln n} \text{ quand } n \rightarrow +\infty,$$

⁽²⁴⁾ À ne pas confondre avec les nombres « P_2 » du crible classique, qui incluent les nombres premiers et leurs carrés.

ce qui suggère que l'espacement entre les e_n est généralement moindre qu'entre nombres premiers⁽²⁵⁾. Pour ce théorème, l'analogie du théorème de Bombieri-Vinogradov pour la répartition des e_n dans les progressions arithmétiques est suffisant.

Le résultat suivant est annoncé dans [13] et [11] :

THÉORÈME 7.2. — *On a*

$$\liminf_{n \rightarrow +\infty} \frac{\gamma(n)}{\sqrt{\ln n} (\ln \ln n)^2} < +\infty.$$

Au vu de la preuve du théorème 1.1, il est naturel d'essayer d'obtenir un tel raffinement, car tous les paramètres disponibles n'ont pas été employés optimalement ; autrement dit, il s'agit de préciser toutes les estimations pour les rendre *uniformes par rapport à k* . Il s'agit d'une tâche extrêmement délicate, et nous n'en dirons pas plus.

Finalement, il est naturel de s'enquérir des espacements $\gamma_r(n) = p_{n+r} - p_n$ pour tout $r \geq 1$. La méthode, avec des adaptations, fournit des résultats, mais pas l'analogie du théorème 1.1. Goldston, Pintz et Yıldırım [11] démontrent :

THÉORÈME 7.3. — *Soit $\theta \geq 0$ tel que la suite $(\Lambda(n))$ ait un exposant de répartition $\geq \theta$; alors on a*

$$\liminf_{n \rightarrow +\infty} \frac{\gamma_r(n)}{\ln n} \leq (\sqrt{r} - \sqrt{2\theta})^2,$$

pour $r \geq 1$, et en particulier

$$\liminf_{n \rightarrow +\infty} \frac{\gamma_r(n)}{\ln n} \leq (\sqrt{r} - 1)^2.$$

Pour $r \geq 11$, ce n'est cependant pas le meilleur résultat connu, qui est

$$\liminf_{n \rightarrow +\infty} \frac{\gamma_r(n)}{\ln n} \leq e^{-\gamma} \left(r - \frac{\sqrt{r}}{2} \right)$$

où γ est la constante d'Euler. Ce résultat de J. Sivak [23] est basé sur la combinaison de la méthode de Goldston-Yıldırım et de celle de Maier qui fournit des intervalles (clairsemés) d'entiers contenant plus de nombres premiers qu'attendu. Goldston, Pintz et Yıldırım [11] ont annoncé pouvoir incorporer les idées de Maier à leurs arguments, ce qui devrait mener à

$$\liminf_{n \rightarrow +\infty} \frac{\gamma_r(n)}{\ln n} \leq e^{-\gamma} (\sqrt{r} - 1)^2.$$

⁽²⁵⁾Toujours afin de préciser le contexte, rappelons que la plupart des nombres P_2 obtenus par les méthodes de crible vérifient une propriété additionnelle plus forte : ils n'ont « pas de petits facteurs premiers », ce qui ramène leur densité moyenne à celle des nombres premiers.

8. QUEL EST L'EXPOSANT DE RÉPARTITION DES NOMBRES PREMIERS ?

Dès la preuve du théorème de Bombieri-Vinogradov et de ses généralisations à d'autres suites, l'importance potentielle d'une amélioration de l'exposant de répartition d'une suite au-delà de la limite du grand crible avait été reconnue. Depuis environ 1980 (le premier exemple est l'article [8] de Fouvry et Iwaniec, puis la thèse d'É. Fouvry est une étape importante), la recherche de telles améliorations pour diverses suites naturelles a fait l'objet d'un effort concerté de grande ampleur de la part essentiellement de Fouvry, Iwaniec, Friedlander et Bombieri.

Au vu du théorème 1.2, la question se pose de nouveau de manière particulièrement brûlante, et nous allons faire ici rapidement un survol de ce qui est connu concernant ce problème.

La première remarque essentielle est de bien distinguer entre exposant de répartition *au sens faible* et *au sens fort* (ce n'est pas toujours le cas dans la littérature). En effet, rappelons que pour un problème de crible classique, c'est l'exposant de répartition au sens faible qui intervient naturellement, alors que pour la méthode de Goldston, Pintz et Yıldırım, il ne peut suffire.

Dans le cas important des nombres premiers, bien que l'exposant au sens fort soit celui le plus discuté dans la littérature, la notion de P -exposant de répartition mérite peut-être l'attention car, en limitant le problème d'uniformité à un ensemble de classes modulo q assez spécial – et, en particulier, de taille raisonnable puisque de l'ordre de $(\ln q)^{k-1}$ en moyenne –, il ne paraît pas impossible qu'il soit plus accessible.

Voici déjà quelques bonnes raisons de croire que l'exposant de répartition des nombres premiers puisse être $> 1/2$:

– D'une part, l'exposant $1/2$ découle, comme on l'a dit, de l'Hypothèse de Riemann Généralisée pour les fonctions L de Dirichlet *de manière triviale* : on part d'une « formule explicite » (voir par exemple [18, 5.66]) telle que

$$\psi(x; q, a) = \frac{x}{\varphi(q)} + \frac{\sqrt{x}}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \sum_{|\gamma_\chi| \leq T} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + O\left(\frac{x(\ln qx)^2}{T}\right),$$

pour $1 \leq T \leq x \leq X$, $q \geq 1$, la constante implicite étant absolue ($\frac{1}{2} + i\gamma_\chi$ parcourt les zéros non triviaux de $L(\chi, s)$), et on estime « trivialement » la somme sur les zéros avec $T = x$. Il ne semble pas déraisonnable de penser que la somme sur χ donne lieu à des compensations significatives permettant d'obtenir un exposant de répartition des nombres premiers $> 1/2$ ⁽²⁶⁾.

– À l'heure actuelle, le premier point relève du « wishful thinking ». Il semble beaucoup plus convaincant à l'auteur qu'il soit *connu* (essentiellement) que l'exposant de répartition *au sens faible* de la suite des nombres premiers est $> 1/2$. Précisément, on a le théorème suivant :

⁽²⁶⁾Dans le monde des fonctions L sur les corps finis, des résultats (relativement) similaires sont connus.

THÉORÈME 8.1. — Il existe $\theta_0 > 1/2$ tel que pour tout $X \geq 2$, tout $x \leq X$, tout $A > 0$ et tout $B > 0$ on ait

$$\sum_{\substack{d \leq D \\ (a,d)=1}} \gamma_d \left\{ \psi(x; d, a) - \frac{x}{\varphi(d)} \right\} \ll \frac{X}{(\ln X)^A}$$

uniformément pour tout $D \leq X^{\theta_0 - \varepsilon}$ avec $\varepsilon > 0$, tout a tel que $1 \leq |a| \leq (\ln X)^B$, et toute fonction arithmétique bien factorisable γ_d de niveau D et d'ordre fixé. La constante implicite dépend au plus de A , B et ε .

Avec $\theta_0 = 9/17$, cela est dû à Fouvry et Iwaniec, et le meilleur résultat connu, $\theta_0 = 4/7$, est dû à Bombieri, Friedlander et Iwaniec [3]. Dans les deux cas, un ingrédient crucial est la théorie spectrale des formes automorphes et ses applications à l'estimation des sommes de sommes de Kloosterman, développées par Deshouillers et Iwaniec⁽²⁷⁾.

La notion de fonction bien factorisable⁽²⁸⁾ est due à Iwaniec ; bien que l'énoncé ne permette pas *a priori* de prendre pour γ_d le signe de $\psi(x; d, a) - \frac{x}{\varphi(d)}$, Iwaniec a démontré que ces fonctions permettent d'exprimer le terme d'erreur du crible « linéaire ». Il en résulte que cet énoncé est – du point de vue des applications usuelles – équivalent à dire que l'exposant de répartition au sens faible (pour a fixé) est $\geq \theta_0 > 1/2$.

Voici une liste (sans doute incomplète) de fonctions dont l'exposant de répartition θ au sens faible est $> 1/2$. Un survol du crible qui met en valeur cet aspect et ses applications a été écrit par Fouvry [7], auquel nous renvoyons pour les références.

- Pour tout $\varepsilon > 0$ fixé, la fonction caractéristique des $n \geq 1$ tels que $p \mid n$ implique $p > n^{1/6 - \varepsilon}$ (Fouvry).
- La multiplicité de représentation de n sous la forme $n = a^2 + b^2$, a appartenant à n'importe quelle suite d'entiers « assez dense » (Fouvry et Iwaniec) ou $n = a^2 + b^4$ (Friedlander et Iwaniec) ; ici l'exposant de répartition au sens faible est optimal.
- La multiplicité de représentation de n sous la forme $n = a^3 + 2b^3$ (Heath-Brown).

Concernant l'exposant de répartition au sens fort, on a :

- La fonction diviseur $\tau(n)$, avec $\theta \geq 2/3$ (Selberg, Linnik, Hooley)⁽²⁹⁾.
- La fonction $\tau_3(n)$, nombre d'écritures $n = abc$ avec $a, b, c \geq 1$ (Friedlander–Iwaniec)⁽³⁰⁾.
- Finalement, le seul (?) exemple connu de « fonction caractéristique » dont l'exposant de répartition *au sens fort* soit $> 1/2$ est la suite (a_n) , fonction caractéristique des

⁽²⁷⁾L'uniformité pour $|a| \leq (\ln X)^B$ n'est pas énoncée explicitement. Il est aussi possible de remplacer la classe fixée a par la classe racine d'une équation linéaire donnée $ma + n = 0 \pmod{d}$, où $m \geq 1$ et n est un entier tel que $(m, n) = 1$; cela peut se voir comme un premier pas vers un P -exposant de répartition. Je dois ces remarques à É. Fouvry.

⁽²⁸⁾Une fonction arithmétique f est bien factorisable de niveau D entier et d'ordre k si $f(d) = 0$ pour $d > D$, $0 \leq f(d) \leq \tau(d)^k$, et si pour toute factorisation $D = D_1 D_2$, il existe f_1, f_2 d'ordre k et de niveau D_1 et D_2 respectivement telles que $f = f_1 \star f_2$.

⁽²⁹⁾Les termes de reste sont même majorés individuellement ici.

⁽³⁰⁾Même remarque.

entiers dont la somme des chiffres en base 2 est pair, et ses variantes évidentes en d'autres bases (Fouvry–Mauduit).

9. QUESTIONS

Comme tout progrès significatif, les idées de Goldston, Pintz et Yıldırım soulèvent un certain nombre de nouvelles questions ; nous terminons en indiquant les plus évidentes. – Peut-on prouver le théorème 1.2 inconditionnellement ? Ce serait un progrès spectaculaire, et cela semble moins imaginable qu'il y a quelques mois. Une raison d'espérer est que l'essentiel de la méthode décrite jusqu'à présent se situe au niveau des « termes principaux », et ne procède à aucun travail nouveau au niveau des termes de reste dans (4). Cela semble laisser un grand potentiel.

– Conditionnellement, quelle est la limite de la méthode ? Peut-on démontrer la conjecture des nombres premiers jumeaux (sous la forme (1)) en supposant valide la conjecture d'Elliott-Halberstam ?

– La méthode peut-elle s'étendre aux écarts entre nombres premiers vérifiant des propriétés supplémentaires ? Comme cela n'a de sens que si l'on sait qu'il existe une infinité de tels nombres premiers, le cas le plus naturel est celui des p vérifiant $p \equiv a \pmod{q}$ (avec q et a fixés) ; cela peut se traiter (c'est annoncé dans [12]) en considérant $G_{\mathbf{h}}(n) = (qn + a + h_1) \cdots (qn + a + h_k)$ au lieu de $F_{\mathbf{h}}(n)$. En particulier, il ne semble pas évident *a priori* de considérer les écarts entre nombres premiers vérifiant une condition galoisienne (par exemple, p totalement scindé dans une extension galoisienne K/\mathbb{Q} fixée), ce qui serait pourtant tout à fait intéressant.

APPENDICE A : SOMMES DE FONCTIONS MULTIPLICATIVES

Voici un des théorèmes standard permettant d'évaluer asymptotiquement une somme de fonction multiplicative.

THÉORÈME A.1. — Soient g une fonction multiplicative positive, $\kappa \geq 0$ un entier tel que

$$\sum_{p < x} g(p) \ln p = \kappa \ln x + O(1).$$

Alors pour tout entier $\ell \geq 0$ et tout $x \geq 2$ on a

$$\sum_{d < x} g(d) \left(\ln \frac{x}{d} \right)^\ell = c \frac{\ell!}{(\ell + \kappa)!} (\ln x)^{\kappa + \ell} + O((\ln x)^{\ell + \kappa - 1})$$

avec

$$c = \prod_p \left(1 - \frac{1}{p} \right)^\kappa (1 + g(p)),$$

la constante implicite dépendant seulement de g et ℓ .

Expliquons en quelques mots d'où vient la constante c , et pourquoi elle est naturelle. On peut essayer de démontrer ce résultat en procédant par intégration complexe, en écrivant

$$\sum_{d < x}^b g(d) \left(\ln \frac{x}{d} \right)^\ell = \frac{\ell!}{2i\pi} \int_{(1)} D_g^b(s) x^s \frac{ds}{s^{\ell+1}}$$

où l'intégrale complexe est prise sur la droite verticale $\operatorname{Re}(s) = 1$ et

$$D_g^b(s) = \sum_{n \geq 1}^b g(n) n^{-s} = \prod_p (1 + g(p) p^{-s}).$$

L'hypothèse peut s'interpréter en disant que $g(p)$ est égal à κ/p «en moyenne», ce qui suggère de comparer $D_g^b(s)$ à $\zeta(s+1)^\kappa$, et de définir une fonction $E_g(s)$ en posant

$$D_g^b(s) = \zeta(s+1)^\kappa E_g(s).$$

C'est effectivement possible pour $\operatorname{Re}(s) > 0$; si $E_g(s)$ se prolonge analytiquement légèrement à gauche de $\operatorname{Re}(s) = 0$ avec une croissance modérée, on peut déplacer le contour d'intégration vers la droite verticale $\operatorname{Re}(s) = \delta < 0$, faisant apparaître un pôle d'ordre $\ell + \kappa + 1$ en $s = 0$. Le résidu se calcule alors facilement :

$$\begin{aligned} \operatorname{res}_{s=0} \left(\zeta(s+1)^\kappa E_g(s) \frac{x^s}{s^{\ell+1}} \right) &= \frac{1}{(\kappa + \ell)!} E_g(0) (\ln x)^{\kappa+\ell} + O((\ln x)^{\kappa+\ell-1}) \\ &= \frac{c}{(\kappa + \ell)!} (\ln x)^{\kappa+\ell} + O((\ln x)^{\kappa+\ell-1}), \end{aligned}$$

ce qui explique la forme du terme principal ci-dessus.

En pratique, le prolongement de $E_g(s)$ n'est pas si évident, et il est en fait possible de démontrer l'estimation annoncée élémentairement, suivant des idées de Wirsing. Voir [18, Th. 1.1] ou [15, Lemma 5.4] pour les détails dans le cas $\ell = 0$ et [13, p. 11–12] pour le cas général.

Terminons enfin avec une version de ce théorème uniforme par rapport à g ; elle est nécessaire pour l'uniformité des estimations par rapport à \mathbf{h} dans la section 4.

THÉORÈME A.2. — *Soit g une fonction multiplicative, $\kappa \geq 0$ un entier. On note*

$$\rho(p) = \frac{g(p)}{1 + g(p)}$$

et on suppose qu'il existe $A_1 > 0$, $A_2 \geq 1$, $L \geq 1$ tels que

$$(30) \quad 0 \leq g(p) \leq A_1$$

$$(31) \quad -L \leq \sum_{w < p \leq z} \rho(p) \ln p - \kappa \ln \frac{z}{w} \leq A_2.$$

Alors on a

$$\sum_{d < x}^b g(d) \left(\ln \frac{x}{d} \right)^\ell = c \frac{\ell!}{(\ell + \kappa)!} (\ln x)^{\kappa+\ell} \left\{ 1 + O\left(\frac{L}{\ln x} \right) \right\}$$

pour tout entier $\ell \geq 0$ et tout $x \geq 2$, avec

$$c = \prod_p \left(1 - \frac{1}{p}\right)^\kappa (1 + g(p)),$$

la constante implicite dépendant seulement de κ , ℓ , A_1 , A_2 et L .

C'est en fait ce qui est démontré (à des changements de notation près) dans [15, Lemma 5.4] et [13, Lemma 5].

RÉFÉRENCES

- [1] E. BOMBIERI – *On twin almost primes*, Acta. Arith. 28 (1975–76), 177–193 ; Correction, id., 457–461.
- [2] E. BOMBIERI et H. DAVENPORT – *Small differences between prime numbers*, Proc. Roy. Soc. A 293 (1966), 1–18 ; aussi dans The Collected Works of H. Davenport, vol. IV, 1639–1656.
- [3] E. BOMBIERI, J. FRIEDLANDER et H. IWANIEC – *Primes in arithmetic progressions to large moduli*, Acta Math. 156 (1986), 203–251.
- [4] N. BOURBAKI – *Fonctions d'une variable réelle*, Paris, Hermann, 1976.
- [5] J.-M. DESHOILLERS – *Progrès récents des petits cribles (d'après Chen, Iwaniec)*, Sémin. Bourbaki (1977/78), Exp. 520, Lect. Notes in Math. 710 (1979), 248–262.
- [6] É. FOUVRY – *Autour du théorème de Bombieri-Vinogradov*, Acta Math. 152 (1984), 219–244.
- [7] É. FOUVRY – *Cinquante ans de théorie analytique des nombres*, dans Development of mathematics 1950–2000, 485–514, Birkhäuser, Basel, 2000.
- [8] É. FOUVRY et H. IWANIEC – *On a theorem of Bombieri-Vinogradov type*, Mathematika 27 (1980), 135–152.
- [9] P.X. GALLAGHER – *On the distribution of primes in short intervals*, Mathematika 23 (1976), 4–9.
- [10] D.A. GOLDSTON – *On Bombieri and Davenport's theorem concerning small gaps between primes*, Mathematika 39 (1992), 10–17.
- [11] D.A. GOLDSTON, J. PINTZ et C.Y. YILDIRIM – *Primes in tuples, I*. Prépublication arXiv:math.NT/0508185
- [12] D.A. GOLDSTON, J. PINTZ et C.Y. YILDIRIM – *The Path to Recent Progress on Small Gaps Between Primes*. Prépublication arXiv:math.NT/0512436
- [13] D.A. GOLDSTON, S.W. GRAHAM, J. PINTZ et C.Y. YILDIRIM – *Small gaps between primes or almost primes*. Prépublication arXiv:math.NT/0506067

- [14] D.A. GOLDSTON, Y. MOTOHASHI, J. PINTZ et C.Y. YILDIRIM – *Small Gaps between Primes Exist*. Prépublication arXiv:math.NT/0505300
- [15] H. HALBERSTAM et H. RICHERT – *Sieve methods*, Academic Press (1974).
- [16] D.R. HEATH-BROWN – *Almost prime k -tuples*, *Mathematika* 44 (1997), 245–266.
- [17] B. HOST – *Progressions arithmétiques dans les nombres premiers*, Sémin. Bourbaki (2004/2005), Exp. 944 (mars 2005), Astérisque 307 (2006).
- [18] H. IWANIEC et E. KOWALSKI – *Analytic Number Theory*, A.M.S Colloquium Publications vol. 53 (2004).
- [19] E. KOWALSKI – *Un cours de théorie analytique des nombres*, S.M.F Cours Spécialisé 13 (2004).
- [20] P. MICHEL – *Progrès récents du crible et applications (d’après Duke, Fouvry, Friedlander, Iwaniec)*, Sémin. Bourbaki (1997/98), Exp. 842, Astérisque 252 (1998), 185–209.
- [21] J. PINTZ et Y. MOTOHASHI – *A smoothed GPY sieve*. Prépublication arXiv:math.NT/0602599
- [22] A. SELBERG – *Lectures on sieves*, dans *Collected Papers*, vol. II, Springer-Verlag 1992 (pp. 65–247).
- [23] J. SIVAK – *Méthodes de crible appliquées aux sommes de Kloosterman et aux petits écarts entre nombres premiers*, Thèse de doctorat (Université Paris-Sud, décembre 2005).

Emmanuel KOWALSKI

Université de Bordeaux I

Laboratoire d’algorithmique arithmétique
et expérimentale (A2X)

UMR 5465 du CNRS

351, Cours de la Libération

F-33405 TALENCE

E-mail : kowalski@math.u-bordeaux1.fr

**MODULATION INVARIANT AND
MULTILINEAR SINGULAR INTEGRAL OPERATORS**
[after Lacey and Thiele]

by Michael CHRIST

Contents

1. Historical background.....	2
2. Localized Fourier coefficients.....	5
3. Almost-orthogonality.....	8
4. Trees, towers, and multitrees.....	10
5. Organizing the total sum into subsums.....	13
6. Counting multitrees.....	15
7. Refinement.....	17
8. Proof of the almost-orthogonality lemma.....	19
9. Carleson's maximal operator, according to Lacey and Thiele.....	19
10. Open problems.....	21
References.....	23

INTRODUCTION

The bilinear Hilbert transform is the operator

$$(1) \quad BH(f_1, f_2)(x) = \int_{\mathbb{R}} f_1(x-t)f_2(x+t)t^{-1}dt$$

where $x, t \in \mathbb{R}$ and $f_j \in L^{p_j}(\mathbb{R})$. If t^{-1} were an integrable function then this integral would become absolutely convergent, for almost every x for appropriate exponents p_j . The question of the finiteness of the conditional integral, and of inequalities in L^p norms, was an open problem from roughly the mid-1960's to the late 1990's, when Michael Lacey and Christoph Thiele showed in a series of breakthrough papers that BH is well-defined and bounded on appropriate L^p spaces. This operator is prototypical for a class of multilinear operators with modulation symmetry, and their work has been followed by significant further developments too numerous to cite in the space available.

In this expository article I discuss the background and origins of the problem, outline the main lines of the analysis, and indicate the connection with the almost everywhere convergence of Fourier integrals. This article is not intended as an exhaustive survey,

but merely as an introduction to the main ideas of the original articles [21],[22],[23]. I deliberately focus on one particular operator in order to emphasize what I consider to be the main concepts.

1. HISTORICAL BACKGROUND

1.1. Singular integrals

The most fundamental example of a Calderón-Zygmund singular integral operator is the Hilbert transform $Hf(x) = \pi^{-1} \int_{\mathbb{R}} f(x-t) t^{-1} dt$ for $x \in \mathbb{R}$. The integral fails to converge absolutely in general, and is defined as the limit as $\varepsilon \rightarrow 0$ of the integral over $|t| > \varepsilon$.

H plays a fundamental role in the theory of convergence of the Fourier transform, as well as in one-dimensional complex analysis. It satisfies $\widehat{Hf}(\xi) = i \operatorname{sgn}(\xi) \widehat{f}(\xi)$ for all $\xi \neq 0$, where $\operatorname{sgn}(\xi) = \pm 1$ according to whether $\xi > 0$ or < 0 . Thus $P = \frac{1}{2}(I - iH)$, where I is the identity, is the projection operator onto positive frequencies: $\widehat{Pf}(\xi) = \widehat{f}(\xi) \chi_{\xi > 0}$. The “partial sum” operators $\widehat{P_N f}(\xi) = \widehat{f}(\xi) \chi_{|\xi| \leq N}$ can be synthesized out of P together with shifts of the Fourier variable, in such a way that uniform boundedness of P_N on L^p is equivalent to boundedness of H on L^p . This is the basis of the classical theorem of M. Riesz on L^p norm convergence of Fourier series.

Somewhat more general CZ operators can be expressed as Fourier multiplier operators

$$(2) \quad \widehat{Tf}(\xi) = m(\xi) \widehat{f}(\xi) \quad \text{where } m(r\xi) \equiv m(\xi) \text{ for all } r > 0$$

and $m \in C^\infty(\mathbb{R}^d \setminus \{0\})$. General Fourier multipliers $\widehat{Tf}(\xi) = m(\xi) \widehat{f}(\xi)$ with $m \in L^\infty$ preserve $L^p(\mathbb{R})$ only for $p = 2$; there is no characterization of L^p functions in terms of the absolute values of their Fourier coefficients for $p \neq 2$.

The most general Calderón-Zygmund operators in \mathbb{R}^d lack convolution structure, taking the form $\int_{\mathbb{R}^d} K(x, y) f(y) dy$ where

$$(3) \quad |K(x, y)| \leq C|x - y|^{-d} \quad \text{and} \quad |\nabla_{x,y} K| \leq C|x - y|^{-d-1};$$

again I slur over the issue of interpretation of this typically absolutely divergent integral. Roughly speaking, (3) says that the portions of f, g microlocalized in phase space near (x, ξ) and (x', ξ') respectively interact quite weakly unless $|\xi| + |\xi'| \leq C|x - x'|^{-1}$. According to the uncertainty principle, any stronger restriction of this general type is meaningless. A basic theorem [2] states that *if* such an operator is bounded on L^2 , then it is also bounded on L^p for all $p \in (1, \infty)$.

The basic symmetries of this theory are translation and dilation; if $K(x, y)$ is a Calderón-Zygmund kernel then so are $K(x - z, y - z)$ and $r^d K(rx, ry)$, *uniformly* for all $r > 0$ and $z \in \mathbb{R}^d$. The individual operators need not exhibit these symmetries, but the class as a whole does.

A third basic symmetry, with respect to *modulation*, is totally lacking in this theory. Multiplying $K(x, y)$ by $e^{i(ax+by)}$ for any nonzero $(a, b) \in \mathbb{R}^2$ destroys the bound on ∇K .

This lack of symmetry is perhaps even more apparent in (2), in the convolution case $K(x - y)$, where $\xi = 0$ plays a privileged role. Of course, such a modulation does not affect L^p estimates, but as we will see, the bilinear Hilbert transform can be regarded as an infinite sum of modulated Calderón-Zygmund operators with different modulating frequencies, in such a way that boundedness of the sum cannot easily be inferred by summing bounds for the individual summands.

1.2. Calderón's commutator

Calderón had an abiding interest in partial differential equations with nonsmooth coefficients and on nonsmooth domains. He had employed algebras of singular integral operators in studying PDE, for instance in his work on uniqueness in the Cauchy problem [3]. Thus he was naturally led to investigate compositions of operators such as the canonical example H , the operator M_A of multiplication by a function A having limited smoothness, and $\frac{d}{dx}$. He showed in 1965 [4] that the commutator $[H, M_A]$ is smoothing, in the sense that $\mathcal{C}_A = \frac{d}{dx} \circ [H, M_A]$ is bounded on $L^2(\mathbb{R}^1)$, whenever A is Lipschitz continuous, that is, whenever $a = dA/dx \in L^\infty$. Formally

$$(4) \quad \mathcal{C}_A f(x) = \int_{\mathbb{R}} f(y) \frac{A(x) - A(y)}{(x - y)^2} dy,$$

which satisfies the Calderón-Zygmund assumptions (3) when $a = \frac{dA}{dx}$ belongs to L^∞ . These operators possess translation and dilation invariance as a family, even though individually they lack it.

Since the commutator operator is not translation-invariant, Plancherel's theorem can not be invoked directly to establish its L^2 boundedness. A key realization of Calderón was that it could profitably be regarded as a *bilinear* operator, and that the full force of Fourier analysis and complex variables methods should be brought to bear on a .

An intriguing alternative expression is obtained by writing $A(x) - A(y) = (x - y) \int_0^1 a(sx + (1 - s)y) ds$ to obtain a decomposition $\mathcal{C}_A(f) = \int_0^1 \mathcal{C}_s(f, a) ds$ where

$$(5) \quad \mathcal{C}_s(f, a)(x) = \int_{\mathbb{R}} f(x - t) a(x + st) t^{-1} dt.$$

Thus bounds for \mathcal{C}_s from $L^2 \times L^\infty$ to L^2 would imply corresponding bounds for the commutator operator. The special case \mathcal{C}_1 is traditionally called the *bilinear Hilbert transform*, but all the operators \mathcal{C}_s for $s \neq 0, -1$ have essentially the same intrinsic qualities and stature. Calderón asked⁽¹⁾ whether these operators do map $L^2 \times L^\infty$ to L^2 . The problem became notorious, but was not resolved until the work of Lacey and Thiele [21],[22] in the late 1990s.

Thought of as linear operators acting on f , \mathcal{C}_s have nonsmooth kernels $K(x, y) = (x - y)^{-1} a(xs + (1 - s)y)$ which satisfy no gradient estimate. Viewed as bilinear operators, they are singular in the sense that $\mathcal{C}_s(f, a)(x)$ depends on $a(y_1)f(y_2)$ only for (y_1, y_2) in a one-dimensional subset of \mathbb{R}^2 .

⁽¹⁾The question is widely attributed to Calderón, though I know of no reference.

It is remarkable that these building blocks \mathcal{C}_s not only retain translation and dilation symmetry, but gain new *modulation* symmetries: defining $\mathcal{M}_\eta f(x) = e^{ix\eta} f(x)$,

$$(6) \quad \boxed{\mathcal{C}_s(\mathcal{M}_{s\eta} f, \mathcal{M}_\eta a) \equiv \mathcal{M}_{(1+s)\eta} \mathcal{C}_s(f, a).}$$

These are partial symmetries; there is no relation for $\mathcal{C}_s(\mathcal{M}_\eta f, \mathcal{M}_{\tilde{\eta}} a)$ unless $s\eta = \tilde{\eta}$.

In terms of the Fourier transform the operator is written

$$(7) \quad \mathcal{C}_s(f, a)(x) = c \iint e^{ix(\xi_1 + \xi_2)} \operatorname{sgn}(s\xi_2 - \xi_1) \widehat{f}(\xi_1) \widehat{a}(\xi_2) d\xi_1 d\xi_2$$

for a certain constant c , and the modulation symmetry is reflected in the invariance of the Fourier multiplier $\operatorname{sgn}(s\xi_2 - \xi_1)$ under $\xi \mapsto \xi + (s\eta, \eta)$. This multiplier is nonsmooth along an entire line, rather than merely at the origin.

It is (perhaps) a general principle that more symmetric operators are more difficult to analyze; a featureless wall presents no cracks which can naturally be enlarged into gaps. A fundamental point to look for in the discussion below is how the symmetry is broken; see §5.

1.3. Carleson’s maximal operator

Carleson [6] proved in 1966 that for any periodic function $f \in L^2$ of one real variable, the partial sums of the Fourier series converge to f almost everywhere. The essentially equivalent statement for the real line is that $(2\pi)^{-1} \int_{|\xi| \leq N} \widehat{f}(\xi) e^{ix\xi} d\xi$ converges to $f(x)$ as $N \rightarrow \infty$, for almost every $x \in \mathbb{R}$. The main ingredient is an estimate for Carleson’s maximal operator $\mathcal{C}^* f(x) = \sup_{N < \infty} \left| \int_{|\xi| \leq N} \widehat{f}(\xi) e^{ix\xi} d\xi \right|$, which is essentially the same as

$$(8) \quad \mathcal{C}^* f(x) = \sup_{N \in \mathbb{R}} \left| \int_{\mathbb{R}} f(x - t) e^{iNt} t^{-1} dt \right|.$$

Carleson proved that \mathcal{C}^* maps L^2 to weak L^2 , that is, $|\{x : \mathcal{C}^* f(x) > \lambda\}| \leq C \lambda^{-2} \|f\|_{L^2}^2$ uniformly for all $\lambda > 0$ and $f \in L^2$. Almost everywhere convergence follows immediately from this inequality since it holds trivially for functions whose Fourier transforms have compact support.

It is equivalent to establish bounds for the linear operators $\int_{\mathbb{R}} f(x - t) e^{iN(x)t} t^{-1} dt$ which are *uniform* over all measurable real-valued selection functions N . Once again these operators enjoy forms of translation, dilation, and modulation invariance. For instance, $\mathcal{L}(\mathcal{M}_\eta f)(x) = \mathcal{M}_\eta \mathcal{L}' f(x)$, where \mathcal{L}' is obtained from \mathcal{L} by replacing the function $N(x)$ by $N(x) - \eta$.

Fefferman [17] later gave a second proof of Carleson’s theorem. Lacey and Thiele used elements of both of these analyses to prove⁽²⁾

THEOREM 1.1. — *Let $p_1, p_2, q \in (1, \infty]$ satisfy $q^{-1} = p_1^{-1} + p_2^{-1}$, and assume that no more than one of these exponents is infinite. Then there exists $C < \infty$ such that $\|BH(f_1, f_2)\|_{L^q} \leq C \|f_1\|_{L^{p_1}} \|f_2\|_{L^{p_2}}$ for all Schwartz class functions.*

⁽²⁾Their theorem actually applies for all $q > \frac{2}{3}$.

1.4. Two roads diverge

Calderón proved the bound he sought for the commutator operator without understanding the bilinear Hilbert transform, and went on to analyze [5] the Cauchy integral associated to Lipschitz curves with small Lipschitz constant by an extension of those ideas. Further developments have included a vast literature on elliptic boundary problems on Lipschitz domains, analytic capacity in one complex variable [27], the work of Coifman-Meyer-McIntosh [12] on the Cauchy integral, and the $T(1)$ theorem of David and Journé [13]. A theory of multilinear Calderón-Zygmund singular operators was developed [9][10], which however does not include \mathcal{C}_s ; it encompasses operators which have a Fourier representation like (7) with $\text{sgn}(s\xi_2 - \xi_1)$ replaced by functions smooth away from $\xi = 0$ and satisfying $m(r\xi) \equiv m(\xi)$ for $r > 0$. These operators lack modulation invariance, and are less singular. Some of that theory provides essential building blocks for the analysis outlined here.

2. LOCALIZED FOURIER COEFFICIENTS

2.1. A frame with a preferred scale

Let $\psi : \mathbb{R}^1 \rightarrow \mathbb{C}$ be an infinitely differentiable function supported in $(0, 2)$ such that $\sum_{n \in \mathbb{Z}} \psi(t - n) \equiv 1$ for all $t \in \mathbb{R}$. Then the set of all functions $\{\psi_{k,n} = e^{ikt}\psi(t - n) : k, n \in \mathbb{Z}\}$ is a *frame* for $L^2(\mathbb{R}^1)$; for any $f \in L^2$,

$$(9) \quad f = c \sum_{k,n} \psi_{k,n} \langle f, \psi_{k,n} \rangle$$

for a certain constant c whose precise value is of no consequence for the type of inequality in question here. The inverse Fourier transform of ψ is a Schwartz function, and multiplying it by $(2\pi)^{-1/2}$ yields a function φ such that $\{\varphi_{k,n}(x) = e^{inx}\varphi(x - k)\}$ is likewise a frame for L^2 . It is good intuition to think of $\varphi_{k,n}(x)$ as being essentially $c_0 e^{inx} e^{-|x-k|^2}$, although this is not quite correct because these functions lack compactly supported Fourier transforms.

One thinks of $\langle f, \varphi_{k,n} \rangle$ as being *localized Fourier coefficients*. Such a frame is quite different from celebrated *wavelet bases*. The lesser difference is that $\{\varphi_{k,n}\}$ is not an orthonormal system; there is some oversampling here. The significant difference is that whereas a wavelet basis treats all scales equally, this frame prefers one scale. An advantage of this frame, not shared by wavelet-type bases, is its invariance under modulation by integral frequencies.

The rank one operator $f \mapsto \langle f, \varphi_{k,n} \rangle \varphi_{k,n}$ heuristically represents the orthogonal projection of $L^2(\mathbb{R})$ onto the subspace consisting of all functions g such that g is supported in $I = [k, k + 1]$ and \widehat{g} is supported in $\omega = [n, n + 1]$, although this is not exactly true. The entire *phase space* $\mathbb{R} \times \mathbb{R}$ is tiled by these sets $I \times \omega$, and corresponding to this geometric decomposition is the analytic decomposition of the identity operator as a sum of projections.

2.2. Scaled frames and tiles in phase space

Other frames can be constructed by scaling: For each integer r we form $\{2^{-r/2}\varphi_{k,n}(2^{-r}x) : k, n \in \mathbb{Z}\}$, which is likewise a frame. This corresponds to tiling \mathbb{R}^2 by rectangles $I \times \omega$ where I has length 2^r and ω has length 2^{-r} .

A *dyadic interval* is a closed bounded subinterval of \mathbb{R}^1 of the form $[k2^n, (k + 1)2^n]$ for arbitrary integers k, n . The set of all dyadic intervals enjoys an often useful combinatorial property: If two such intervals do overlap, then one is contained in the other.

DEFINITION 2.1. — A tile is a subset of the phase space \mathbb{R}^2 of the form $I \times \omega$, where I, ω are arbitrary dyadic intervals satisfying $|I| \cdot |\omega| = 1$.

In contrast to dyadic intervals, no tile is properly contained in another. Tiles and dyadic intervals are said to be *nonoverlapping* if their interiors are disjoint⁽³⁾.

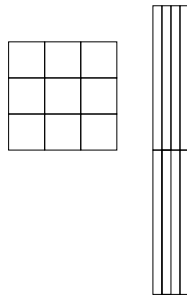


FIGURE 1. Two tilings of phase space. According to the uncertainty principle, no significant refinement of either tiling is meaningful.

2.3. Decomposition of the bilinear Hilbert transform by scales

The bilinear Hilbert transform has no preferred scale; with respect to the operators $D_\lambda f(x) = f(\lambda x)$, there is the dilation symmetry $BH(D_\lambda f, D_\lambda g) = D_\lambda(BH(f, g))$. Thus none of our frames is well adapted to the operator.

Let $\eta : \mathbb{R} \rightarrow \mathbb{C}$ be a smooth function supported⁽⁴⁾ in (8, 32) such that $\sum_{r \in \mathbb{Z}} \eta(2^r \xi) \equiv 1$ for all $\xi > 0$. BH is thus decomposed as $\sum_{r \in \mathbb{Z}} BH_r$ where

$$(10) \quad BH_r(f_1, f_2)(x) = c \int_{\mathbb{R}^2} e^{ix(\xi_1 + \xi_2)} \operatorname{sgn}(\xi_1 - \xi_2) \eta(2^r(\xi_1 - \xi_2)) \widehat{f}_1(\xi_1) \widehat{f}_2(\xi_2) d\xi_1 d\xi_2,$$

plus a second, very similar, infinite sum obtained by replacing $\eta(s)$ by $\eta(-s)$.

There is an alternative expression

$$(11) \quad BH_r(f_1, f_2)(x) = \int_{\mathbb{R}} f_1(x + t) f_2(x - t) 2^{-r} h(2^{-r}t) dt$$

⁽³⁾I will sometimes abuse language and notation by saying that intervals are disjoint when they merely do not overlap.

⁽⁴⁾The precise numbers 8, 32 are of no significance; one could replace these by any A, B such that $B > 2A$ and $A > 0$ is sufficiently large.

for a certain Schwartz function h . An immediate consequence of Hölder’s inequality is that BH_r maps $L^{p_1} \times L^{p_2}$ to L^q whenever p_1, p_2, q are ≥ 1 and satisfy $q^{-1} = p_1^{-1} + p_2^{-1}$. and the scaling symmetry ensures that the operator norm is independent of r . The sole issue is the summation over r .

BH_r clearly has a preferred spatial scale, 2^r , and retains the modulation invariance of BH . For each $r \in \mathbb{Z}$, BH_r will be analyzed in terms of the frame described above with the same parameter r . Thus one works simultaneously with infinitely many frames, one for each scale.

2.4. Phase space decomposition of the bilinear Hilbert transform

For each $r \in \mathbb{Z}$ denote by \mathcal{P}_r the set of all phase space tiles $P = I \times \omega$ of dimensions $2^r \times 2^{-r}$. Consider the trilinear form $\mathcal{H}_r(f_1, f_2, f_3) = \langle f_3, BH_r(f, f_2) \rangle$ for an arbitrary test function f_3 . Then $\langle BH(f_1, f_2), f_3 \rangle = \sum_{r=-\infty}^{\infty} \mathcal{H}_r(f_1, f_2, f_3)$.

Decompose each $f_k = \sum_{P^k \in \mathcal{P}_r} \langle f_k, \phi_{P^k} \rangle \phi_{P^k}$ to obtain

$$\mathcal{H}_r(f_1, f_2, f_3) = \sum_{P=(P^1, P^2, P^3) \in \mathcal{P}_r^3} \beta_P \prod_{k=1}^3 \langle f_k, \phi_{P^k} \rangle$$

for certain coefficients β_P independent of $\{f_k\}$. Upper bounds for the “interaction amplitudes” β_P are required, and elementary estimates combined with the information that η is supported in (8, 32) give:

LEMMA 2.2. — *Let $P = (P^1, P^2, P^3)$ where each $P^k = I_k \times \omega_k$ is a phase space tile of dimensions $2^r \times 2^{-r}$. Then for any finite exponent N ,*

$$(12) \quad |\beta_P| \leq C_N 2^{-r/2} \left(\max_{k, l \in \{1, 2, 3\}} (1 + 2^{-r} \text{distance}(I_k, I_l)) \right)^{-N}$$

and

$$(13) \quad \beta_P \equiv 0$$

unless the centerpoints c_k of the frequency space intervals ω_k satisfy $c_2 - c_1 \in [2 \cdot 2^{-r}, 38 \cdot 2^{-r}]$ and $c_3 - 2c_1 \in [10 \cdot 2^{-r}, 70 \cdot 2^{-r}]$.

This is a reflection of a fundamental characteristic of the bilinear Hilbert transform: $BH(e^{i\xi_1 \cdot}, e^{i\xi_2 \cdot})(x) = \pi i \operatorname{sgn}(\xi_1 - \xi_2) e^{i\xi_3 x}$ where $\boxed{\xi_3 = \xi_1 + \xi_2}$.

The numbers 2, 38, 10, 70 are insignificant artifacts of certain nearly arbitrary choices. What is important is the following consequence, which will be a source of orthogonality in the analysis.

FACT 2.3. — *Let $P, Q \in \cup_r \mathcal{P}_r^3$ and suppose that β_P, β_Q are both nonzero. Suppose that for some index $k \in \{1, 2, 3\}$, $\omega_{P^k}, \omega_{Q^k}$ overlap, and that $|\omega_{P^k}| < |\omega_{Q^k}|$. Then for each $i \neq k$, ω_{P^i} is disjoint from ω_{Q^i} , and they are separated by a distance comparable to $|\omega_{Q^i}|$.*

2.5. Nuisance technicalities involving tiles

If I, J are dyadic intervals, then either one is contained in the other, or they do not overlap. This makes dyadic intervals well suited to stopping time arguments, in which one begins with such an interval, subjects it to a test, and if it fails, subdivides it into halves and subjects the two halves separately to (rescaled versions of) the same test, repeating indefinitely.

On the other hand, the set of all dyadic intervals has no reasonable translation invariance. Consider the interval $[0, 1]$, which for any integer $N \geq 1$ is contained in the larger dyadic interval $[0, 2^N]$. These larger intervals have the unnatural feature that they extend only to the right of $[0, 1]$, never to the left; their union is only half of the real axis. Thus analysis based on these intervals is likely to disregard interactions between the two halves of the real axis. This defect is essentially reprised at every dyadic point $j2^n$, $j, n \in \mathbb{Z}$.

This difficulty arises commonly and has been sidestepped in various ways by various authors; see [6], [18], [26]. It is also helpful to thin out the sum by partitioning the set of all dyadic intervals into finitely many subfamilies, so that for any two intervals I, J belonging to any common subfamily, if $I \subset J$ and $I \neq J$ then $|J| \geq 2^K |I|$ where K is a large constant. This leads to a decomposition into finitely many suboperators, all having the same structure.

In this exposé I will systematically slur over these technicalities, which are of no intrinsic interest. I do not pretend to give a full proof, only a conceptually accurate outline. Statements made below are correct, but under the proviso that these technicalities have been dealt with.

3. ALMOST-ORTHOGONALITY

3.1. Introduction

The space L^2 plays a special role in the classical singular integral operator theory, partly because methods relying on Hilbert space structure are available. In particular, Plancherel's theorem can be applied to easily establish L^2 estimates for translation-invariant operators; but it is not directly applicable to the commutator operator or its more degenerate relatives.

A rather flexible almost-orthogonality principle was introduced by Knapp and Stein [20], who were motivated by problems in the representation theory of semisimple groups to establish L^2 bounds for singular integral operators invariant with respect to certain (nilpotent) Lie group structures. They showed that if an abstract operator T is decomposed as a sum of bounded operators T_j such that $\|T_i T_j^*\|, \|T_i^* T_j\| \leq c_{i-j}$ for all $i, j \in \mathbb{Z}$, then $\|T\| \leq \sum_j c_j^{1/2}$. A trivial case is when the summands have pairwise orthogonal ranges, and likewise for their adjoints.

A satisfying explanation of the L^2 boundedness of Calderón-Zygmund operators was finally obtained by David and Journé, who showed that any such (bounded) operator can be decomposed as a sum of three parts, one of which has a natural almost-orthogonal decomposition in the sense of Knapp and Stein. The other two parts have a different structure, related to the concepts of Carleson measures and paraproducts [10]. It is remarkable that on this level the ultimate understanding of the fundamental L^2 estimate rests on the theory surrounding BMO, which is one limit of L^p as $p \rightarrow \infty$.

3.2. Orthogonality via phase space disjointness

If two functions f, g have disjoint supports, then of course $\langle f, g \rangle = 0$. The same goes if \widehat{f}, \widehat{g} have disjoint supports. If P, Q are nonoverlapping tiles, then either ω_P and ω_Q are nonoverlapping, or I_P and I_Q are. In the former case, $\langle \phi_P, \phi_Q \rangle = 0$, but in the latter case the two supports cannot be disjoint (both functions are real analytic). Nonetheless, since $\phi_P(x)$ decays rapidly as x moves away from I_P , $\langle \phi_P, \phi_Q \rangle$ is relatively small if I_P, I_Q are far apart. Therefore one hopes to retain some form of orthogonality. The following lemma⁽⁵⁾ is analogous to Bessel’s inequality for Fourier coefficients, but is slightly weakened by the necessity of a supplementary hypothesis⁽⁶⁾ (14).

LEMMA 3.1. — *Let \mathcal{S} be any set of pairwise nonoverlapping tiles. Let $f \in L^2$ and $\lambda > 0$. Suppose that for every $P \in \mathcal{S}$,*

$$(14) \quad \lambda |I_P|^{1/2} \leq |\langle f, \phi_P \rangle| \leq 2\lambda |I_P|^{1/2}.$$

Then $\boxed{\sum_{P \in \mathcal{S}} |\langle f, \phi_P \rangle|^2 \leq C \|f\|_{L^2}^2}$.

The constant C is independent of f, λ, \mathcal{S} . An equivalent statement of the conclusion is $\sum_{P \in \mathcal{S}} |I_P| \leq C \lambda^{-2} \|f\|_{L^2}^2$; $\sum_P |I_P|$ is a weighted count of the number of tiles satisfying (14). For a complete proof of this fundamental fact see §8.

A localized variant is often useful: Suppose that there is given an interval J such that $I_P \subset J$ for all $P \in \mathcal{S}$. Then under hypothesis (14),

$$(15) \quad \sum_{P \in \mathcal{S}} |\langle f, \phi_P \rangle|^2 \leq C_M \int_{\mathbb{R}} |f(x)|^2 \left(1 + \frac{\text{distance}(x, J)}{|J|} \right)^{-M} dx$$

for any finite M . This is natural, since $|\phi_P(x)|$ decays rapidly as x moves away from J .

⁽⁵⁾The original proofs of Lacey and Thiele [21],[22] were more complicated than the one outlined here, in large part because they had only a weaker version of Lemma 3.1 at their disposal. Lemma 3.1, in a slightly more sophisticated form, is in their paper [23].

⁽⁶⁾Without a supplementary hypothesis of some kind, Lemma 3.1 is false.

4. TREES, TOWERS, AND MULTITREES

The set of all tiles is endowed with a partial ordering.

DEFINITION 4.1. — $P \leq Q$ if and only if $I_P \subset I_Q$ and $\omega_P \supset \omega_Q$. Also $P < Q$ means that $P \leq Q$ and $P \neq Q$.

DEFINITION 4.2. — A multitile P is an ordered 3-tuple of tiles (P^1, P^2, P^3) such that $I_{P^i} = I_{P^j}$ for all indices i, j , and the centers of the associated frequency intervals ω_{P^j} satisfy the constraints listed following (13).

This common interval I_{P^k} is denoted by I_P . To any multitile are associated three functions ϕ_{P^k} , $k \in \{1, 2, 3\}$. For any single P , the associated three rectangles $I_P \times \omega_P^k$ all share the same dimensions.

4.1. Model operator

A simplified model for the sum that represents the bilinear Hilbert transform is

$$(16) \quad \mathcal{H}_{\text{model}}(f_1, f_2, f_3) = \sum_P |I_P|^{-1/2} \prod_{k=1}^3 |\langle f_k, \phi_{P^k} \rangle|$$

where the sum ranges over all multitiles, or some large collection of multitiles. These terms are normalized so that $|I_P|^{-1/2} \prod_k |\langle f_k, \phi_{P^k} \rangle| \leq C \prod_k \|f_k\|_{L^{p_k}}$ whenever $\sum_k p_k^{-1} = 1$, with C independent of P . The theorem asserts that this sum of infinitely many uniformly bounded operators is bounded.

The condition (12) that β_P is small whenever I_{P^k}, I_{P^i} are far apart has been simplified in this model to the condition that $\beta_P = 0$ whenever $I_{P^k} \neq I_{P^i}$. Only this model will be discussed further; but see [21] for an explanation of how the bilinear Hilbert transform can actually be realized as a limit of averages of such model operators.

One should regard each factor $\langle f_k, \phi_{P^k} \rangle$ as expressing one irreducible bit of information, and each term in the sum as being likewise irreducible. This minimality would be lost if a single frame for $L^2(\mathbb{R})$ were used.

4.2. Trees, towers, and multitrees

The analysis proceeds by decomposing the full sum (16) over all multitiles into sums over various subfamilies enjoying additional structure.

DEFINITION 4.3. — A tower is a nonempty finite set T of tiles such that there exists a tile $\mathbf{top}(T) \in T$, called the top of T , such that every $P \in T$ satisfies $P \leq \mathbf{top}(T)$, that is, $I_P \subset I_{\mathbf{top}(T)}$ and $\omega_P \supset \omega_{\mathbf{top}(T)}$.

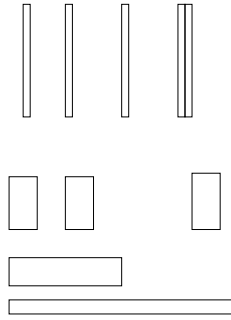


FIGURE 2. A tree. The x axis is horizontal; the ξ axis is vertical. Towers are similar, except that there exists some horizontal line which crosses every rectangle.

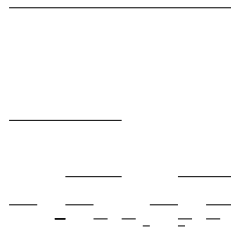


FIGURE 3. All intervals I_P associated to tiles $P \in T$, where T is either a tree or a tower. These intervals are contained in \mathbb{R} , but each is displaced vertically for purposes of illustration. $I_{\mathbf{top}(T)}$ is the topmost interval. The corresponding frequency space intervals ω_P have lengths inversely proportional to $|I_P|$. These form a nested set if T is a tower, whereas any two of different lengths are nonoverlapping if T is a tree.

DEFINITION 4.4. — A tree T is a finite set of tiles with an element $\mathbf{top}(t) \in T$, such that for any $P \neq Q \in T$:

- (i) $I_P \subset I_{\mathbf{top}(T)}$,
- (ii) P, Q do not overlap,
- (iii) if $|I_P| \neq |I_Q|$ then $\omega_P \cap \omega_Q = \emptyset$.

Moreover,

- (iv) there exists $\xi_0 \in \mathbb{R}$ such that $\text{distance}(\omega_P, \xi_0) \sim |\omega_P|$ for all $P \in T$.

DEFINITION 4.5. — Let $k \in \{1, 2, 3\}$. A finite set T of multitiles is said to be a k -tower if $T^k = \{P^k : P \in T\}$ is a tower. A multitree is a set of multitiles which is a k -tower for some index k .

A k -tower has a unique topmost element $\mathbf{top}(T)$, such that $\mathbf{top}(T)^k$ is the top of the tower T^k .

4.3. One origin of orthogonality

Plancherel’s and Parseval’s theorems and Bessel’s inequality lie at the heart of many more sophisticated arguments in classical Fourier analysis. Here a fundamental role is played by inequalities based on orthogonality, reflecting the principle that ϕ_P, ϕ_Q are nearly orthogonal when the subsets P, Q of phase space do not overlap. We now discuss one property of the set of all multitiles, which leads to families of nonoverlapping tiles in the analysis of multilinear singular integrals.

Towers T are sets of tiles which totally lack useful almost-orthogonality, since $\omega_P \cap \omega_Q \neq \emptyset$ for any $P, Q \in T$. In the case where $0 \in \omega_{\text{top}(T)}$, the rank one operators $f \mapsto \langle f, \phi_Q \rangle \phi_Q$ for $Q \in T$ are closely analogous to averaging operators $f \mapsto (|I_Q|^{-1} \int_{I_Q} f) \cdot \chi_{I_Q}$ (and more generally to frequency-modulated generalizations $f \mapsto e^{ix\xi_0} (|I_Q|^{-1} \int_{I_Q} e^{-ix\xi_0} f) \cdot \chi_{I_Q}(x)$ for $\xi_0 \in \omega_{\text{top}(T)}$). A tower could include a large family of Q such that $|I_Q| \rightarrow 0$ and I_Q approaches some point x_0 . For almost all such x_0 , $|I_Q|^{-1} \int_{I_Q} f \rightarrow f(x_0)$ by Lebesgue’s differentiation theorem. Thus for a tower, no upper bound may be available for the coefficients $|I_Q|^{-1/2} \langle f, \phi_Q \rangle$ beyond their uniform boundedness.

However, in the context of multitiles, there is substantial compensation for this lack of orthogonality.

FACT 4.6. — *If a family T of multitiles is an i -tower for some $i \in \{1, 2, 3\}$ then $\{P^j : P \in T\}$ is a tree for any $j \neq i$ in $\{1, 2, 3\}$.*

This crucial consequence of Fact 2.3 partly explains the terminology “multitree”.

Expressions $(\sum_{Q \in T} |\langle f, \phi_Q \rangle|^2)^{1/2}$, where T is a tree, are closely analogous to classical Littlewood-Paley expressions and are central to the analysis. An important part of this analogy is the next bound.

LEMMA 4.7. — *Uniformly for all trees T and all functions f ,*

$$(17) \quad \sum_{P \in T} |\langle f, \phi_P \rangle|^2 \leq C \min (\|f\|_{L^2}^2, |I_{\text{top}(T)}| \|f\|_{L^\infty}^2).$$

These inequalities lend Fact 4.6 great significance. The second conclusion is a localized version of the first.

The basic principle underlying Lemma 4.7 is that since the tiles $P \in T$ form a nearly disjoint family of subsets of phase space, $\{\phi_P : P \in T\}$ is an almost-orthogonal family of functions. The lemma is not quite a consequence of Lemma 3.1 on almost-orthogonality via nonoverlapping tiles, but follows easily from a direct examination of the magnitudes of matrix coefficients $\langle \phi_P, \phi_Q \rangle$ for $P, Q \in T$.

4.4. Energy and mass of multitrees

The j -energy $\mathcal{E}_j(T)$ and j -mass $\mathcal{M}_j(T)$ of a multitree T are defined for $j \in \{1, 2, 3\}$ to be

$$(18) \quad \mathcal{E}_j(T) = |I_{\text{top}(T)}|^{-1/2} \left(\sum_{P \in T} |\langle f_j, \phi_{P^j} \rangle|^2 \right)^{1/2}$$

$$(19) \quad \mathcal{M}_j(T) = \sup_{P \in T} |I_P|^{-1/2} |\langle f_j, \phi_{P^j} \rangle|.$$

The individual terms are normalized so that $|I_P|^{-1/2} |\langle f_j, \phi_{P^j} \rangle| \leq CM f_j(x)$ for $x \in I_P$, and if $\{P^j : P \in T\}$ is a tree then $\mathcal{E}_j(T) \leq C \|f_j\|_{L^\infty}$ (by Lemma 4.7).

If i, j, k are the three elements of $\{1, 2, 3\}$ in any order then

$$(20) \quad |\mathcal{H}_T(f_1, f_2, f_3)| \leq \mathcal{E}_j(T) \mathcal{E}_k(T) \mathcal{M}_i(T) |I_{\text{top}(T)}|.$$

(20) is a direct consequence of definitions via Cauchy-Schwarz, and should be regarded as a manifesto of intent rather than as a genuine estimate. Proposition 7.2 below provides alternative upper bounds for \mathcal{H}_T , not subsumed in the mass-energy bound.

5. ORGANIZING THE TOTAL SUM INTO SUBSUMS

The stage has been set for a discussion, in this section and the next, of the heart of the proof: a sorting algorithm, a counting problem, and a counting estimate based on phase space orthogonality. To any finite set \mathcal{S} of multitiles and any three functions $f_j : \mathbb{R}^1 \rightarrow \mathbb{C}$ is associated the operator expression

$$(21) \quad \mathcal{H}_{\mathcal{S}}(f_1, f_2, f_3) = \sum_{P \in \mathcal{S}} |I_P|^{-1/2} \prod_{j=1}^3 |\langle f_j, \phi_{P^j} \rangle|.$$

We will partition the set of all multitiles into subsets having special structure, derive a reasonable bound for the contribution of each subset, and sum those bounds.

5.1. The sorting algorithm's output

In the proof we examine the sum $\mathcal{H}_{\mathcal{S}}$ of contributions of an arbitrary finite set of multitiles \mathcal{S} . The sorting procedure detailed below partitions the collection \mathcal{S} into multitrees. It constructs families $\mathcal{F}_{n,i,j}$ of multitrees, indexed by $n \in \mathbb{Z}$ and $i, j \in \{1, 2, 3\}$, with the following main properties:

- (i) \mathcal{S} is the disjoint union, over all n and all ordered pairs (i, j) , of all $T \in \mathcal{F}_{n,i,j}$.
- (ii) Each $T \in \mathcal{F}_{n,i,j}$ is an i -tower.
- (iii) For any $T \in \mathcal{F}_{n,i,j}$,

$$(22) \quad |\mathcal{H}_T(f_1, f_2, f_3)| \leq C 2^{3n} |I_{\text{top}(T)}|.$$

(iv) If $j = i$ then

$$(23) \quad |I_{\text{top}(T)}| \leq 2^{-2n} |\langle f_i, \phi_{\text{top}(T)^i} \rangle|^2.$$

(v) If $j \neq i$ then

$$(24) \quad |I_{\text{top}(T)}| \leq 2^{-2n} \sum_{P \in T} |\langle f_j, \phi_{P^j} \rangle|^2.$$

(vi) Each $T \in \mathcal{F}_{n,i,j}$ enjoys certain maximality properties.

5.2. The sorting algorithm

The algorithm seeks out the enemy in the form of subsums corresponding to i -towers; such subsums are potentially unfavorable because of the lack of orthogonality discussed above. We proceed to construct families $\mathcal{F}_{n,i,j}$ of multitrees $T \subset \mathcal{S}$, for all $n \in \mathbb{Z}$ and all $(i, j) \in \{1, 2, 3\}^2$. The construction proceeds by descending induction on n . Begin with a very large positive n . Order the 9 ordered pairs of indices $(i, j) \in \{1, 2, 3\}$ arbitrarily.

Fix a small constant $c_0 > 0$. For the first pair (i, j) , if $i = j$ then consider all i -towers $T \subset \mathcal{S}$ such that

(i) $|I_P|^{-1/2} |\langle f_i, \phi_{P^i} \rangle| \geq c_0 2^n$ for every $P \in T$

(ii) T is maximal with respect to set inclusion. That is, there exists no i -tower $T' \subset \mathcal{S}$ properly containing T which satisfies (i).

If there exists a nonempty i -tower $T \subset \mathcal{S}$ satisfying (i), then there also exists one satisfying both criteria. Choose any one, T , put it into $\mathcal{F}_{n,i,j}$, and delete all tiles $P \in T$ from \mathcal{S} . Repeat the procedure with this reduced set \mathcal{S} for the same pair (i, j) until no multitrees satisfying (i) remain. Then move on to the next pair (i, j) . Observe how modulation symmetry is broken; the decomposition depends on the localized Fourier coefficients $\langle f_i, \phi_{P^i} \rangle$ of f_i .

If $j \neq i$ then do the same, retaining (ii) but replacing (i) by

(i)* $\mathcal{E}_j(T) \geq 2^n$,

and imposing a supplementary condition (iii)* whose role will not be visible in this exposé⁽⁷⁾

Continue with a given index n until all 9 pairs (i, j) have been fully examined, and no multitrees satisfying the criteria remain. Then replace n by $n - 1$, and repeat the selection again, with n replaced by $n - 1$ in criteria (i) and (i)*. Continue by descending induction on n until only tiles satisfying $\langle f_k, \phi_{P^k} \rangle = 0$ for all $k \in \{1, 2, 3\}$ remain. Those contribute nothing to the operator, and may be discarded.

⁽⁷⁾Criterion (iii)*: If $i < j$, then a candidate i -tower T is not selected if there is some other candidate multitree T' such that ω_T^i lies strictly to the left of $\omega_{T'}^i$. (An interval $[a, b]$ is said to lie strictly to the left of $[c, d]$ if $b \leq c$.) The same goes if $i > j$, with “left” replaced by “right”. The reader is urged to disregard this point for the present. Likewise the smallness of the constant c_0 in (i) plays a role which will not be apparent at the level of detail of this exposé.

Given a finite set \mathcal{S} , there exists n_0 so large that (i), (i)* cannot possibly hold with factors of 2^{n_0} . The induction begins with such an n_0 .

(23) and (24) follow directly from (i), (i)*. Moreover $\mathcal{M}_k(T) \leq c_0 2^{n+1}$ and $\mathcal{E}_k(T) \leq 2^{n+1}$ for all $k \in \{1, 2, 3\}$, because all tiles of T were available throughout the selection of $\mathcal{F}_{n,i,k}$, yet T was not selected. Therefore T did not satisfy the selection criteria (i), (i)* at stage $n + 1$. (22) follows by Cauchy-Schwarz.

This purely formal discussion applies to any expression $\sum_{P \in \mathcal{S}} |I_P|^{-1/2} \prod_{k=1}^3 |a_k(P)|$; the actual meaning of the coefficients $a_k = \langle f_k, \phi_{P^k} \rangle$ has not yet been exploited.

6. COUNTING MULTITREES

The total sum is the sum of the contributions of all T , and we have a bound⁽⁸⁾ of $2^{3n} |I_{\text{top}}(t)|$ for each T , so it would suffice to have a suitable bound for

$$\boxed{\text{the weighted count } \sum_T |I_{\text{top}(T)}| \text{ of the number of multitrees}}$$

in each family $\mathcal{F}_{n,i,j}$. Here an estimate with genuine content must finally be established.

Since an upper bound for $\sum_T |I_{\text{top}(T)}|$ is required, it is advantageous to place as many tiles as possible into each multitree, consistent with the selection criteria (i), (i)*. This motivates the maximality criterion (ii).

6.1. A small reduction

For each $j \in \{1, 2, 3\}$ let $E_j \subset \mathbb{R}$ be an arbitrary measurable set with $|E_j| < \infty$. Let f_j be any function which is $\equiv 0$ on $\mathbb{R} \setminus E_j$, and $|f_j(x)| \leq 1$ for all $x \in E_j$. The theorem is a consequence of the following inequality via a simple interpolation argument⁽⁹⁾.

PROPOSITION 6.1. — *There exists $C < \infty$ such that for any finite set \mathcal{S} of multitiles, sets E_j , and functions f_j ,*

$$(25) \quad \sum_{P \in \mathcal{S}} |I_P|^{-1/2} \prod_{k=1}^3 |\langle f_k, \phi_{P^k} \rangle| \leq C \prod_{k=1}^3 |E_k|^{1/p_k}$$

for any exponents $p_k \in (1, \infty)$ satisfying $\sum_k p_k^{-1} = 1$.

⁽⁸⁾An important alternative bound is also available; see Proposition 7.2.

⁽⁹⁾Any function can be decomposed as $f = \sum_{m=-\infty}^{\infty} 2^m f_m$ where the sets f_m have pairwise disjoint supports and satisfy $\|f_m\|_{L^\infty} \leq 1$. Interpolation amounts to making this substitution for f_k for each $k \in \{1, 2, 3\}$ to produce an infinite sum of expressions of the type controlled by (25), and summing the resulting bounds.

6.2. Counting multitrees

Here is the crux of the entire analysis.

LEMMA 6.2 (Counting multitrees). — *For any $n \in \mathbb{Z}$ and $i, j \in \{1, 2, 3\}$, for any measurable sets E_1, E_2, E_3 , the multitrees selected by the sorting algorithm satisfy*

$$(26) \quad \sum_{T \in \mathcal{F}_{n,i,j}} |I_{\mathbf{top}(T)}| \leq C2^{-2n}|E_j|.$$

We will discuss only the case $j = i$ in detail; see §7.1 for a brief discussion of the case $j \neq i$. The proof combines structural information built into the sorting algorithm with almost-orthogonality of the rank one operators $f \mapsto \langle f, \phi_{P^k} \rangle \phi_{P^k}$ associated to collections of nonoverlapping (scalar) tiles P . This structural information is:

FACT 6.3. — *The tiles $\{\mathbf{top}(T)^i : T \in \mathcal{F}_{n,i,i}\}$ do not overlap.*

Proof. — If not, there exist two distinct i -towers $T, T' \in \mathcal{F}_{n,i,i}$ satisfying $\mathbf{top}(T)^i < \mathbf{top}(T')^i$. If T' was chosen before T , we reach a contradiction because $T' \cup \{\mathbf{top}(T)\}$ is an i -tower which properly contains T' , contradicting the maximality criterion (ii) for T' . If T was chosen before T' then (ii) is again contradicted, since $T \cup \{\mathbf{top}(T')\}$ is an i -tower (whose top is $\mathbf{top}(T')$), which properly contains T . \square

By the sorting algorithm, $|I_{\mathbf{top}(T)}| \leq 2^{-2n} |\langle f_i, \phi_{\mathbf{top}(T)^i} \rangle|^2 \leq 4|I_{\mathbf{top}(T)}|$ for all $T \in \mathcal{F}_{n,i,i}$, uniformly in n, i and in $\{f_k\}, \{E_k\}$. Since these tiles $\mathbf{top}(T)^i$ are nonoverlapping and $|f_i| \leq \chi_{E_i}$, the almost-orthogonality lemma gives

$$\sum_{T \in \mathcal{F}_{n,i,i}} |I_{\mathbf{top}(T)}| \leq C2^{-2n} \sum_{T \in \mathcal{F}_{n,i,i}} |\langle f_i, \phi_{\mathbf{top}(T)^i} \rangle|^2 \leq C2^{-2n} \|f_i\|_{L^2}^2 \leq C2^{-2n}|E_i|,$$

establishing Lemma 6.2. \square

6.3. Summation with respect to n

By the bound of Lemma 4.7 in terms of $\|f\|_{L^\infty}$, the selection criteria (i), (i)* for parameters (n, i, j) cannot be satisfied unless n does not exceed a certain finite n_0 , independent of the functions f_k and sets E_k . Therefore $\mathcal{F}_{n,i,j}$ is empty for all larger n . Thus

$$(27) \quad \sum_n \sum_{i,j} \sum_{T \in \mathcal{F}_{n,i,j}} \mathcal{H}_T(f_1, f_2, f_3) \leq C \sum_{n \leq n_0} 2^{3n} 2^{-2n} \sum_{k=1}^3 |E_k| \leq C \sum_{k=1}^3 |E_k|.$$

In the case where all three sets E_j have comparable measures, this is the bound of Proposition 6.1. Although it remains to treat the general case, this suffices to exhibit essential elements of the analysis⁽¹⁰⁾: localized Fourier decomposition, the connection between phase space disjointness and orthogonality, sorting of tiles into towers, the

⁽¹⁰⁾It also fully proves a nontrivial result, an L^1 inequality for $BH(f_1, f_2)$ for arbitrary L^∞ functions f_j supported on sets of boundedly finite measures.

relation between towers and trees, the key role of the weighted count of all resulting multitrees, and the role of orthogonality in establishing that count.

7. REFINEMENT

In this more technical section we briefly discuss two steps omitted above. The first is the counting of multitrees of type (n, i, j) for $j \neq i$. The second is the refinement of the above argument to replace the crude bound $\sum_{k=1}^3 |E_k|$ by $\prod_{k=1}^3 |E_k|^{1/p_k}$.

7.1. Counting multitrees of type (n, i, j) with $j \neq i$

The total “number” $\sum_{T \in \mathcal{F}_{n,i,j}} |I_{\text{top}(T)}|$ of multitrees in $\mathcal{F}_{n,i,j}$ is likewise $\leq C2^{-2n}|E_j|$. The proof relies on the following variant of the almost-orthogonality Lemma 3.1. A collection \mathcal{F} of trees is said to be *strongly disjoint* if for any $T \neq T' \in \mathcal{F}$, any $P \in T$ and any $Q \in T'$, if $\omega_Q \supset \omega_P$ then not only must $I_Q \cap I_P = \emptyset$, but furthermore $I_Q \cap I_{\text{top}(T)} = \emptyset$.

LEMMA 7.1. — *Let $f \in L^2$ and $\lambda > 0$ be arbitrary. Let \mathcal{F} be a finite collection of strongly disjoint trees satisfying $\lambda \leq \text{energy}(T) \leq 2\lambda$ for every $T \in \mathcal{F}$. Then*

$$(28) \quad \sum_{T \in \mathcal{F}} |I_{\text{top}(T)}| \leq C\lambda^{-2} \|f\|_{L^2}^2.$$

There is a localization in the same spirit as (15).

With this lemma in hand, the reasoning for the case $j \neq i$ is parallel to that for $j = i$, though a bit more complicated. It is almost but not quite true that $\{T^j : T \in \mathcal{F}_{n,i,j}\}$, where $T^j = \{P^j : P \in T\}$, is a strongly disjoint family of trees for any fixed n, i, j with $j \neq i$. See [21] for details.

7.2. Classical trilinear bound

So far, the analysis has relied entirely on L^2 estimates, but genuine L^p inequalities for linear operators do come into play. L^p estimates are available because expressions \mathcal{H}_T associated to multitrees T are subsumed⁽¹¹⁾ by the theory of singular integral operators, as developed by Coifman and Meyer [9],[10]. If $f_k \in L^\infty$ then such a sum can be rewritten (with $\{1, 2, 3\} = \{i, j, k\}$) as $\langle \mathcal{T}f_i, f_j \rangle$ where \mathcal{T} is a classical singular integral operator, associated to a kernel K which satisfies (3) with a constant C proportional to $\|f_k\|_{L^\infty}$.

Let E_j be measurable sets satisfying $|E_j| < \infty$, and let f_j be measurable functions supported on E_j satisfying $|f_j(x)| \leq 1$ for almost every $x \in E_j$ and $f_j(x) = 0$ for $x \notin E_j$.

⁽¹¹⁾More exactly, these expressions are *modulated* singular integral operators; the associated kernels K are multiplied by factors $e^{i(x-y)\xi}$ for some arbitrary ξ .

PROPOSITION 7.2. — *Let T be any multitree. For all 3-tuples of exponents $p_k \in (1, \infty]$ satisfying the scaling relation $\sum_k p_k^{-1} = 1$ with at most one exponent equal to ∞ ,*

$$(29) \quad \sum_{P \in T} |I_P|^{-1/2} \prod_{k=1}^3 |\langle f_k, \phi_{P^k} \rangle| \leq C \prod_{k=1}^3 |E_k|^{1/p_k}.$$

The constant C depends on the exponents but not on the sets, functions, tower, or index l . This is of course a very particular case of Proposition 6.1.

As is often the case in this subject, a localized version is also available: $|E_k|$ can be replaced by $\int_{E_k} (1 + \frac{\text{distance}(x, I_{\text{top}(T)})}{|I_{\text{top}(T)}|})^{-2} dx$.

7.3. An alternative bound

Consider for simplicity only the case $j = i$. If $\{i, l, m\} = \{1, 2, 3\}$ then using (29) in place of the mass-energy bound for each multitree leads to an alternative bound

$$(30) \quad \sum_{T \in \mathcal{F}_{n,i,i}} \mathcal{H}_T(f_1, f_2, f_3) \leq C_\varepsilon 2^{-(1+\varepsilon)n} |E_m|^{1-2\varepsilon} |E_l|^\varepsilon |E_i|^\varepsilon$$

for arbitrarily small $\varepsilon > 0$.

To prove this write $\{1, 2, 3\} = \{i, l, m\}$. Application of the almost-orthogonality lemma as above shows that uniformly for all dyadic intervals J ,

$$(31) \quad \sum_{T \in \mathcal{F}_{n,i,i}: I_{\text{top}(T)} \subset J} |I_{\text{top}(T)}| \leq C 2^{-2n} \int_{E_i} (1 + \frac{\text{distance}(x, J)}{|J|})^{-2} dx \leq C 2^{-2n} |J|.$$

The *John-Nirenberg lemma* says that this “self-similar” inequality implies a stronger version of itself, to the effect that $\sum_{T \in \mathcal{F}_{n,i,i}} \chi_{I_{\text{top}(T)}}$ is nearly a bounded function:

$$(32) \quad \int_J (\sum_{T \in \mathcal{F}_{n,i,i}} \chi_{I_{\text{top}(T)}})^r \leq C_r 2^{-2nr} |J|$$

uniformly for all dyadic intervals J , for any finite exponent r .

Write $\{1, 2, 3\} = \{i, l, m\}$. Concerning the contributions of f_m, f_l , classical Calderón-Zygmund theory almost gives

$$(33) \quad \sum_{P \in T} |\langle f_m, \phi_{P^m} \rangle| |\langle f_l, \phi_{P^l} \rangle| \leq C |E_l \cap I_{\text{top}(T)}|^{1/p_l} |E_m \cap I_{\text{top}(T)}|^{1/p_m}$$

for any exponents in $(1, \infty)$ satisfying $p_l^{-1} + p_m^{-1} = 1$; in particular, p_m can be taken to be arbitrarily close to 1. This inequality is merely almost true; $|E_j \cap I_{\text{top}(T)}|$ must be replaced by $\int_{E_j} (1 + \frac{\text{distance}(x, I)}{|I|})^{-2} dx$, where $I = I_{\text{top}(T)}$. This sort of routine complication is controlled satisfactorily by the Hardy-Littlewood maximal function.

Consider all intervals J which are maximal among the collection $\{I_{\text{top}(T)} : T \in \mathcal{F}_{n,i,i}\}$ with respect to set inclusion. Combining (33) with (32) and Hölder’s inequality for each interval J , summing over J , and finally exploiting the inclusions $J \subset \{x : M(\chi_{E_i})(x) \geq c 2^{2n}\}$ leads to (30). Details are left to the experts.

(30) is unfavorable for n near $-\infty$, but the bound $2^n|E_i|$ is still favorable for all sufficiently negative n . To conclude the argument, modify the selection algorithm by replacing the thresholds 2^n in criteria (i), (i)* by 2^{na_j} where $a_j > 0$ are parameters. Adjusting these appropriately, depending on the relative sizes of $|E_i|, |E_m|, |E_l|$, makes the minimum of our two bounds favorable in all cases, and permits summation of the infinite series over n .

8. PROOF OF THE ALMOST-ORTHOGONALITY LEMMA

I have emphasized the leading role played by L^2 arguments based on almost-orthogonality. The required almost-orthogonality lemma has a relatively simple proof. First, a preliminary fact: Define $h_P(x) = |I_P|^{-1}(1 + \frac{\text{distance}(x, I_P)}{|I_P|})^{-2}$. These functions belong to $L^1(\mathbb{R})$ uniformly in P . If $|I_Q| \leq |I_P|$ then

$$|\langle \phi_P, \phi_Q \rangle| \leq C|I_P|^{1/2}|I_Q|^{-1/2} \int_{I_Q} h_P.$$

$h_P(x)$ decays rapidly as x moves away from P , on a scale comparable to $|I_P|$.

Proof of Lemma 3.1. — Set $\beta_P = \langle f, \phi_P \rangle$ and $X = \sum_{P \in \mathcal{F}} |\beta_P|^2$. Then

$$X = \sum_P \langle f, \phi_P \rangle \langle \phi_P, f \rangle = \langle \sum_P \beta_P \phi_P, f \rangle \leq \|f\|_{L^2} \|\sum_P \beta_P \phi_P\|_{L^2}.$$

Now

$$\|\sum_P \beta_P \phi_P\|_{L^2}^2 \leq 2 \sum_P \sum_{Q: \omega_P \subset \omega_Q} |\beta_P \beta_Q \langle \phi_P, \phi_Q \rangle|$$

because $\langle \phi_P, \phi_Q \rangle = 0$ unless ω_P, ω_Q overlap, and dyadic intervals cannot overlap unless one contains the other. Thus

$$\begin{aligned} \|\sum_P \beta_P \phi_P\|_{L^2}^2 &\leq 8\lambda^2 \sum_P |I_P|^{1/2} \sum_{Q: \omega_P \subset \omega_Q} |I_Q|^{1/2} |\langle \phi_P, \phi_Q \rangle| \\ &\leq C\lambda^2 \sum_P |I_P| \sum_{Q: \omega_P \subset \omega_Q} \int_{I_Q} h_P \leq C\lambda^2 \sum_P |I_P| \int_{\mathbb{R}} h_P \leq C \sum_P |\beta_P|^2 \end{aligned}$$

since the collection of all intervals I_Q such that $Q \in \mathcal{F}$ and $\omega_P \subset \omega_Q$ is nonoverlapping by hypothesis. Thus $X^2 \leq C\|f\|_{L^2}^2 X$. □

9. CARLESON’S MAXIMAL OPERATOR, ACCORDING TO LACEY AND THIELE

Here I give only a brief outline of a rather condensed treatment [23], in order to exhibit the parallel between the analyses of the bilinear Hilbert transform and Carleson’s maximal operator.

Expression in terms of localized Fourier coefficients. For any dyadic interval ω , denote by ω^b and ω^\sharp the left and right halves of ω , respectively. Likewise to any tile $P = I \times \omega$ are associated the two semitiles $P^b = I \times \omega^b$ and $P^\sharp = I \times \omega^\sharp$.

Let functions ϕ_P be as above, with the single change that $\widehat{\phi_P}$ is supported in ω_P^b . The linearized Carleson operator $\int_{\mathbb{R}} e^{iN(x)t} f(x-t) t^{-1} dt$ is modeled by sums

$$(34) \quad \mathcal{C}^{**}(f)(x) = \sum_P \langle f, \phi_P \rangle \phi_P(x) \chi(x, P)$$

where $\chi(x, P) = 1$ if $N(x)$ lies in ω_P^\sharp , and $= 0$ otherwise. In fact, this operator can be realized as a limit of averages of such models. The goal is an inequality $|\langle \mathcal{C}^{**}(f), \chi_E \rangle| \leq C \|f\|_{L^2} |E|^{1/2}$ for all measurable sets E .

Towers. A tower T is a set of tiles for which there exists some tile $\mathbf{top}(T)$ such that for every $P \in T$, $P \leq \mathbf{top}(T)$; in this discussion $\mathbf{top}(T)$ is *not* required to be an element of T . \sharp -towers and b -towers are defined in the same way, with the requirements $P^\sharp \leq \mathbf{top}(T)^\sharp$ and $P^b \leq \mathbf{top}(T)^b$, respectively.

There are two fundamental sources of orthogonality. Firstly, if T is a b -tower, then for any $P, Q \in T$ satisfying $|I_P| \neq |I_Q|$, the intervals $\omega_P^\sharp, \omega_Q^\sharp$ do not overlap⁽¹²⁾. Therefore the sets $\{x \in E : N(x) \in \omega_P^\sharp\}$ and $\{x \in E : N(x) \in \omega_Q^\sharp\}$ are disjoint. Secondly, any \sharp -tower T is a b -tree: whenever $P, Q \in T$ satisfy $|I_P| \neq |I_Q|$, ω_P^b and ω_Q^b do not overlap. Therefore $\langle \phi_P, \phi_Q \rangle = 0$.

The energy of a tower T is

$$(35) \quad \mathcal{E}(T) = \sup_{T' \subset T} |I_{\mathbf{top}(T')}|^{-1/2} \left(\sum_{P \in T'} |\langle f, \phi_P \rangle|^2 \right)^{1/2},$$

where the supremum is taken over all \sharp -towers $T' \subset T$. The mass of a tower is

$$(36) \quad \mathcal{M}(T) = \sup_{P \in T} \sup_{Q \geq P} |I_Q|^{-1} \int_{\{x \in E : N(x) \in \omega_Q\}} (1 + |I_Q|^{-1} \text{distance}(x, I_Q))^{-2} dx.$$

Classical mass-energy bound for a single tower. In the analysis of the bilinear Hilbert transform, the contribution of a single multitree was estimated by a trivial mass-energy bound. While there is a mass-energy bound here, it is not trivial. For any tower,

$$(37) \quad \sum_{P \in T} |\langle f, \phi_P \rangle| \cdot |\langle \phi_P, \chi_{N(x) \in \omega_P^\sharp} \chi_E \rangle| \leq C \mathcal{E}(T) \mathcal{M}(T) |I_{\mathbf{top}(T)}|.$$

There is also a very simple alternative bound $\mathcal{M}(T) \leq C < \infty$, uniformly for all sets E , because $\chi_E \in L^\infty$ with norm 1.

⁽¹²⁾This is correct under the usual proviso that technicalities concerning dyadic intervals have been appropriately dealt with.

(37) is roughly on the level of Proposition 7.2, and is proved as follows. Any \sharp -tower is a \flat -tree, and the associated sum is essentially a (frequency-modulated generalization of a) truncated singular integral operator

$$(38) \quad \int_{|x-y| \geq \varepsilon(x)} K(x, y) f(y) dy$$

for some function $\varepsilon(x)$ and Calderón-Zygmund kernel K . Such truncations are basic objects in the classical theory. On the other hand, any \flat -tower enjoys the first form of orthogonality described above, and this leads quite easily to the upper bound (37) for its contribution.

Sorting algorithm. As in the analysis of BH , a selection algorithm partitions any finite set of tiles by selecting certain collections $\mathcal{F}_{n,\sharp}$ of towers, for each integer n and index $\sharp \in \{\flat, \sharp\}$. For $\mathcal{F}_{n,\flat}$ the main selection criterion is that $\mathcal{M}(T) \geq 2^{2n}$, while for $\mathcal{F}_{n,\sharp}$ it is that $\mathcal{E}(T) \geq 2^n$. In both cases T is required to be maximal, and for $\mathcal{F}_{n,\sharp}$ there is an analogue of criterion (iii)*.

Counting the towers. The key point is again an upper bound for the weighted number of trees in $\mathcal{F}_{n,\sharp}$.

LEMMA 9.1. — *Uniformly for all functions f and all measurable sets $E \subset \mathbb{R}$ satisfying $\|f\|_{L^2} \leq 1$ and $|E| \leq 1$,*

$$(39) \quad \sum_{T \in \mathcal{F}_{n,\sharp} \cup \mathcal{F}_{n,\flat}} |I_{\text{top}(T)}| \leq C2^{-2n}.$$

For $\mathcal{F}_{n,\flat}$, (39) is essentially a bound for the Hardy-Littlewood maximal function of χ_E . The proof for $\mathcal{F}_{n,\sharp}$ is essentially the same as that of the counting bound for $\mathcal{F}_{n,i,j}$ with $i \neq j$ in the bilinear Hilbert transform analysis; the primary ingredient is the almost-orthogonality lemma for strongly disjoint trees, Lemma 7.1. The strong disjointness of the collection $\mathcal{F}_{n,\sharp}$ of trees is a consequence of the selection algorithm.

In this argument there is no *a priori* upper bound on n , but summation over all $n \in \mathbb{Z}$ yields the desired uniform upper bound anyway (the alternative bound $\mathcal{M}(T) \leq C < \infty$ is used for $n \geq 0$).

10. OPEN PROBLEMS

Higher-degree multilinear operators. The bilinear Hilbert transform can be generalized to

$$(40) \quad T(f_1, \dots, f_m)(x) = \int_{\mathbb{R}} \prod_{k=1}^m f_k(x - \alpha_k t) t^{-1} dt$$

where the α_j are pairwise distinct and nonzero. The analysis outlined above fails to apply to this operator for $m \geq 3$, and it is an open problem whether this formal expression has any meaning for functions in appropriate spaces; scaling dictates that

the natural estimate would be $\|T(f_1, f_2, f_3)\|_{L^q} \leq C \prod \|f_k\|_{L^{p_k}}$ where $q^{-1} = \sum_k p_k^{-1}$. In fact, for $m \geq 4$, when this operator is expanded in terms of localized Fourier coefficients, the resulting sum actually fails to converge absolutely. See [24] and the references cited there for some positive results in this direction.

I mention in passing that for nonsingular expressions $T(f_1, \dots, f_m)(x) = \int_{-1}^1 \prod_{k=1}^m f_k(x - \alpha_k t) dt$, with no singular factor t^{-1} , there are interesting questions (for $m \geq 3$) [7] which make contact with work of Bourgain [1] and of Katz-Tao [19] on a problem in additive number theory related to the Kakeya problem.

Scattering for one-dimensional Dirac operators. Associated to any potential $V : \mathbb{R} \rightarrow \mathbb{C}$ is the Dirac operator $D_V = \begin{pmatrix} -i\frac{d}{dx} & \bar{V} \\ V & i\frac{d}{dx} \end{pmatrix}$. If V has compact support then for each $\lambda \in \mathbb{R}$ there exist unique scattering coefficients $a(\lambda), b(\lambda)$ for which there exists a solution $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ of the generalized eigenfunction equation $D_V u = \lambda u$ taking the form $u(x) \equiv \begin{pmatrix} e^{i\lambda x} \\ 0 \end{pmatrix}$ as $x \rightarrow -\infty$ and $u(x) \equiv \begin{pmatrix} a(\lambda)e^{i\lambda x} \\ b(\lambda)e^{-i\lambda x} \end{pmatrix}$ as $x \rightarrow +\infty$.

For potentials $V \in L^1(\mathbb{R})$, it is elementary that there still exist solutions with these asymptotics, for every λ . For $V \in L^p$ for $1 < p < 2$, this continues to hold for Lebesgue-almost every λ . However, there are indications [16] that the natural class of potentials is $V \in L^2$.

For any x , $u(x)$ can be written in the form $u(x) = \begin{pmatrix} a(x, \lambda)e^{i\lambda x} \\ b(x, \lambda)e^{-i\lambda x} \end{pmatrix}$ where $b(\lambda) = \lim_{x \rightarrow +\infty} b(x, \lambda)$ and likewise for a . The mapping $V \rightarrow \begin{pmatrix} a(x, \lambda) \\ b(x, \lambda) \end{pmatrix}$ is fully nonlinear. A Taylor-type expansion about $V = 0$ gives the linearized expression

$$(41) \quad b(x, \lambda) = i \int_{-\infty}^x e^{2i\lambda y} V(y) dy \quad \text{plus higher-order terms.}$$

The question is whether $\lim_{x \rightarrow +\infty} b(x, \lambda)$ exists for almost every λ , for every $V \in L^2(\mathbb{R})$. By writing $V = \widehat{f}$ we see that the linearization of this problem about $V = 0$ is simply a restatement of Carleson's theorem. Thus the almost everywhere existence of these scattering coefficients, for general $V \in L^2$, is a nonlinear extension of the problem of almost everywhere convergence of Fourier integrals.

For $1 < p < 2$, this was proved [8] using an expansion of the mapping $V \mapsto b(x, \lambda)$ as an infinite sum of multilinear expressions acting on V . However, Muscalu, Tao, and Thiele [25] have shown that even the first nonlinear expression in this series diverges for general $V \in L^2$, and the problem remains open despite an interesting positive result of those authors concerning a related model problem.

There is an almost identical problem for one-dimensional Schrödinger operators, with the added complication that a WKB-type phase correction must be incorporated [8].

Multilinear maximal operators. The bilinear Hilbert transform is to the Hilbert transform as the following multilinear maximal operator is to the Hardy-Littlewood

maximal function, which acts on functions f, g defined on \mathbb{R} :

$$\mathcal{M}(f, g)(x) = \sup_{r>0} r^{-1} \int_{|t|\leq r} |f(x-t)g(x+t)| dt.$$

This operator obviously maps $L^p \times L^q$ to weak L^1 whenever $1 < p, q$ and $p^{-1} + q^{-1} = 1$, but a much deeper result of Lacey is that it maps $L^p \times L^q$ to L^1 . Despite the “positive” nature of the operator, the approach of Lacey relies on a combination of the machinery described in this article, with an inequality developed by Bourgain in his work on averages over subsequences in ergodic theory. Various extensions are known, but most variants of this inequality remain open, including for instance trilinear variants $\sup_{r>0} r^{-1} \int_{|t|\leq r} |f(x-t)g(x+t)h(x+ct)| dt$. See [14],[15] for recent work and references on this topic.

Epilogue. There is a great more to be said, both about subsequent developments to which other authors have made important contributions, and about other types of multilinear operators which appear in contemporary analysis. Space-time inequalities prevent this author from discussing those matters here. I particularly regret having had to give short shrift to other authors’ contributions in order to discuss the themes chosen in the space allotted.

REFERENCES

- [1] J. BOURGAIN – *On the dimension of Kakeya sets and related maximal inequalities*. *Geom. Funct. Anal.* 9 (1999), no. 2, 256–282.
- [2] A.P. CALDERÓN, A. ZYGMUND – *On the existence of certain singular integrals*. *Acta Math.* 88, (1952). 85–139.
- [3] A.P. CALDERÓN – *Uniqueness in the Cauchy problem for partial differential equations*. *Amer. J. Math.* 80 1958 16–36.
- [4] A.P. CALDERÓN – *Commutators of singular integral operators*. *Proc. Nat. Acad. Sci. U.S.A.* 53 1965 1092–1099.
- [5] A.P. CALDERÓN – *Cauchy integrals on Lipschitz curves and related operators*. *Proc. Nat. Acad. Sci. U.S.A.* 74 (1977), no. 4, 1324–1327.
- [6] L. CARLESON – *On convergence and growth of partial sums of Fourier series*, *Acta Math.* 116 (1966), 135–157.
- [7] M. CHRIST – *On certain elementary trilinear operators*. *Math. Res. Lett.* 8 (2001), no. 1-2, 43–56.
- [8] M. CHRIST, A. KISELEV – *WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials*. *J. Funct. Anal.* 179 (2001), no. 2, 426–447.

- [9] R.R. COIFMAN, Y. MEYER – *Au-delà des opérateurs pseudo-différentiels*. Astérisque, 57. Société Mathématique de France, Paris, 1978.
- [10] R.R. COIFMAN, Y. MEYER – *Ondelettes et opérateurs. III. Opérateurs multilinéaires*. Actualités Mathématiques. Hermann, Paris, 1991.
- [11] R.R. COIFMAN, Y. MEYER – *Le théorème de Calderón par les “méthodes de variable réelle”*. Seminar on Harmonic Analysis, 1978–1979, pp. 49–55, Publ. Math. Orsay 79, 7, Univ. Paris XI, Orsay, 1979.
- [12] R.R. COIFMAN, Y. MEYER, A. MCINTOSH – *L’intégrale de Cauchy définit un opérateur borné sur L^2 pour les courbes lipschitziennes*. Ann. of Math. (2) 116 (1982), no. 2, 361–387.
- [13] G. DAVID, J-L. JOURNÉ – *A boundedness criterion for generalized Calderón-Zygmund operators*. Ann. of Math. (2) 120 (1984), no. 2, 371–397.
- [14] C. DEMETER, T. TAO, and C. THIELE – *Maximal multilinear operators*, preprint, math.CA/0510581
- [15] C. DEMETER, M. LACEY, T. TAO, C. THIELE – *Breaking the duality in the return times theorem*, preprint, math.DS/0601455
- [16] P. DEIFT, R. KILLIP – *On the absolutely continuous spectrum of one-dimensional Schrödinger operators with square summable potentials*. Comm. Math. Phys. 203 (1999), no. 2, 341–347.
- [17] C. FEFFERMAN – *Pointwise convergence of Fourier series*, Ann. of Math. 98 (1973), 551–571.
- [18] J.B. GARNETT, P.W. JONES – *BMO from dyadic BMO*. Pacific J. Math. 99 (1982), no. 2, 351–371.
- [19] N.H. KATZ, T. TAO – *Bounds on arithmetic projections, and applications to the Keakeya conjecture*. Math. Res. Lett. 6 (1999), no. 5-6, 625–630.
- [20] A. KNAPP, E.M. STEIN – *Intertwining operators for semisimple groups*. Ann. of Math. (2) 93 (1971), 489–578.
- [21] M.T. LACEY, C.M. THIELE – *L^p bounds for the bilinear Hilbert transform, $p > 2$* , Ann. of Math. 146 (1997), 693–724.
- [22] M.T. LACEY, C.M. THIELE – *On Calderón’s Conjecture*, Ann. of Math. 149 (1999), 475–496.
- [23] M.T. LACEY, C.M. THIELE – *A proof of boundedness of the Carleson operator*, Math. Res. Lett. 7 (2000), no. 4, 361–370.
- [24] C. MUSCALU, T. TAO, C. THIELE – *Multi-linear operators given by singular multipliers*. J. Amer. Math. Soc. 15 (2002), no. 2, 469–496.
- [25] C. MUSCALU, T. TAO, C. THIELE – *A counterexample to a multilinear endpoint question of Christ and Kiselev*. Math. Res. Lett. 10 (2003), no. 2-3, 237–246.
- [26] F. NAZAROV, S. TREIL, A. VOLBERG – *The Tb-theorem on non-homogeneous spaces*. Acta Math. 190 (2003), no. 2, 151–239.

- [27] H. PAJOT – *Capacité analytique et le problème de Painlevé*. (French), [Analytic capacity and the Painlevé problem], Séminaire Bourbaki, Vol. 2003/2004, Astérisque No. 299 (2005), Exp. No. 936, ix, 301–328.

Michael CHRIST

Department of Mathematics

University of California

Berkeley, CA 94720-3840 – U.S.A.

E-mail : `mchrist@math.berkeley.edu`

**SUR LES REPRÉSENTATIONS DE DIMENSION FINIE
DE LA SUPER ALGÈBRE DE LIE $\mathfrak{gl}(m, n)$
[d'après Serganova]**

par **Caroline GRUSON**

INTRODUCTION

Les représentations de dimension finie d'une algèbre de Lie semi-simple complexe de dimension finie sont sommes directes de leurs composantes irréductibles. Si on fixe une sous-algèbre de Cartan \mathfrak{h} dans une algèbre de Lie simple et une représentation de dimension finie V de cette algèbre de Lie, les éléments de \mathfrak{h} agissent de manière simultanément diagonalisable dans V et les valeurs propres s'organisent en des formes linéaires sur \mathfrak{h} . Connaître ces formes linéaires, c'est-à-dire les *poids* de la représentation V , et leurs multiplicités revient à connaître V si V est irréductible sur l'algèbre de Lie initiale : c'est ce qu'on appelle le caractère de V et Hermann Weyl en a donné une formule générale, qui est une somme alternée indexée par le groupe de Weyl de la situation.

Les algèbres de Lie $\mathbb{Z}/2\mathbb{Z}$ -graduées (ou super algèbres de Lie) simples ont été classifiées par Victor Kac en 1977, [8]. Dans [9], Kac étudie la théorie des représentations de ces algèbres et, dans le cas de $\mathfrak{gl}(m, n)$, il introduit les modules de Kac, qui sont de dimension finie et possèdent une formule des caractères analogue à la formule de Weyl. Malheureusement (?), la catégorie \mathcal{F} des représentations de $\mathfrak{gl}(m, n)$ de dimension finie sur lesquelles une sous-algèbre de Cartan de la partie paire $\mathfrak{gl}(m) \times \mathfrak{gl}(n)$ agit de manière diagonalisable n'est pas semi-simple : certaines représentations irréductibles, dites atypiques, ne sont pas scindées dans les représentations où elles apparaissent comme sous-quotients. Les autres représentations irréductibles, dites typiques, sont en fait des modules de Kac et on dispose donc d'une formule des caractères pour eux.

Le problème de trouver le caractère des représentations atypiques de $\mathfrak{gl}(m, n)$, resté ouvert depuis 1977, étudié par Bernstein et Leites dans [2], a été résolu par Vera Serganova en 1996 ([13]). Chaque représentation irréductible atypique L est quotient propre d'un module de Kac V , et si on regarde une suite de Jordan-Hölder de ce module de Kac, on voit apparaître des représentations irréductibles atypiques, qui sont dans la même composante connexe (ou dans le même bloc) de \mathcal{F} que L et V . Le premier théorème (théorème 2.2) explicite ces représentations irréductibles, qui sont en nombre fini. Le second théorème (théorème 2.3) exprime le caractère d'une représentation irréductible atypique comme une somme infinie, il est donc d'un maniement plus compliqué. Ce sont ces résultats qui sont exposés ici.

La difficulté principale du cas $\mathbb{Z}/2\mathbb{Z}$ -gradué provient du fait que le centre de l'algèbre enveloppante $\mathcal{Z}(\mathfrak{gl}(m, n))$ a une structure beaucoup plus compliquée que dans le cas classique. Sergeev, dans [15], donne une description de ce centre dans le cas de $\mathfrak{gl}(m, n)$. Il s'agit des polynômes super symétriques. D'autre part, $\mathcal{Z}(\mathfrak{gl}(m, n))$ intervient dans des questions de géométrie algébrique étudiées par Fulton et Pragacz dans [7]. On remarquera que les difficultés soulevées ressemblent beaucoup aux ennuis que l'on a quand on étudie les représentations des algèbres de Lie simples en caractéristique p (voir [5]). Plusieurs constructions établies par Serganova se font de manière analogue en caractéristique p , même si les conclusions sont assez différentes.

Notons que Jonathan Brundan fait un lien dans [4] entre certaines catégories de représentations du groupe quantique $\mathcal{U}_q(\mathfrak{gl}_\infty)$ et certaines catégories de $\mathfrak{gl}(m, n)$ -modules. Ses résultats répondent aux mêmes questions que ceux de Serganova, avec une approche très différente, mais ne simplifient pas, hélas, les difficultés calculatoires qu'on rencontre.

Le plan de ce texte est le suivant : dans les deux premières parties, je donne les définitions, les notations et les résultats nécessaires pour énoncer les deux principaux théorèmes de [13], et je traite des exemples. Les parties 3 et 4 contiennent des éléments de la démonstration de Vera Serganova, l'induction géométrique pour le paragraphe 3 et un principe de récurrence pour 4 ; elles sont basées sur les articles [12] et [13]. Pour terminer, au paragraphe 5, j'essaie de dépeindre l'état actuel du sujet, en particulier les résultats de [14] et [4].

À partir du paragraphe 3, je suppose que le lecteur a une certaine familiarité avec la théorie des représentations.

J'ai choisi de ne citer en bibliographie que les textes qui m'ont servi directement. L'article de Brundan ([4]) contient une bibliographie très bien faite sur le sujet.

Je suis très reconnaissante envers tous ceux avec lesquels j'ai discuté de ce texte, tout particulièrement Corinne Blondel, Michel Duflo, Laurent Gruson, Séverine Leidwanger, Jean-François Robinet et Jerzy Weyman.

Je remercie très chaleureusement Vera Serganova pour ses réponses claires et rapides à mes questions, pour sa relecture attentive de la présente rédaction et pour sa disponibilité et sa gentillesse pendant la préparation de cet exposé.

1. LE CONTEXTE : DÉFINITIONS ET NOTATIONS

Soient m et n deux entiers supérieurs ou égaux à 1. On définit la super algèbre de Lie $\mathfrak{g} = \mathfrak{gl}(m, n)$; elle est constituée des matrices $(m + n) \times (m + n)$ à coefficients dans \mathbb{C} , munies de la $\mathbb{Z}/2\mathbb{Z}$ -graduation suivante :

$$\mathfrak{g}_0 = \left\{ \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, A \in \mathfrak{gl}(m, \mathbb{C}), D \in \mathfrak{gl}(n, \mathbb{C}) \right\},$$

$$\mathfrak{g}_1 = \left\{ \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}, B \in \text{Hom}(\mathbb{C}^n, \mathbb{C}^m), C \in \text{Hom}(\mathbb{C}^m, \mathbb{C}^n) \right\},$$

et du crochet de Lie défini sur les éléments homogènes u dans $\mathfrak{g}_{p(u)}$ et v dans $\mathfrak{g}_{p(v)}$ par :

$$[u, v] := uv - (-1)^{p(u)p(v)}vu$$

et prolongé par bilinéarité.

Ce crochet vérifie les axiomes des crochets de super algèbres de Lie (super anti-symétrie et identité de Jacobi graduée : pour tous u, v, w homogènes dans \mathfrak{g} , on a $[u, [v, w]] + (-1)^{p(u)(p(v)+p(w))}[v, [w, u]] + (-1)^{p(w)(p(u)+p(v))}[w, [u, v]] = 0$).

Le contexte étant $\mathbb{Z}/2\mathbb{Z}$ -gradué, on utilise une indéterminée ε , de carré 1, pour différencier les parties paires et impaires.

DÉFINITION 1.1. — Soit $V = V_0 \oplus V_1$ un espace vectoriel $\mathbb{Z}/2\mathbb{Z}$ -gradué, de (super) dimension $k + \varepsilon l$ (i.e. la dimension de la partie paire est k et la dimension de la partie impaire est l) ; on notera $\mathfrak{gl}(V)$ la super algèbre de Lie des endomorphismes gradués de V (qui est isomorphe à $\mathfrak{gl}(k, l)$ par le choix d'une base homogène de V).

On dira que V est un **\mathfrak{g} -module**, ou une **représentation** de \mathfrak{g} , si on s'est donné un morphisme de super algèbres de Lie $\rho : \mathfrak{g} \rightarrow \mathfrak{gl}(V)$.

On dira que la représentation V est **irréductible** ou **simple** si elle ne contient aucune sous-représentation non triviale.

Soit \mathfrak{h} la sous-algèbre de Cartan de \mathfrak{g}_0 constituée des matrices diagonales (remarquons que \mathfrak{h} contient le centre de dimension 1 de \mathfrak{g}).

Racines de \mathfrak{g}

L'ensemble des racines de \mathfrak{g} est par définition constitué de l'ensemble Δ_0 des racines (au sens de [3]) de \mathfrak{g}_0 et de l'ensemble Δ_1 des poids de \mathfrak{g}_1 , vu comme \mathfrak{g}_0 -module.

Remarquons que l'on peut munir \mathfrak{g} d'une graduation sur \mathbb{Z} , compatible avec la structure de super algèbre de Lie, en posant :

$$\begin{aligned} \mathfrak{g}^0 &= \mathfrak{g}_0, \\ \mathfrak{g}^{+1} &= \left\{ \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}, B \in \text{Hom}(\mathbb{C}^n, \mathbb{C}^m) \right\}, \\ \mathfrak{g}^{-1} &= \left\{ \begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}, C \in \text{Hom}(\mathbb{C}^m, \mathbb{C}^n) \right\}. \end{aligned}$$

Notons \mathfrak{b}_0 la sous-algèbre de Borel de \mathfrak{g}_0 formée des matrices triangulaires supérieures par blocs ; alors $\mathfrak{b} := \mathfrak{b}_0 \oplus \mathfrak{g}^{+1}$ est une *sous-algèbre de Borel* de \mathfrak{g} ce qui permet, comme dans le cas classique, de séparer les racines de \mathfrak{g} en racines positives et négatives.

On note $(E_{i,j})_{1 \leq i, j \leq m+n}$ la base usuelle de \mathfrak{g} et $(E_{i,j}^*)_{1 \leq i, j \leq m+n}$ la base duale. On appelle *super trace* la forme linéaire $Str = E_{1,1}^* + \dots + E_{m,m}^* - E_{m+1,m+1}^* - \dots - E_{m+n,m+n}^*$. Elle est \mathfrak{g} -invariante, et permet de définir une forme bilinéaire \mathfrak{g} -invariante non dégénérée sur \mathfrak{g} par $B(u, v) := Str(v.u)$ pour u et v dans \mathfrak{g} . Sa restriction à \mathfrak{h} est encore non dégénérée et on note $\langle ; \rangle$ le produit scalaire sur \mathfrak{h}^* qui s'en déduit.

Si on note ε_i (resp. δ_j) l'élément $E_{i,i}^*$ (resp. $E_{m+j,m+j}^*$) de \mathfrak{g}^* , alors les ε_i , $1 \leq i \leq m$, et les δ_j , $1 \leq j \leq n$, forment une base de \mathfrak{h}^* et on a $\langle \varepsilon_i, \varepsilon_j \rangle = \delta_{ij}$, $\langle \delta_i, \delta_j \rangle = -\delta_{ij}$ et $\langle \varepsilon_i, \delta_j \rangle = 0$. On pose :

$$\Delta_0^+ = \{(\varepsilon_i - \varepsilon_j), 1 \leq i < j \leq m\} \cup \{(\delta_i - \delta_j), 1 \leq i < j \leq n\},$$

$$\Delta_1^+ = \{(\varepsilon_i - \delta_j), 1 \leq i \leq m, 1 \leq j \leq n\},$$

avec $\Delta_0 = \Delta_0^+ \cup -\Delta_0^+$ et $\Delta_1 = \Delta_1^+ \cup -\Delta_1^+$. Remarquons que toutes les racines de Δ_1 , qu'on appelle racines *impaires*, sont isotropes. Les racines de Δ_0 sont dites *paires*.

Enfin, on notera W le groupe de Weyl de \mathfrak{g}_0 , identifié au produit de groupes symétriques $\mathfrak{S}_m \times \mathfrak{S}_n$.

Modules et poids

On étudie ici les \mathfrak{g} -modules dans lesquels la sous-algèbre de Cartan \mathfrak{h} agit de manière diagonalisable. Tout \mathfrak{g} -module est un \mathfrak{g}_0 -module, par restriction. Les \mathfrak{g} -modules de dimension finie ont des *poids*, qui correspondent aux valeurs propres de \mathfrak{h} agissant sur ceux-ci. Ce sont des éléments de \mathfrak{h}^* .

Soit $\lambda \in \mathfrak{h}^*$; on dit que λ est un *poids entier* si les coordonnées de λ dans la base ci-dessus sont entières (ce qui implique que $\langle \lambda, \alpha \rangle \in \mathbb{Z}, \forall \alpha \in \Delta_0$), et on note P l'ensemble des poids entiers.

On pose $\rho = \frac{1}{2}(\sum_{\alpha \in \Delta_0^+} \alpha - \sum_{\alpha \in \Delta_1^+} \alpha)$. On a

$$\rho = \frac{1}{2}((m - n - 1)\varepsilon_1 + (m - n - 3)\varepsilon_2 + \dots + (-m - n + 1)\varepsilon_m +$$

$$+ (m + n - 1)\delta_1 + (m + n - 3)\delta_2 + \dots + (m - n + 1)\delta_n)$$

et, pour toute racine paire α , on pose $\alpha^\vee = \frac{2\alpha}{\langle \alpha, \alpha \rangle}$. On dira que λ est un *poids dominant* si $\langle \lambda, \alpha^\vee \rangle \in \mathbb{N}, \forall \alpha \in \Delta_0^+$. On note P^+ l'ensemble des poids dominants.

Si $\lambda = a_1\varepsilon_1 + \dots + a_m\varepsilon_m + b_1\delta_1 + \dots + b_n\delta_n$, on a :

$$\lambda \in P^+ \Leftrightarrow \lambda \in P \text{ et } a_1 \geq \dots \geq a_m, b_1 \geq \dots \geq b_n.$$

De plus, si $\lambda \in P^+$, on a alors $\lambda + \rho = x_1\varepsilon_1 + \dots + x_m\varepsilon_m + y_1\delta_1 + \dots + y_n\delta_n$ avec $x_1 > \dots > x_m, x_{i+1} - x_i \in \mathbb{Z}$ et $y_1 > \dots > y_n, y_{i+1} - y_i \in \mathbb{Z}$.

On dispose par ailleurs d'un *ordre partiel* sur les poids entiers défini par : si λ et μ sont dans P , on dit que $\lambda \geq \mu$ si $\lambda - \mu$ peut s'écrire comme une combinaison linéaire à coefficients dans $\mathbb{N}, \sum_{\alpha_k \in \Delta_0^+ \cup \Delta_1^+} n_k \alpha_k$.

Remarque 1.2. — La catégorie des \mathfrak{g} -modules \mathfrak{h} -diagonalisables de dimension finie, \mathcal{F} , munie des morphismes de \mathfrak{g} -modules de degré 0, est abélienne mais n'est pas semi-simple, contrairement à la catégorie correspondante pour \mathfrak{g}_0 . D'autre part, un \mathfrak{g} -module irréductible V de \mathcal{F} a un *plus haut poids*, c'est-à-dire qu'il existe $\lambda \in P^+$ tel que, pour tout poids μ de V , on a $\lambda \geq \mu$.

Algèbre enveloppante et caractère infinitésimal

Comme dans le cas classique (voir [6]), on dispose de l'algèbre enveloppante de \mathfrak{g} , $\mathcal{U}(\mathfrak{g})$, d'un théorème de Poincaré-Birkhoff-Witt, et du centre $\mathcal{Z}(\mathfrak{g})$ de $\mathcal{U}(\mathfrak{g})$. Contrairement à ce dont nous avons l'habitude, $\mathcal{Z}(\mathfrak{g})$ n'est plus un anneau de polynômes. L'article [15] donne une description complète de la structure de $\mathcal{Z}(\mathfrak{g})$. Par le lemme de Schur, si V est un \mathfrak{g} -module irréductible de dimension finie, tout élément z de $\mathcal{Z}(\mathfrak{g})$

agit par un scalaire $z(V)$ sur V et on peut donc définir le *caractère infinitésimal* associé à V sur $\mathcal{Z}(\mathfrak{g})$:

$$\chi_V : \mathcal{Z}(\mathfrak{g}) \rightarrow \mathbb{C}, z \mapsto z(V).$$

De même, si V est un module indécomposable avec plus haut poids (par exemple un module de Kac ou un module de Verma, voir les définitions dans l’alinéa suivant), tout élément z de $\mathcal{Z}(\mathfrak{g})$ agit par un scalaire sur V et χ_V est bien défini.

Construction de modules de plus haut poids

Soit $\lambda \in P$. On construit le *module de Verma* M_λ de plus haut poids λ : soit \mathbb{C}_λ le caractère de \mathfrak{b} qui étend trivialement le caractère λ de \mathfrak{h} . On pose :

$$M_\lambda = \text{Ind}_{\mathcal{U}(\mathfrak{b})}^{\mathcal{U}(\mathfrak{g})} \mathbb{C}_\lambda.$$

C’est un \mathfrak{g} -module de dimension infinie qui admet un unique sous-module maximal, il est donc indécomposable. On note L_λ son unique quotient simple, c’est le *\mathfrak{g} -module simple de plus haut poids λ* , on démontre ([9]) que L_λ est de dimension finie si et seulement si $\lambda \in P^+$.

Remarquons maintenant que, si $\lambda \in P^+$, alors il existe un \mathfrak{g}_0 -module simple de plus haut poids λ et de dimension finie, que nous noterons $L_\lambda(\mathfrak{g}_0)$. On étend trivialement l’action à \mathfrak{g}^{+1} , ce qui nous donne un $\mathfrak{g}_0 \oplus \mathfrak{g}^{+1}$ -module de dimension finie. On fait enfin une induction en posant :

$$V_\lambda = \text{Ind}_{\mathcal{U}(\mathfrak{g}_0 \oplus \mathfrak{g}^{+1})}^{\mathcal{U}(\mathfrak{g})} L_\lambda(\mathfrak{g}_0).$$

Or l’algèbre enveloppante de \mathfrak{g}^{-1} est isomorphe, comme \mathbb{C} -algèbre, à l’algèbre extérieure de l’espace vectoriel de dimension finie \mathfrak{g}^{-1} . En faisant cette induction, on obtient donc un \mathfrak{g} -module de dimension finie V_λ indécomposable qui s’appelle le *module de Kac* de plus haut poids λ . Le module simple L_λ est un quotient de V_λ .

Modules typiques et atypiques

Rappelons que, dans le cas classique, si \mathfrak{a} est une algèbre de Lie réductive et si on a choisi $\mathfrak{h} \subset \mathfrak{b} \subset \mathfrak{a}$ une sous-algèbre de Cartan contenue dans une sous-algèbre de Borel de \mathfrak{a} , soient λ et μ deux poids entiers ; notons χ_λ le caractère infinitésimal associé au \mathfrak{a} -module simple de plus haut poids λ , W le groupe de Weyl de la situation et ρ la demi-somme des racines positives. On a :

$$\chi_\lambda = \chi_\mu \iff \exists w \in W \text{ t.q. } \mu = w(\lambda + \rho) - \rho.$$

Revenons à $\mathfrak{g} = \mathfrak{gl}(m, n)$, en gardant la notation χ_λ pour le caractère infinitésimal associé au module simple de plus haut poids λ , L_λ .

On introduit une relation d’équivalence sur l’ensemble des poids entiers P :

$$\lambda \sim \mu \iff \chi_\lambda = \chi_\mu.$$

DÉFINITION 1.3. — PROPOSITION - Soit $\lambda \in P^+$.

i) Si pour toute racine α dans Δ_1^+ , on a $\langle \lambda + \rho, \alpha \rangle \neq 0$, alors λ (et le module simple L_λ) est dit **typique**. Pour tout μ , on a alors

$$\chi_\lambda = \chi_\mu \iff \exists w \in W \text{ t.q. } \mu = w(\lambda + \rho) - \rho.$$

ii) Si il existe α dans Δ_1^+ telle que $\langle \lambda + \rho, \alpha \rangle = 0$, alors λ (et le module simple L_λ) est dit **atypique** et on a, pour tout k dans \mathbb{Z} , $\chi_\lambda = \chi_{\lambda+k\alpha}$.

DÉFINITION 1.4. — Soit λ un poids dominant atypique. On appelle **degré d'atypie** de λ et on note $\#\lambda$ le cardinal de l'ensemble $A(\lambda)$ suivant :

$$A(\lambda) := \{\alpha \in \Delta_1^+, \langle \lambda + \rho, \alpha \rangle = 0\}.$$

Remarquons que $A(\lambda)$ est constitué de racines deux à deux orthogonales et indépendantes, ce qui fait que $\#\lambda \leq \inf(m, n)$.

Si λ est typique, l'ensemble $A(\lambda)$ est vide et on pose $\#\lambda = 0$.

PROPOSITION 1.5. — Soit λ un poids dominant atypique, soit μ dans P . On a

$$\begin{aligned} \chi_\lambda = \chi_\mu &\iff \lambda \sim \mu \\ \lambda \sim \mu &\iff \lambda \sim' \mu, \end{aligned}$$

où \sim' est la relation d'équivalence engendrée par :

- i) $\lambda \sim' \mu$ si $\exists w \in W$ t.q. $\mu = w(\lambda + \rho) - \rho$,
 - ii) Soit $\alpha \in \Delta_1^+$ tel que $\langle \lambda + \rho, \alpha \rangle = 0$; alors $\mu = \lambda + \alpha$ vérifie $\lambda \sim' \mu$.
- De plus, si $\lambda \sim' \mu$ et si μ est dominant, alors $\#\lambda = \#\mu$.

Ceci justifie la définition plus générale suivante :

DÉFINITION 1.6. — Soit $\mu \in P$; on dira que μ est atypique de degré d'atypie $\#\mu$ si il existe λ dans P^+ atypique de degré d'atypie $\#\mu$ tel que $\lambda \sim \mu$.

PROPOSITION 1.7. — Soit $\lambda \in P^+$, λ typique. Alors le module de Kac V_λ est égal au module simple L_λ .

Remarque 1.8. — Les modules de Kac étant des modules induits à partir de \mathfrak{g}_0 -modules, on calcule aisément leur décomposition en \mathfrak{h} -modules irréductibles (ce qu'on appelle leur *caractère*), donc le caractère des modules simples typiques est connu ([9]).

Blocs de \mathcal{F}

La catégorie \mathcal{F} se décompose en composantes connexes, les *blocs* de \mathcal{F} . Dans [14], Serganova montre que chaque bloc correspond à une valeur du caractère infinitésimal χ_λ . Les modules typiques forment donc chacun un bloc, et si λ est un poids atypique, tous les modules de dimension finie dont le caractère infinitésimal est χ_λ sont dans le même bloc que le module simple L_λ . Elle montre de plus que si le degré d'atypie de λ est égal à k , alors le bloc est équivalent au bloc maximale atypique de $\mathfrak{gl}(k, k)$ (c'est-à-dire l'unique bloc contenant les modules simples de degré d'atypie égal à k ; l'unicité de ce bloc est montrée dans [14]). Les modules de Kac étant indécomposables, tous les

quotients simples d’une suite de Jordan-Hölder d’un module de Kac V sont dans le même bloc que V .

2. ÉNONCÉ DES RÉSULTATS ET EXEMPLES

Soit λ un poids dominant atypique. On garde les notations du premier paragraphe.

On veut trouver les quotients simples d’une suite de Jordan-Hölder du module de Kac V_λ . On utilise pour ce faire des opérateurs définis dans le \mathbb{Z} -module libre $\mathcal{K}[\mathcal{F}]$ ayant pour base les classes d’isomorphismes $[L_\mu]$ de \mathfrak{g} -modules simples \mathfrak{h} -diagonalisables de dimension finie (c’est le *groupe de Grothendieck* de la catégorie \mathcal{F}). Soit q une indéterminée. Si $V = V_0 \oplus V_1$ est un super espace vectoriel de dimension $m + \varepsilon n$, on note V^Π le super espace vectoriel $(V^\Pi)_0 \oplus (V^\Pi)_1$ avec $(V^\Pi)_0 = V_1$ et $(V^\Pi)_1 = V_0$, et on pose, pour tout V dans \mathcal{F} , $\varepsilon V = V^\Pi$. C’est un opérateur dont le carré vaut l’identité.

Soit \mathcal{H} le $\mathbb{C}[q, \varepsilon]$ -module libre ayant pour base des éléments T_λ , où λ parcourt l’ensemble des poids entiers, P .

DÉFINITION 2.1. — *Soit α une racine impaire positive. On définit un opérateur σ_α sur \mathcal{H} par les axiomes suivants :*

i) $\sigma_\alpha(T_\lambda) = 0$ si $\langle \lambda, \alpha \rangle \neq 0$.

ii) Si $\langle \lambda, \alpha \rangle = 0$,

$$\sigma_\alpha(T_\lambda) = \varepsilon[q^{l(\lambda, \alpha)-1} \sigma_\alpha(T_{\lambda-\alpha})]_+ + \varepsilon q T_{\lambda-\alpha},$$

où $[]_+$ désigne la partie polynomiale en q de degré strictement positif et $l(\lambda, \alpha) = rk(W_{\lambda-\alpha}) - rk(W_{\lambda-\alpha} \cap W_\lambda)$, où $rk(W_\mu)$ désigne le rang (comme groupe de Coxeter) du stabilisateur de μ dans W .

iii) Si $\langle \lambda, \alpha \rangle = 0$ et si λ et $-\alpha$ sont dans la même chambre (dans le même domaine fondamental pour l’action de W), alors $\sigma_\alpha(T_\lambda) = \varepsilon q T_{\lambda-\alpha}$.

On obtient ainsi les σ_α par récurrence. Leur définition est directement liée à la démonstration des théorèmes, comme on le voit en analysant l’énoncé du théorème 4.3 de ce texte. Nous donnerons un peu plus loin un algorithme permettant de les calculer.

On définit ensuite des opérateurs \tilde{s}_α et s_α de $\mathbb{C}[q] \otimes \mathcal{K}[\mathcal{F}]$ et de $\mathcal{K}[\mathcal{F}]$ respectivement : on définit d’abord l’opérateur $\mathbb{C}[q, \varepsilon]$ -linéaire $\Xi : \mathcal{H} \rightarrow \mathbb{C}[q] \otimes \mathcal{K}[\mathcal{F}]$ tel que $\Xi(T_\lambda) = [L_{\tilde{\lambda}-\rho}]$ (où $\tilde{\lambda}$ est un représentant dans P^+ de la W -orbite de λ) si $\tilde{\lambda} - \rho \in P^+$, et $\Xi(T_\lambda) = 0$ si $\tilde{\lambda} - \rho \notin P^+$.

On pose : $\tilde{s}_\alpha[L_\lambda] = \Xi(\sigma_\alpha(T_{\lambda+\rho}))$, l’opérateur s_α étant alors obtenu en évaluant \tilde{s}_α en $q = -1$.

THÉORÈME 2.2. — *On a la relation suivante dans l’anneau de Grothendieck $\mathcal{K}[\mathcal{F}]$:*

$$[V_\lambda] = \left(\prod_{\alpha \in \Delta_1^+}^{\leftarrow} (1 - s_\alpha) \right) [L_\lambda],$$

où l’ordre du produit est compatible avec l’ordre partiel \geq sur Δ_1^+ .

On connaît donc les modules simples qui interviennent dans les suites de Jordan-Hölder des modules de Kac. De plus, on a une égalité de \mathfrak{g}_0 -modules entre la somme directe des modules en question et le module de Kac.

On note $\mathcal{Ch}(\mathcal{F})$ l’anneau des caractères de la catégorie \mathcal{F} : on le considère comme un sous-anneau de $\mathbb{Z}[\varepsilon, \{e^\mu\}_{\mu \in P}]$. En effet, le caractère d’un \mathfrak{g} -module $V = V_0 \oplus V_1$ est $\mathcal{Ch}(V) = \sum (m_i e^{\mu_i} + \varepsilon n_j e^{\nu_j})$ où μ_i parcourt les poids de V_0 (avec multiplicité m_i) et ν_j parcourt les poids de V_1 (avec multiplicité n_j).

On a vu que le caractère du module de Kac V_λ est connu car il s’agit d’un module induit. On a la formule suivante :

$$\mathcal{Ch}(V_\lambda) = \mathcal{Ch}(L_\lambda(\mathfrak{g}_0)) \prod_{\alpha \in \Delta_1^+} (1 + \varepsilon e^{-\alpha}).$$

Soit maintenant l’application $\mathbb{Z}[\varepsilon]$ -linéaire de $\mathcal{K}[\mathcal{F}]$ dans $\mathcal{Ch}(\mathcal{F})$, ψ , telle que $\psi[L_\lambda] := \mathcal{Ch}(V_\lambda)$, qui à une classe d’isomorphisme de modules simples de plus haut poids λ associe le caractère du module de Kac de plus haut poids λ . On a le

THÉORÈME 2.3. — *Soit λ un poids dominant ; on a :*

$$\mathcal{Ch}(L_\lambda) = \psi\left(\overrightarrow{\prod}_{\alpha \in \Delta_1^+} (1 - s_\alpha)^{-1}[L_\lambda]\right),$$

où l’ordre du produit est compatible avec l’ordre partiel \leq sur Δ_1^+ .

Exemples

Le théorème 2.3 correspond à une inversion formelle dans la formule du théorème 2.2 ; la formule fait apparaître des sommes infinies, car $(1 - s_\alpha)^{-1} = 1 + s_\alpha + s_\alpha^2 + \dots$. Elle n’aide donc pas beaucoup au calcul effectif du caractère d’un module simple atypique donné.

Le théorème 2.2 est plus accessible au calcul ; Vera Serganova donne un algorithme pour le calcul des s_α ([13], remarque 2.5) qui se déduit de l’énoncé du théorème. Je vais décrire cet algorithme en utilisant des diagrammes de partitions pour un poids maximale atypique (de degré d’atypie k) dans $\mathfrak{gl}(k, k)$ pour un entier $k \geq 2$. On peut toujours se ramener à cette situation-là d’après le dernier alinéa du paragraphe 1.

Dans les notations précédentes, un tel poids s’écrit $\lambda = a_1 \varepsilon_1 + \dots + a_k \varepsilon_k + b_1 \delta_1 + \dots + b_k \delta_k$ et on peut supposer qu’on a les relations $a_i = -b_{k+1-i}$ pour $1 \leq i \leq k$, de sorte que les racines $\alpha_1, \dots, \alpha_k$ telles que $\langle \lambda + \rho, \alpha_i \rangle = 0$ sont les $\alpha_i := \varepsilon_{k+1-i} - \delta_i$ (on a $\alpha_1 \leq \dots \leq \alpha_k$, ce qui fait que l’ordre du produit dans le théorème 2.2 est $(1 - s_{\alpha_k}) \dots (1 - s_{\alpha_1})$). Pour éviter les lourdeurs de notations, on note $\lambda = (a_1, \dots, a_k)$.

Comme λ est dominant, on a $a_1 \geq a_2 \geq \dots \geq a_k$. Notons D_λ le diagramme obtenu en mettant a_i cases dans la i -ième ligne (avec la notation anglaise). Comme les a_i ne sont pas supposés positifs, on considère qu’on peut ajouter autant de colonnes qu’on veut à gauche du diagramme.

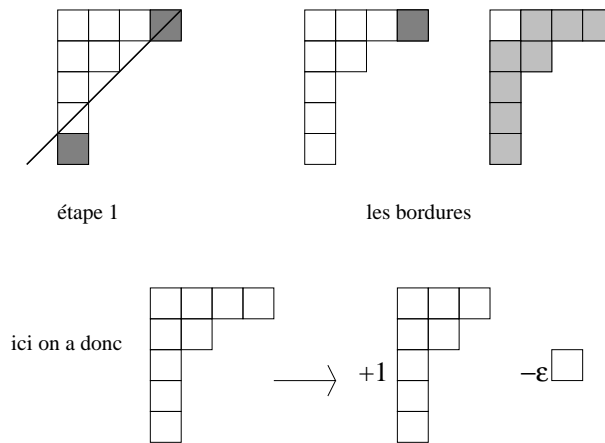
Pour calculer s_{α_k} , (ou plus exactement $-s_{\alpha_k}$) on fait la chose suivante : soit Δ l’antidiagonale issue de la case C_1 en haut à droite de D_λ . On regarde toutes les cases

C_j qui sont à droite (au sens large) de cette antidiagonale, à droite de leur horizontale, en bas de leur verticale (les cases Sud-Est sous l'antidiagonale Δ) et telles que la parallèle à Δ issue de C_j est au-dessous (au sens large) de toutes les cases Sud-Est apparues avant C_j . Elles correspondent à des lignes i_1, \dots, i_r de D_λ , la ligne i_j contenant a_{i_j} cases (dont C_j est la dernière). On en déduit r nouveaux diagrammes $D_{\lambda_1}, \dots, D_{\lambda_r}$, où D_{λ_j} est obtenu en retirant à D_λ la bordure reliant la case C_j à la case C_1 . Cela nous donne r nouveaux poids $\lambda_1, \dots, \lambda_r$ et on calcule

$$-s_{\alpha_k}[L_\lambda] = \sum_{j=1}^r (-\varepsilon)^{a_{i_j} - a_1 + i_j - 1} [L_{\lambda_j}]$$

(le nombre $a_{i_j} - a_1 + i_j - 1$ est la distance de la case C_j à Δ).

Exemple pour le poids $\lambda = (4, 2, 1, 1, 1)$:



Pour calculer $s_{\alpha_{k-i}}$, on refait la même chose en oubliant les $i - 1$ premières lignes de D_λ .

Si par chance λ est tel que, pour tout i , $a_i - a_{i+1} \geq 2$, alors les opérations sur le diagramme sont triviales et on obtient la belle formule suivante :

$$[V_\lambda] = [L_\lambda] + \varepsilon \sum_{1 \leq i \leq k} [L_{\lambda - \alpha_i}] + \sum_{1 \leq i < j \leq k} [L_{\lambda - \alpha_i - \alpha_j}] + \dots + \varepsilon^k [L_{\lambda - \sum_{i=1}^k \alpha_i}].$$

Ces poids sont « loin des murs », ils ont un comportement régulier qu'on peut considérer comme admirable...

Dès qu'on s'approche des murs (i.e. quand un $a_i - a_{i+1}$ est égal à 0 ou 1), les calculs se compliquent fortement, des signes font leur apparition et le nombre de termes peut croître beaucoup.

L'exemple 2.6 de [13] consiste à faire tous les calculs pour le module trivial de $\mathfrak{gl}(3, 3)$, $\lambda = (0, 0, 0)$, on trouve (après des calculs où interviennent 14 modules!!!) un résultat plutôt élégant :

$$[V_{(0,0,0)}] = [L_{(0,0,0)}] + \varepsilon [L_{(0,0,-1)}] + [L_{(0,-2,-2)}] + \varepsilon [L_{(-3,-3,-3)}].$$

Plus généralement, on trouve pour le module trivial de $\mathfrak{gl}(k, k)$:

$$[V_{(0, \dots, 0)}] = \sum_{i=0}^k [L_{-i \sum_{j=1}^i \alpha_j}].$$

3. INDUCTION GÉOMÉTRIQUE

Soit \mathcal{F}^{free} la sous-catégorie pleine de \mathcal{F} constituée des modules qui sont libres comme $\mathcal{U}(\mathfrak{g}^{-1})$ -modules : par construction, les modules de Kac sont dans \mathcal{F}^{free} . On montre facilement que tout module de \mathcal{F}^{free} admet une filtration finie dont les quotients successifs sont des modules de Kac, qui, comme dans une suite de Jordan-Hölder, sont uniquement déterminés à l'ordre près. On démontre de plus que tout module M de \mathcal{F} admet une résolution par des modules de \mathcal{F}^{free} , et cette résolution est de longueur finie si et seulement si $M \in Ob(\mathcal{F}^{free})$. En particulier, les \mathfrak{g} -modules simples L_λ de \mathcal{F} ont une résolution dans \mathcal{F}^{free} qui est de longueur infinie quand λ est un poids atypique. On peut choisir la résolution de L_λ de sorte que le premier terme soit le module de Kac V_λ ; on a ainsi :

$$\dots \rightarrow M_i \rightarrow \dots \rightarrow M_1 \rightarrow V_\lambda \rightarrow L_\lambda \rightarrow 0$$

et de plus (c'est le lemme 1.14 de [13]), en désignant par $[M_i : V_\mu]$ la multiplicité du module de Kac V_μ dans une \mathcal{F}^{free} -filtration de M_i , si $i > 0$, $[M_i : V_\mu] \neq 0$ impose $\mu < \lambda$ et $\chi_\mu = \chi_\lambda$.

Ceci signifie qu'on peut choisir la résolution de L_λ dans le bloc de L_λ .

Serganova introduit un opérateur $\mathbb{Z}[q]$ -linéaire de $\mathcal{K}[\mathcal{F}] \otimes_{\mathbb{Z}} \mathbb{Z}[q]$, $\mathbf{U}(q)$, dont elle donne deux définitions (elle démontre dans [12] qu'elles sont équivalentes).

Première définition de $\mathbf{U}(q)$

Soit \mathfrak{p} la sous-algèbre parabolique maximale de \mathfrak{g} obtenue en annulant toutes les coordonnées non diagonales de la première colonne. On note $\mathfrak{g}(m-1, n)$ la sous super algèbre de Lie de \mathfrak{p} obtenue en annulant toutes les coordonnées non diagonales de la première ligne. C'est une sous super algèbre de Lie de \mathfrak{g} qui est isomorphe à $\mathfrak{gl}(m-1, n) \oplus \mathbb{C}\varepsilon_1$; elle a donc un centre de dimension deux.

Soit t un entier. Soit $M \in Ob(\mathcal{F})$, soit M^t le sous-espace vectoriel de M engendré par les vecteurs de poids μ avec $\langle \mu, \varepsilon_1 \rangle \geq t$: c'est un sous- \mathfrak{p} -module de M et on note $U_t(M)$ le quotient de M par le \mathfrak{g} -module engendré par M^t . Ceci définit un foncteur U_t de la catégorie \mathcal{F} dans elle-même qui est exact à droite ([13], lemme 4.3), ce qui fait qu'on peut considérer la suite $(U_t)_i, i \geq 1$ de ses foncteurs dérivés à gauche.

DÉFINITION 3.1. — *L'opérateur $\mathbf{U}(q)$ associe au module simple de plus haut poids λ dans P^+ la somme formelle*

$$\sum_{i \geq 1} q^i [(U_t)_i(L_\lambda)],$$

où $t = \langle \lambda, \varepsilon_1 \rangle$.

On démontre (cf. [12]) que cette somme définit un polynôme en q sur $\mathbb{Z}[\varepsilon]$.

Remarque 3.2. — On montre que le foncteur U_t est exact dans \mathcal{F}^{free} , ce qui fait que, si $V \in Ob(\mathcal{F})$ et si

$$\dots \rightarrow M_i \rightarrow \dots \rightarrow M_0 \rightarrow V \rightarrow 0$$

est une résolution de V par des modules de \mathcal{F}^{free} , on a $(U_t)_i(V) = H_i(U_t(M_\bullet))$. Dans [13], toutes les notations sont cohomologiques, ce qui rend parfois l'interprétation hasardeuse. J'ai pris la liberté de descendre les indices pour comprendre.

Seconde définition de $U(q)$

Soit (G, \mathcal{O}_G) le super groupe algébrique complexe $GL(m, n)$: sa super algèbre de Lie est \mathfrak{g} ; les points G_0 de G sont les éléments de $GL(m, \mathbb{C}) \times GL(n, \mathbb{C})$ et l'algèbre de Hopf \mathcal{O}_G est isomorphe à l'algèbre (super) symétrique de \mathfrak{g} , localisée en $det|_{\mathfrak{g}_0}$, la restriction du déterminant à la partie paire.

Soit (P, \mathcal{O}_P) le sous super groupe de (G, \mathcal{O}_G) d'algèbre de Lie \mathfrak{p} : c'est le sous super groupe fermé de (G, \mathcal{O}_G) qui fixe le premier vecteur de base de \mathbb{C}^m à homothéties près. On note P_0 les points de P . Comme P est fermé dans G , il existe une structure de super variété algébrique sur le quotient G/P notée (X, \mathcal{O}_X) . Les points de cette super variété sont les points de l'espace projectif $\mathbb{P}^{m-1}(\mathbb{C})$, qu'on notera aussi parfois G_0/P_0 , et les fonctions constituent l'algèbre extérieure de $\mathbb{C}^n \otimes_{\mathcal{O}_{\mathbb{P}^{m-1}}} \mathcal{O}_{\mathbb{P}^{m-1}}(-1)$ où $\mathcal{O}_{\mathbb{P}^{m-1}}$ désigne l'anneau des fonctions régulières (au sens non gradué) sur l'espace projectif $\mathbb{P}^{m-1}(\mathbb{C})$.

Soit V un (P, \mathcal{O}_P) -module de dimension finie (il se dérive en un \mathfrak{p} -module). On construit, de la manière habituelle ([5], I.5.8), un super fibré vectoriel \mathbf{V} sur X , dont on note \mathcal{O}_V le faisceau des sections, sur lequel le super groupe (G, \mathcal{O}_G) agit de manière équivariante. On note $H^i(\mathcal{O}_V)$ le i -ième groupe de cohomologie de ce faisceau ; c'est un (G, \mathcal{O}_G) -module de dimension finie, qui se dérive en un \mathfrak{g} -module de dimension finie que l'on note de la même manière.

PROPOSITION 3.3. — *Soit λ un poids dominant de \mathfrak{g} .*

i) *On note $V_\lambda(\mathfrak{p})$ le $\mathfrak{g}(m-1, n)$ -module de Kac de plus haut poids λ étendu à \mathfrak{p} et V_λ le \mathfrak{g} -module de Kac de plus haut poids λ . On a alors :*

$$H^i(\mathcal{O}_{V_\lambda(\mathfrak{p})^*}) = 0, \quad \forall i > 0,$$

$$H^0(\mathcal{O}_{V_\lambda(\mathfrak{p})^*}) = V_\lambda^*.$$

ii) *On note $L_\lambda(\mathfrak{p})$ le $\mathfrak{g}(m-1, n)$ -module simple de plus haut poids λ étendu à \mathfrak{p} et L_λ le \mathfrak{g} -module simple de plus haut poids λ . On a alors :*

$$H^i(\mathcal{O}_{L_\lambda(\mathfrak{p})^*}) = ((U_t)_{i+1}(L_\lambda))^* \text{ où } t = \langle \lambda, \varepsilon_1 \rangle \text{ si } i > 0,$$

et une suite exacte :

$$0 \rightarrow (U_t)_1(L_\lambda) \rightarrow (H^0(\mathcal{O}_{L_\lambda(\mathfrak{p})^*}))^* \rightarrow L_\lambda \rightarrow 0.$$

Remarque 3.4. — Cette proposition montre que la somme formelle de la définition 3.1 de $\mathbf{U}(q)$ est bien un polynôme en q sur $\mathbb{Z}[\varepsilon]$ puisqu'elle porte sur un nombre fini de termes.

COROLLAIRE 3.5. — *Soit λ un poids dominant, on a l'égalité suivante dans $\mathcal{K}[\mathcal{F}]$:*

$$[L_\lambda] - \mathbf{U}(-1)[L_\lambda] = \text{Eu}((\mathcal{O}_{L_\lambda(\mathfrak{p})^*})^*),$$

où Eu désigne la caractéristique d'Euler-Poincaré.

Lien avec la formule des caractères

Serganova introduit dans [12] des polynômes en la variable q présentant des analogies avec les polynômes de Kazhdan-Lusztig tels qu'ils apparaissent dans l'étude de la catégorie \mathcal{O} de [1] dans le cas d'une algèbre de Lie semi-simple complexe.

DÉFINITION 3.6. — *Soit λ un poids dominant. Considérons le complexe de Koszul $S^\bullet(\mathfrak{g}^{-1}) \otimes L_\lambda$; soit $\mu \in P^+$. On pose :*

$$K_{\lambda,\mu}(q) := \sum_{i \geq 0} \dim(H_i(\mathfrak{g}^{-1}, L_\lambda))_\mu q^i,$$

où $(H_i(\mathfrak{g}^{-1}, L_\lambda))_\mu$ désigne la composante \mathfrak{g}_0 -isotypique de poids μ de $H_i(\mathfrak{g}^{-1}, L_\lambda)$. On appelle $K_{\lambda,\mu}$ le **polynôme de Kazhdan-Lusztig associé à λ et μ** .

Remarque 3.7. — La série formelle qui définit $K_{\lambda,\mu}$ est un polynôme car, pour i assez grand, le poids μ n'apparaît pas dans $S^i(\mathfrak{g}^{-1}) \otimes L_\lambda$.

Remarque 3.8. — Soient λ et μ dans P^+ ; on a :

$$K_{\lambda,\lambda} = 1,$$

$$K_{\lambda,\mu} \neq 0 \Rightarrow \mu \leq \lambda \text{ et } \chi_\lambda = \chi_\mu.$$

Attention, il peut y avoir des suites infinies décroissantes de poids ayant même caractère infinitésimal qui restent dans P^+ , par exemple, pour k dans \mathbb{N} , $\lambda_k = -k\varepsilon_m + k\delta_1$.

Vera Serganova aimerait beaucoup avoir une méthode théorique de calcul de ces polynômes, mais elle ne sait pour le moment le faire que pour $q = -1$ (c'est une des raisons pour lesquelles j'ai dit dans le résumé que la compréhension de la structure des blocs de la catégorie \mathcal{F} est presque complète).

La démonstration du lemme qui suit (numéroté 3.4 dans [13]) est une adaptation d'arguments de la théorie de Bernstein-Gelfand-Gelfand [1].

LEMME 3.9. — *Soit λ un poids dominant ; on a l'égalité suivante dans $\tilde{\mathcal{K}}[\mathcal{F}]$ (où $\tilde{\mathcal{K}}[\mathcal{F}]$ est un complété convenable de $\mathcal{K}[\mathcal{F}]$, voir [13], p. 622) :*

$$[L_\lambda] = \sum_{\mu \in P^+} K_{\lambda,\mu}(-1)[V_\mu],$$

ce qui donne le caractère de L_λ si les valeurs en -1 des polynômes de Kazhdan-Lusztig sont connues.

Notons \mathbf{K} la matrice (infinie) constituée des $K_{\lambda,\mu}(-1)$: elle est triangulaire inférieure avec des 1 sur la diagonale. D'après ce qui précède, son inverse calcule les multiplicités des modules simples $[L_\mu]$ dans une suite de Jordan-Hölder d'un module de Kac $[V_\lambda]$.

Notons \mathbf{U} l'opérateur $\mathbf{U}(q)$ de la définition 3.1 pris en $q = -1$. Si on désigne par $\mathbf{K}[\mathfrak{g}(m-1, n)]$ l'analogue de la matrice \mathbf{K} pour l'algèbre $\mathfrak{g}(m-1, n)$ contenue dans \mathfrak{p} , on a le théorème suivant :

THÉORÈME 3.10. — *On a l'égalité de matrices :*

$$\mathbf{K}[\mathfrak{g}(m-1, n)] = \mathbf{K} - \mathbf{K}\mathbf{U}.$$

Ce résultat permet de calculer \mathbf{K} par récurrence si on connaît l'opérateur \mathbf{U} .

Remarque 3.11. — Tout le contenu de ce paragraphe peut se faire à l'identique si on remplace \mathfrak{p} par $\tilde{\mathfrak{p}}$, la sous-algèbre parabolique maximale de \mathfrak{g} obtenue en annulant toutes les coordonnées non diagonales de la dernière ligne. Le rôle de ε_1 est alors joué par δ_n . Quand on utilisera $\tilde{\mathfrak{p}}$, on notera tout avec un tilde sauf $\mathfrak{g}(m-1, n)$ transformée alors en $\mathfrak{g}(m, n-1)$.

4. DESCRIPTION PAR RÉCURRENCE DE L'OPÉRATEUR $\mathbf{U}(q)$ ET FIN DE LA DÉMONSTRATION

Notations et définitions

Pour $0 \leq k \leq m$ et $0 \leq l \leq n$, on notera $\mathfrak{g}(k, l)$ la sous-algèbre de \mathfrak{g} constituée des matrices dont tous les coefficients qui ne sont pas sur la diagonale sont nuls sur les $m-k$ premières (resp. $n-l$ dernières) lignes et colonnes (ces notations sont compatibles avec les notations $\mathfrak{g}(m-1, n)$ et $\mathfrak{g}(m, n-1)$ introduites au paragraphe 3).

Reprenons la sous-algèbre \mathfrak{p} (resp. $\tilde{\mathfrak{p}}$). Soit λ un poids de P^+ ; on dira que λ est *atypique relativement à \mathfrak{p}* (resp. $\tilde{\mathfrak{p}}$) si il existe une racine impaire positive α (resp. $\tilde{\alpha}$) avec $\alpha = \varepsilon_1 - \delta_i$ (resp. $\tilde{\alpha} = \varepsilon_j - \delta_n$) telle que $\langle \lambda + \rho, \alpha \rangle = 0$ (resp. $\langle \lambda + \rho, \tilde{\alpha} \rangle = 0$).

Dans le cas contraire, λ est dit *typique relativement à \mathfrak{p}* (resp. $\tilde{\mathfrak{p}}$). Notons que, comme l'ensemble $A(\lambda)$ de la définition 1.4 est constitué de racines deux à deux orthogonales, si λ est atypique relativement à \mathfrak{p} , alors la racine α est unique.

On note $U_{\lambda,\mu}$ (resp. $\tilde{U}_{\lambda,\mu}$) le polynôme en q qui correspond à l'entrée (λ, μ) de la matrice de l'opérateur $\mathbf{U}(q)$ (resp. $\tilde{\mathbf{U}}(q)$) dans $\mathcal{K}[\mathcal{F}]$ (cf. définition 3.1). On note $U_{\lambda,\mu}[\mathfrak{g}(k, l)]$ le polynôme en q défini de manière analogue pour $\mathfrak{g}(k, l)$.

Dans la catégorie \mathcal{F} , pour un poids dominant λ , on notera p_λ la projection sur le bloc correspondant à la valeur χ_λ du caractère infinitésimal. Si V est un (P, \mathcal{O}_P) -module et si on note \mathcal{O}_V le faisceau correspondant sur la super grassmannienne G/P , on peut considérer les $p_\lambda(\mathcal{O}_V)$ (qui ne sont pas nécessairement les mêmes que les $\mathcal{O}_{p_\lambda(V)}$).

Il existe un foncteur (qui dépend du choix de la sous-algèbre de Cartan, \mathfrak{h}) de \mathcal{F} dans elle-même, contravariant et de carré l'identité

$$V \rightarrow V^\vee,$$

qui conserve la décomposition en \mathfrak{h} -modules irréductibles (et donc le caractère) : soit τ l'automorphisme de \mathfrak{g} qui, à toute matrice M , associe $-{}^{str}M$, où, si M est une matrice par blocs $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, ${}^{str}M = \begin{pmatrix} {}^tA & {}^tC \\ -{}^tB & {}^tD \end{pmatrix}$. Si V est un \mathfrak{g} -module de \mathcal{F} , on construit un module V^τ , isomorphe à V comme super espace vectoriel, sur lequel une matrice M de \mathfrak{g} agit sur x dans V par $-{}^{str}M.x$. On pose $V^\vee = (V^\tau)^*$.

Principe de la récurrence

On veut calculer les $U_{\lambda,\mu}$. Pour le faire, on suppose d'une part qu'on a calculé les opérateurs $\mathbf{U}[\mathfrak{g}(k, l)](q)$, et d'autre part qu'on connaît $U_{\lambda-\alpha,\mu}$ si λ est atypique relativement à \mathfrak{p} avec $\alpha = \varepsilon_1 - \delta_j$ et $\langle \lambda + \rho, \alpha \rangle = 0$. On montre dans le même mouvement que les \mathfrak{g} -modules $(U_t)_i(L_\lambda)$ ($t = \langle \lambda, \varepsilon_1 \rangle$, comme dans la partie 3) sont semi-simples.

La récurrence s'initialise pour $\mathfrak{gl}(1, 1)$; les poids atypiques sont de la forme $\lambda = a\varepsilon_1 - a\delta_1$ (a est supposé entier). Ils sont tous dominants et ils sont tous dans le même bloc. Tous les modules simples associés sont de dimension 1 ; on a la suite exacte :

$$0 \rightarrow (L_{\lambda-\alpha})^\Pi \rightarrow V_\lambda \rightarrow L_\lambda \rightarrow 0,$$

où α est la racine impaire $\varepsilon_1 - \delta_1$. Ceci décrit les $U_{\lambda,\mu}$, qui sont tous nuls sauf pour $\mu = \lambda - \alpha$ pour lequel on a $U_{\lambda,\lambda-\alpha}(q) = \varepsilon q$.

Revenons au cas général.

On veut se ramener à étudier les poids λ qui sont atypiques relativement à \mathfrak{p} , la racine α correspondante étant $\varepsilon_1 - \delta_n$. On montre relativement facilement les deux énoncés suivants :

PROPOSITION 4.1 ([13], théorème 6.2). — *Supposons que λ est typique relativement à \mathfrak{p} (resp. $\tilde{\mathfrak{p}}$) alors, si on note U_λ (resp. \tilde{U}_λ) la matrice ligne formée des polynômes $U_{\lambda,\mu}$ (resp. $\tilde{U}_{\lambda,\mu}$), on a $U_\lambda = 0$ (resp. $\tilde{U}_\lambda = 0$).*

PROPOSITION 4.2. — *Supposons que λ est atypique relativement à \mathfrak{p} , la racine α correspondante étant $\varepsilon_1 - \delta_k$; alors, lorsque les polynômes en q correspondants ont tous un sens, on a :*

$$U_{\lambda,\mu} = U_{\lambda,\mu}[\mathfrak{g}(m, k)].$$

De plus, si on fait l'hypothèse que les $(U_t)_i$ ($t = \langle \lambda, \varepsilon_1 \rangle$) du paragraphe 3 appliqués au $\mathfrak{g}(m, k)$ -module simple $L_\lambda[\mathfrak{g}(m, k)]$ sont des $\mathfrak{g}(m, k)$ -modules semi-simples, alors $(U_t)_i(L_\lambda)$ est un \mathfrak{g} -module semi-simple.

On a un résultat analogue en transformant \mathfrak{p} en $\tilde{\mathfrak{p}}$ et α en $\tilde{\alpha} = \varepsilon_k - \delta_n$, avec les modifications évidentes.

On suppose maintenant que λ est atypique relativement à \mathfrak{p} pour la racine $\alpha = \varepsilon_1 - \delta_n$.

Trois cas de comportements relativement différents surgissent :

THÉORÈME 4.3 ([13], corollaire 6.26). — Soit λ un poids dominant atypique relativement à \mathfrak{p} pour la racine $\alpha = \varepsilon_1 - \delta_n$. On a les relations suivantes :

i) Si $\lambda - \alpha$ est dominant,

$$U_{\lambda,\mu} = \varepsilon[q^{-1}U_{\lambda-\alpha,\mu}]_+ \text{ si } \mu \neq \lambda - \alpha$$

où $[\]_+$ a le même sens que dans la définition 2.1, et

$$U_{\lambda,\lambda-\alpha} = \varepsilon q.$$

ii) Si $\lambda - \alpha$ n'est pas dominant et si

a) $\lambda - \varepsilon_1$ est dominant,

$$U_{\lambda,\mu} = \varepsilon U_{\lambda-\alpha,\mu}[\mathfrak{g}(m, n-1)],$$

b) $\lambda + \delta_n$ est dominant,

$$U_{\lambda,\mu} = \varepsilon U_{\lambda-\alpha,\mu}[\mathfrak{g}(m-1, n)].$$

iii) Si $\lambda - \alpha$, $\lambda - \varepsilon_1$, $\lambda + \delta_n$ ne sont pas dominants,

$$U_{\lambda,\mu} = \varepsilon q U_{\lambda-\alpha,\mu}[\mathfrak{g}(m-1, n-1)].$$

Je ne prétends pas donner ici une démonstration, même incomplète, de ce théorème ; je veux juste indiquer quelques articulations de l'argument de Vera Serganova pour i), sachant que ii) utilise le même type d'éléments. La preuve de iii) est plus technique et compliquée.

On remarque que L_{ε_1} (resp. $L_{-\delta_n}$) est la représentation standard (resp. standard duale) de \mathfrak{g} . L'étude du produit tensoriel avec L_{ε_1} joue un rôle crucial dans la démonstration :

LEMME 4.4 ([13], théorème 5.9). — Soient λ un poids dominant, χ un caractère de $\mathcal{Z}(\mathfrak{g})$ provenant d'un \mathfrak{g} -module simple de dimension finie et E le module L_{ε_1} (resp. $L_{-\delta_n}$). On pose :

$$T = p_\chi(L_\lambda \otimes E).$$

On a :

i) $T^\vee \simeq T$.

ii) Si T est non nul, alors $\chi = \chi_{\lambda+\varepsilon_i}$ ou $\chi_{\lambda+\delta_j}$ (resp. $\chi = \chi_{\lambda-\varepsilon_i}$ ou $\chi_{\lambda-\delta_j}$) pour un certain $i \in \{1, \dots, m\}$, ou $j \in \{1, \dots, n\}$.

iii) Si $\#\chi$ est égal à $\#\lambda$, alors T est simple.

iv) Si T n'est pas irréductible, il a un unique sous-module simple S et, si on note X le noyau de la projection de T^\vee ($= T$) vers S^\vee , alors X contient S et X est engendré comme \mathfrak{g} -module par un vecteur de plus haut poids, dont on note μ le poids. On a $\chi_\mu = \chi$, et il existe une racine α dans Δ_1^+ vérifiant $\langle \mu + \rho, \alpha \rangle = 0$ et telle que le plus haut poids de S est égal à $\mu - \alpha$.

Notons p_{λ}^* la projection de \mathcal{F} sur le bloc dont le caractère infinitésimal est celui du dual de L_{λ} .

Le lien entre $U_{\lambda,\mu}$ et $U_{\lambda-\alpha,\mu}$ provient, avec les notations du paragraphe 3, de la suite exacte de faisceaux au-dessus du super espace projectif $X = G/P$:

$$0 \rightarrow \mathcal{O}_{((L_{\lambda-\alpha}(\mathfrak{p}))^{\Pi})^*} \rightarrow p_{\lambda}^*(\mathcal{O}_{(L_{\lambda-\varepsilon_1}(\mathfrak{p}))^*} \otimes_{\mathcal{O}_X} \mathcal{O}_{(L_{\varepsilon_1})^*}) \rightarrow \mathcal{O}_{(L_{\lambda}(\mathfrak{p}))^*},$$

qui provient du dévissage de L_{ε_1} comme \mathfrak{p} -module :

$$0 \rightarrow L_{\varepsilon_1}(\mathfrak{p}) \rightarrow L_{\varepsilon_1} \rightarrow L_{\varepsilon_2}(\mathfrak{p}) \rightarrow 0$$

(suite d'Euler pour le fibré tangent à X).

Le terme du milieu de cette suite exacte de faisceaux s'écrit

$$p_{\lambda}^*(\mathcal{O}_{(L_{\lambda-\varepsilon_1}(\mathfrak{p}))^*} \otimes_{\mathbb{C}} (L_{\varepsilon_1})^*).$$

Comme $\lambda - \varepsilon_1$ est typique relativement à \mathfrak{p} , la proposition 4.1 montre que la cohomologie de ce faisceau est concentrée en degré 0 et est égale à $(p_{\lambda}(L_{\varepsilon_1} \otimes L_{\lambda-\varepsilon_1}))^*$.

La suite exacte longue de cohomologie, associée à la suite exacte de faisceaux, donne des isomorphismes

$$H^i(\mathcal{O}_{(L_{\lambda}(\mathfrak{p}))^*}) \simeq H^{i+1}(\mathcal{O}_{((L_{\lambda-\alpha}(\mathfrak{p}))^*)^{\Pi}}) \text{ si } i \geq 1;$$

donc, compte tenu de la proposition 3.3, ii), on obtient des isomorphismes

$$(U_{t'})_{i+2}(L_{\lambda-\alpha}) \simeq (U_t)_{i+1}(L_{\lambda}) \text{ si } i \geq 1,$$

où $t = \langle \lambda, \varepsilon_1 \rangle$, $t' = \langle \lambda - \alpha, \varepsilon_1 \rangle$ (on a $t' = t - 1$).

Pour $i = 0$, on a une suite exacte :

$$0 \rightarrow (U_{t'})_2(L_{\lambda-\alpha}) \rightarrow (H^0(\mathcal{O}_{(L_{\lambda}(\mathfrak{p}))^*}))^* \rightarrow p_{\lambda}(L_{\varepsilon_1} \otimes L_{\lambda-\varepsilon_1}) \rightarrow (H^0(\mathcal{O}_{(L_{\lambda-\alpha}(\mathfrak{p}))^*}))^* \rightarrow 0.$$

Ces isomorphismes donnent l'égalité de i) pour les coefficients des puissances strictement supérieures à 1 de q (le terme constant est nul par la définition 3.1). Le coefficient de q est plus difficile à trouver.

On montre dans un premier temps qu'on a l'égalité suivante dans le groupe de Grothendieck :

$$[(U_t)_1(L_{\lambda})] = \varepsilon[(U_{t'})_2(L_{\lambda-\alpha})] + \varepsilon[L_{\lambda-\alpha}] + [R_{\lambda}], \quad (4.5(\lambda))$$

où R_{λ} est un \mathfrak{g} -module tel que $[R_{\lambda} : L_{\mu}] \leq \varepsilon[(U_t)_1(L_{\lambda-\alpha}) : L_{\mu}]$ pour $\mu \in P^+$.

La preuve est algébrique et elle repose sur le lemme 4.4.

Pour obtenir i), il reste à voir que $R_{\lambda} = 0$.

On a l'égalité (4.5(λ)), et donc l'égalité (4.5($\lambda + k\alpha$)) est vérifiée pour tout $k \in \mathbb{N}$.

On montre que $R_{\lambda+k\alpha} = 0$ pour k assez grand : je donne ici quelques détails.

Soit k un entier assez grand, on pose $\mu = \lambda + k\alpha$, montrons que $R_{\mu} = 0$. On a la suite exacte :

$$0 \rightarrow (U_t)_1(L_{\mu}) \rightarrow (H^0(\mathcal{O}_{L_{\mu}(\mathfrak{p})^*}))^* \rightarrow L_{\mu} \rightarrow 0,$$

qui provient de la proposition 3.3. On montre d'abord que :

(*) les plus hauts poids des facteurs de composition de $(U_t)_1(L_\mu)$ font partie des plus hauts poids des facteurs de composition de $L_{\varepsilon_1} \otimes L_\mu$.

On peut regarder le faisceau $\mathcal{O}_{L_\mu(\mathfrak{p})^*}$ comme un faisceau M au-dessus de $G_0/P_0 = \mathbb{P}^{m-1}(\mathbb{C})$; les résultats de Penkov [11] permettent de construire une filtration finie G_0 -équivariante de M dont les quotients M_i sont des \mathcal{O}_{G_0/P_0} -modules, chacun provenant d'un \mathfrak{g}_0 -module simple dont on notera μ_i le plus haut poids. De plus, les poids $\mu - \mu_i$ doivent être les plus hauts poids de \mathfrak{g}_0 -composantes de l'algèbre extérieure $\Lambda(\mathfrak{g}^{+1})$, donc on a $\mu - \mu_i = \sum_{\beta \in S_i} \beta$ où S_i est un sous-ensemble de Δ_1^+ .

On applique le théorème de Borel-Weil-Bott aux faisceaux M_i : comme k est assez grand, il n'y a pas de cohomologie en degré strictement positif, et on a :

$$H_{G_0/P_0}^0(M_i) = (L_{\mu_i}(\mathfrak{g}_0))^*.$$

La cohomologie de M est donc connue comme \mathfrak{g}_0 -module : elle est concentrée en degré 0 et on a :

$$H_{G_0/P_0}^0(M) = \bigoplus_i (L_{\mu_i}(\mathfrak{g}_0))^*.$$

De plus, $H_{G_0/P_0}^0(M)$ est un \mathfrak{g} -module égal à $H^0(\mathcal{O}_{L_\mu(\mathfrak{p})^*})$, donc les \mathfrak{g}_0 -plus hauts poids de $(H^0(\mathcal{O}_{L_\mu(\mathfrak{p})^*}))^*$ sont les μ_i . On doit chercher parmi les μ_i , qui sont les \mathfrak{g} -plus hauts poids, c'est-à-dire les poids ν tels que $\chi_\mu = \chi_\nu$ et qui font partie des μ_i . Un tel ν est nécessairement de la forme $\mu - \sum_{\beta \in S_\nu} \beta$ avec S_ν contenu dans Δ_1^+ et, si β appartient à S_ν et si $\beta = \varepsilon_1 - \delta_k$, alors $k = n$. Ceci fait que toute racine de S_ν est soit orthogonale, soit égale à α . Alors $\langle \mu - \nu, \varepsilon_1 \rangle$ est égal à 0 ou 1, ce qui permet de montrer (*).

En utilisant le lemme 4.4, iii), on en déduit que $R_\mu = 0$ pour k assez grand. Il reste à voir que

$$R_\mu = 0 \Rightarrow R_{\mu-\alpha} = 0,$$

ce qui se fait en reprenant les arguments algébriques aboutissant à l'égalité (4.5(λ)), auxquels on incorpore deux ingrédients supplémentaires : l'hypothèse $R_\mu = 0$ et le fait que, si V est un \mathfrak{g} -module dont les facteurs de composition sont de multiplicité 1 et vérifiant $V = V^\vee$, alors V est semi-simple.

Pour obtenir les théorèmes 2.2 et 2.3, on remarque que les résultats de ce paragraphe nous permettent de calculer la matrice de \mathbf{U} et donc de calculer l'opérateur \mathbf{K} par récurrence (théorème 3.10). Le lien avec les opérateurs s_α s'obtient en remarquant que si $\alpha = \varepsilon_1 - \delta_n$, on a les mêmes matrices pour \mathbf{U} et s_α .

5. ÉTAT DES LIEUX

Depuis [13], je connais deux développements récents du sujet.

Dans [14], Vera Serganova cite un certain nombre de résultats nouveaux, sans démonstration. D'une part elle étudie les blocs des différentes catégories \mathcal{F} (quand on fait varier m et n), et montre qu'à équivalence près ils sont caractérisés par le degré d'atypie. D'autre part, elle utilise des techniques d'induction géométrique pour obtenir des résultats analogues à la formule des caractères pour les super algèbres de Lie de

type orthosymplectique : dans ce cadre, les modules de Kac n'existent plus, ce qui oblige à formuler des énoncés plus compliqués que je ne chercherai pas à décrire ici.

Le travail de Jonathan Brundan [4] décrit le groupe de Grothendieck $\mathcal{K}[\mathcal{F}]$ comme $\Lambda^m(V^*) \otimes \Lambda^n(V)$ où V désigne la représentation naturelle du groupe quantique $\mathcal{U}_q(\mathfrak{gl}_\infty)$.

En utilisant la base canonique (resp. canonique duale) de Lusztig ([10]) de ce module, il paramètre les classes d'objets simples (resp. basculants, qui sont en l'occurrence les objets projectifs indécomposables, qui sont aussi injectifs) de \mathcal{F} .

Cela lui permet d'obtenir une formule de multiplicité d'un module simple L_λ dans un module de Kac V_μ , qui montre que ces multiplicités sont toujours égales à 0 ou 1.

Voici son résultat, qui est très élégant ; je le cite dans les notations que j'ai utilisées, dans le cas maximale atypique pour $\mathfrak{gl}(n, n)$.

THÉORÈME 5.1 ([4], Main theorem, p. 3). — *Soit $\lambda = a_1\varepsilon_1 + a_2\varepsilon_2 + \dots + a_n\varepsilon_n - a_n\delta_1 - \dots - a_1\delta_n$ un poids dominant maximale atypique de $\mathfrak{gl}(n, n)$. Soit (k_1, \dots, k_n) un n -uplet d'entiers strictement positifs minimal dans l'ordre lexicographique pour la propriété suivante :*

pour tous $(\theta_1, \dots, \theta_n)$ dans $\{0, 1\}^n$, $\lambda + \sum_{s=1}^n \theta_s k_s (\varepsilon_s - \delta_{k-s})$ est conjugué pour l'action $w(\lambda + \rho) - \rho$ du groupe de Weyl W à un élément $R_\theta(\lambda)$ de P^+ .

Alors la multiplicité $[V_\mu : L_\lambda]$ est égale à 1 si μ est un $R_\theta(\lambda)$, à 0 sinon.

Ce théorème permet d'obtenir un algorithme qui détermine, pour un poids dominant maximale atypique λ donné, quels sont les modules de Kac V_μ dont L_λ est un sous-quotient. On remarquera toutefois que cet algorithme ne fournit pas directement les facteurs de composition d'un module de Kac donné.

La démonstration utilise très peu la structure super, le principal argument $\mathbb{Z}/2\mathbb{Z}$ -gradué provient de l'étude des produits tensoriels d'un module de Kac V_μ par les puissances symétriques de L_{ε_1} et $L_{-\delta_n}$, dont il décompose les projections dans des blocs convenables.

Dans le même article, Brundan conjecture des résultats analogues pour la catégorie \mathcal{O} de $\mathfrak{gl}(m, n)$.

Parmi les nombreuses questions soulevées par les travaux de Brundan et de Serganova, il en est une que je souhaite poser : toutes les constructions que nous avons vues ici dépendent du choix de la sous-algèbre de Borel \mathfrak{b} de \mathfrak{g} ; il se trouve que, contrairement au cas classique, les sous-algèbres de Borel ne sont pas toutes \mathfrak{g} -conjuguées. Si on choisit une autre sous-algèbre de Borel \mathfrak{b}' de \mathfrak{g} contenant \mathfrak{b}_0 , la catégorie \mathcal{F} reste elle-même en tant que catégorie, mais comme *catégorie de modules de plus haut poids*, elle change de nature. Quel parti peut-on en tirer ? En effet, le choix de \mathfrak{b} semble privilégié à cause de la \mathbb{Z} -gradation de longueur 3 de $\mathfrak{gl}(m, n)$, mais cette propriété disparaît dans le cas orthosymplectique.

RÉFÉRENCES

- [1] J.N. BERNSTEIN, I.M. GELFAND, S.I. GELFAND – *Categories of \mathfrak{g} -modules*, *Funct. Anal. and Appl.* 10 (1976), pp. 87-92.
- [2] J. BERNSTEIN, D. LEITES – *A formula for the characters of the irreducible finite dimensional representations of Lie superalgebras of series \mathfrak{gl} and \mathfrak{sl}* , *CR Acad. Bulgare Sci.* 33 (1980), pp. 1049-1051.
- [3] N. BOURBAKI – *Groupes et algèbres de Lie, Chapitres 4, 5 et 6*, Masson, Paris, 1981.
- [4] J. BRUNDAN – *Kazhdan-Lusztig polynomials and character formulae for the Lie superalgebra $\mathfrak{gl}(m, n)$* , *J. Amer. Math. Soc.* 16 no. 1 (2003), pp. 185-231 (electronic).
- [5] J.C. JANTZEN – *Representations of algebraic groups*, second edition, *Mathematical surveys and monographs vol. 107*, AMS 2003.
- [6] J. DIXMIER – *Algèbres enveloppantes*, Gauthier-Villars, Paris, 1974.
- [7] W. FULTON, P. PRAGACZ – *Schubert varieties and degeneracy loci*, LNM 1689, Springer, 1998.
- [8] V.G. KAC – *Lie superalgebras*, *Advances in Math.* 26 (1977), pp. 8-96.
- [9] V.G. KAC – *Representations of classical Lie superalgebras*, LNM 676, Springer (1978), pp. 597-626.
- [10] G. LUSZTIG – *Introduction to quantum groups*, *Progress in Math.* vol. 110, Birkhäuser (1993).
- [11] I. PENKOV – *Borel-Weil-Bott theory for classical Lie supergroups*, *Contemporary problems in mathematics vol. 32*, VINITI, Moscou 1988 (en russe). Traduit dans *J. Soviet Math.* 51 (1990), no. 1, 2108–2140.
- [12] V.V. SERGANOVA – *Kazhdan-Lusztig polynomials for the Lie superalgebra $\mathfrak{gl}(m, n)$* , *Advances in Soviet Math.* 16 (1993), pp. 151-165.
- [13] V.V. SERGANOVA – *Kazhdan-Lusztig polynomials and character formula for the Lie superalgebra $\mathfrak{gl}(m, n)$* , *Selecta Math.* vol. 2 n. 4 (1996), pp. 607-651.
- [14] V.V. SERGANOVA – *Characters of irreducible representations of simple Lie superalgebras*, *Documenta Math. Extra volume ICM 1998 II*, pp. 583-596.
- [15] A. SERGEEV – *The invariant polynomials on simple Lie superalgebras*, *Representation theory, AMS*, vol. 3 (1999), pp. 250-280.

Caroline GRUSON

Institut Élie Cartan (IÉCN)

Faculté des Sciences

Université de Nancy I (Henri Poincaré)

F-54506 Vandœuvre-lès-Nancy

E-mail : `Caroline.Gruson@iecn.u-nancy.fr`

TRAVAUX DE ZINK

by William MESSING

1. INTRODUCTION AND PRELIMINARY DEFINITIONS AND RESULTS

Fix a prime number p . All rings considered will be $\mathbf{Z}_{(p)}$ -algebras. If R is a ring we will consider p -divisible groups over R and in particular those which are formal groups. If $\frac{1}{p} \in R$, then p -divisible groups are étale and consequently given by continuous representations $\rho : \pi_1(\mathrm{Spec}(R)) \rightarrow \mathrm{GL}_h(\mathbf{Z}_p)$. Hence we shall assume p is either nilpotent in R or R is separated and complete for a topology having a neighborhood basis of 0 consisting of ideals and that p is topologically nilpotent.

With these conventions, the aim of the various Dieudonné theories is to classify the category of p -divisible groups over R via functors to categories living in the realm of (semi)linear algebra. One should think of them as analogous to the functor $G \mapsto \mathrm{Lie}(G)$ which establishes an equivalence of categories between formal groups and Lie algebras when R is a \mathbf{Q} -algebra. We will not give an overview of the various Dieudonné theories, but rather concentrate on the most recent, Zink's theory of displays. Nevertheless it will be necessary for us to relate Zink's theory to Cartier's theory and to the crystalline theory. We refer the reader to [Ta], [Ser], [Gr1], [Gr2], [Dem], [Fon1] for p -divisible groups, to [Car1], [Car2], [Haz], [Laz], [Z1], [Z2] for Cartier theory, to [Gr1], [Gr2], [MM], [M], [BBM], [BM1], [BM2], [dJ2], [dJM] for crystalline Dieudonné theory.

If R is a perfect field of characteristic p , these theories are, for p -divisible groups (formal in the case of Cartier's theory), all equivalent. Indeed it was one of Zink's motivations in developing his theory to relate the Cartier theory to the crystalline theory. But, in establishing properties of his theory, he uses both the Cartier and the crystalline theories. Hence there is a symbiotic relationship between the three theories.

We refer to [Bour] for the standard facts about the Witt vector ring, $W(R)$. We write $w_n : W(R) \rightarrow R$ for the ghost component maps, $f : W(R) \rightarrow W(R)$ for the Frobenius ring endomorphism and $v : W(R) \rightarrow W(R)$ for the additive Verschiebung endomorphism. Let $I_R = \mathrm{Ker}(w_0) = \mathrm{im}(v)$. If $a \in R$, $[a]$ denotes its Teichmüller representative.

LEMMA 1.1. — *If R is separated and complete in the p -adic topology, then $W(R)$ is separated and complete in both its p -adic and I_R -adic topologies. If p is nilpotent in R , these topologies coincide and it is finer than the v -adic topology.*

DEFINITION 1.2. — *A display \mathcal{P} over R is a quadruple (P, Q, F, F_1) where P is a finitely generated projective $W(R)$ -module, Q a submodule, $F : P \rightarrow P$, $F_1 : Q \rightarrow P$ are f -semilinear such that*

- (i) $I_R P \subset Q$.
- (ii) $0 \rightarrow Q/I_R P \rightarrow P/I_R P \rightarrow P/Q \rightarrow 0$ is a split sequence of R -modules.
- (iii) P is generated by $\text{im}(F_1)$.
- (iv) $F_1(v(\xi)x) = \xi F(x)$ for $\xi \in W(R)$, $x \in P$.

If $u : M \rightarrow N$ is a f -semilinear map of $W(R)$ -modules, we set $M^{(1)} = W(R) \otimes_{f, W(R)} M$ for the extension of scalars using f and denote by $u^\sharp : M^{(1)} \rightarrow N$ the associated linear map.

With the obvious notion of morphisms, displays form an additive category and, if we define a morphism of displays $u : \mathcal{P}' \rightarrow \mathcal{P}$ to be an admissible monomorphism (resp. epimorphism) provided $u : P' \rightarrow P$ is injective (resp. surjective) and $u^{-1}(Q) = Q'$ (resp. $u(Q') = Q$), we equip Displays_R with the structure of an exact category.

DEFINITION 1.3. — *A normal decomposition for a display \mathcal{P} over R is a direct sum decomposition $P = L \oplus T$ such that $Q = L \oplus I_R T$.*

If R is a p -adic ring, in particular if p is nilpotent in R , normal decompositions always exist. This is a consequence of the fact that finitely generated projective modules can always be lifted for surjections $A \rightarrow B$ whose kernel is a nilideal or such that A is separated and complete for the topology given by powers of the kernel.

Examples.— (i) The display corresponding to the formal multiplicative group $\mathcal{G} = (W(R), I_R, f, v^{-1})$.

(ii) If $R = k$, a perfect field of characteristic p , $M \mapsto \mathcal{P}_M = (M, V(M), F, V^{-1})$ establishes an equivalence of categories between Dieudonné modules over k and displays over k .

From now on we assume p is nilpotent in R , unless we explicitly state the contrary.

If $u : R \rightarrow R'$ is a ring homomorphism and \mathcal{P} is a display over R , the base changed display $u_*(\mathcal{P})$ is the display over R' , $\mathcal{P}' = (P', Q', F', F'_1)$, where $P' = W(R') \otimes_{W(R)} P$, $Q' = \text{Ker}(P' \rightarrow R' \otimes_R P/Q)$, $F' = f \otimes F$ and F'_1 is determined by

$$F'_1(v(\xi) \otimes x) = \xi \otimes F(x), \xi \in W(R'), x \in P$$

and

$$F'_1(\xi \otimes y) = f(\xi) \otimes F_1(y), \xi \in W(R'), y \in Q.$$

Using a normal decomposition, it is easy to show that F'_1 exists and \mathcal{P}' is a display.

DEFINITION 1.4. — Let $\mathcal{P}, \mathcal{P}'$ be displays over R . A bilinear form of displays $(\ , \) : \mathcal{P} \times \mathcal{P}' \rightarrow \mathcal{G}$ is a bilinear map $P \times P' \rightarrow W(R)$ such that $v(F_1 y, F_1' y') = (y, y')$ for $y \in Q, y' \in Q'$.

If \mathcal{P} is a display over R , its dual display $\mathcal{P}^t = (P^\vee, \widehat{Q}, F, F_1)$ where $P^\vee = \text{Hom}_{W(R)}(P, W(R)), \widehat{Q} = \{z \in P^\vee \mid z(Q) \subset I_R\}$ and F and F_1 are determined by

$$\begin{aligned} (F_1 x, Fz) &= f(x, z) && \text{for } x \in Q, z \in P^\vee \\ (Fx, Fz) &= pf(x, z) && \text{for } x \in P, z \in P^\vee \\ (Fx, F_1 z) &= f(x, z) && \text{for } x \in P, z \in \widehat{Q} \\ v(F_1 x, F_1 z) &= (x, z) && \text{for } x \in Q, z \in \widehat{Q}. \end{aligned}$$

We have a canonical isomorphism

$$\text{Bil}(\mathcal{P}, \mathcal{P}'; \mathcal{G}) \simeq \text{Hom}(\mathcal{P}', \mathcal{P}^t).$$

PROPOSITION 1.5. — There is a unique linear map $V^\sharp : P \rightarrow P^{(1)}$ determined by $V^\sharp(\xi Fx) = p\xi \otimes x, V^\sharp(\xi F_1 y) = \xi \otimes y$, for $\xi \in W(R), x \in P, y \in Q$.

This is established by taking a normal decomposition $P = L \oplus T$, showing that $F_1^\sharp \oplus F^\sharp : L^{(1)} \oplus T^{(1)} \rightarrow P$ is bijective and defining V^\sharp to be the composite

$$(\text{id} \oplus p \cdot \text{id}) \circ (F_1^\sharp \oplus F^\sharp)^{-1} : P \rightarrow L^{(1)} \oplus T^{(1)} = P^{(1)}.$$

One has $F^\sharp \circ V^\sharp = p \cdot \text{id}_P, V^\sharp \circ F^\sharp = p \cdot \text{id}_{P^{(1)}}$. If $P^{(i)}$ is the scalar extension of P using f^i , then V^\sharp gives rise to $V_i^\sharp : P^{(i)} \rightarrow P^{(i+1)}$.

DEFINITION 1.6. — \mathcal{P} satisfies the nilpotence condition or \mathcal{P} is a nilpotent display provided there is an N such that $V_N^\sharp \circ V_{N-1}^\sharp \circ \dots \circ V^\sharp$ is zero modulo $I_R + pW(R)$.

Remark 1.7. — In [Z5], displays were called $3n$ -displays ($3n$ for “not necessarily nilpotent”) and nilpotent displays were called displays. We follow Zink’s more recent terminology (cf. his Paris 13 lectures of February, 2006) here. Also in [Z5], F_1 was denoted by V^{-1} . Zink and Langer have initiated a theory of higher displays, [LZ2], in which $P = P_0, Q = P_1$ and there are higher P_i and $F_i : P_i \rightarrow P$. For this reason we write, following Zink, F_1 instead of his original V^{-1} .

Remark 1.8. — Locally on $\text{Spec}(R)$, if $L \oplus T$ is a normal decomposition we will have L and T free modules and if T has basis $\{e_1, \dots, e_d\}$ and L has basis $\{e_{d+1}, \dots, e_h\}$, the map $F_1^\sharp \oplus F^\sharp$ will be expressed in terms of these bases by a matrix $(\alpha_{ij}) \in \text{GL}_h(W(R))$. Conversely any such invertible matrix will determine a display. If the matrix (α_{ij}) has inverse $(\beta_{k\ell})$, and B is the $(h - d) \times (h - d)$ matrix with entries in R/pR given by $B = (w_0(\beta_{k\ell})) \bmod p)_{k,\ell=d+1,\dots,h}$, then \mathcal{P} is nilpotent if and only if there is an N such that

$$B^{(p^N)} \dots B = 0,$$

where $B^{(p^i)}$ is the matrix obtained by applying the i -th iterate of Frobenius to B .

If e_i is a basis for a free module over the Cartier ring, then the relations

$$\begin{aligned} Fe_i &= \sum \alpha_{ji}e_j, & i = 1, \dots, d; \\ e_i &= V\left(\sum \alpha_{ji}e_j\right), & i = d + 1, \dots, h \end{aligned}$$

define a reduced Cartier module. Relations of this form were called by Norman [N] “displayed structural equations” of a reduced Cartier module. This is the origin of Zink’s use of the term display.

Remark 1.9. — Let $S \xrightarrow{u} R$ be a surjection whose kernel is a nilideal. Let \mathcal{P} be a display over R . Then there is a display \mathcal{P}' over S and an isomorphism $u_*(\mathcal{P}') \xrightarrow{\sim} \mathcal{P}$.

This is proven using the fact that finitely generated projective $W(R)$ -modules can be lifted to finitely generated projective $W(S)$ -modules and using normal decompositions. Nakayama’s lemma then shows that lifting modules are determined up to isomorphism (non-unique!).

If \mathcal{P}/R is a nilpotent display and \mathcal{P}' is a lifting to S , then \mathcal{P}' is nilpotent too. This is clear as $\text{Ker}(S \rightarrow R)$ is a nilideal.

We ask about the ambiguity in the lifting \mathcal{P}' of \mathcal{P} . If $\mathcal{P}' = (P', Q', F', F'_1)$, $J = \text{Ker}(S \xrightarrow{u} R)$ and $\alpha : P' \rightarrow W(J) \otimes_{W(S)} P'$, we define a display \mathcal{P}'_α over S lifting \mathcal{P} by $\mathcal{P}'_\alpha = (P', Q', F'_\alpha, F'_{1\alpha})$, where $F'_\alpha(x) = F'x - \alpha(F'x)$, for $x \in P'$, $F'_{1\alpha}(y) = F'_1y - \alpha(F'_1y)$, for $y \in Q'$. Then \mathcal{P}'_α is a display and Zink shows any lifting of \mathcal{P} is isomorphic to a \mathcal{P}'_α .

Remark 1.10. — Assume $p \cdot 1_R = 0$. Let \mathcal{P} be a display over R , $\mathcal{P}^{(p)}$ be the display over R given by $(\text{Frob})_*\mathcal{P}$. Then V^\sharp commutes with F and F_1 and hence defines a morphism of display $\mathcal{F}_\mathcal{P} : \mathcal{P} \rightarrow \mathcal{P}^{(p)}$. Similarly F^\sharp defines a morphism of displays $\mathcal{V}_\mathcal{P} : \mathcal{P}^{(p)} \rightarrow \mathcal{P}$. Of course both composites are multiplications by p .

If $R \rightarrow R'$ is a ring homomorphism, there is an obvious notion of a descent datum for \mathcal{P}' a R' -display and, if \mathcal{P} is a R -display, $\mathcal{P}_{R'}$ has a canonical descent datum, can.

Zink proves:

PROPOSITION 1.11. — *If $R \rightarrow R'$ is faithfully flat and p is nilpotent in R , then $\mathcal{P} \mapsto (\mathcal{P}_{R'}, \text{can})$ is an equivalence of categories between $\text{Displays}/R$ and the category of R' -displays equipped with descent data. The same is true for nilpotent displays.*

2. THE CRYSTALS ASSOCIATED TO DISPLAYS

We refer to [Ber] for a detailed discussion of crystals, crystalline cohomology,... and recall the bare minimum here. An ideal $J \subset A$ has divided powers if we are given maps $\gamma_n : J \rightarrow J$, $n \geq 1$, satisfying axioms imposed by thinking of $\gamma_n(x)$ as $\frac{x^n}{n!}$. The ideal $(p) \subset \mathbf{Z}_{(p)}$ has unique divided powers since $\frac{p^n}{n!} \in (p)$. It follows that for any ring A , $p \cdot A$ has divided powers. If $J \subset A$ is an ideal with divided powers we require that its divided

powers agree with those on $J \cap pA$. This is called the compatibility condition. If R is a $\mathbf{Z}_{(p)}$ -algebra, then $I_R \subset W(R)$ has canonical divided powers which are compatible with those on $p \cdot W(R)$. These are determined by $\gamma_n(v(x)) = \frac{p^{n-1}}{n!}v(x^n)$, [Gr2]. The ideals $v^m(W(R))$ are sub-divided power ideals. We refer to [Ber] for the definition of nilpotent divided powers and to [M], [Z3] for a weaker notion.

We continue to assume p is nilpotent in R . If A is an R -algebra, a divided power thickening of A is a surjection $A' \xrightarrow{\pi} A$ such that p is nilpotent in A' and $\text{Ker}(\pi)$ is equipped with divided powers (satisfying the compatibility condition). A morphism of divided power thickenings is a commutative diagramm

$$(*) \quad \begin{array}{ccc} A' & \xrightarrow{\pi} & A \\ \psi \downarrow & & \downarrow \phi \\ B' & \xrightarrow{\tilde{\pi}} & B \end{array}$$

such that $\text{Ker}(\pi)$, $\text{Ker}(\tilde{\pi})$ have divided powers, $\psi(\gamma_n(x)) = \gamma_n(\psi(x))$, $n \geq 1$ for $x \in \text{Ker} \pi$.

A *crystal in modules* M on R is the giving for every divided power thickening $A' \xrightarrow{\pi} A$ of a A' -module, $M_{(A' \xrightarrow{\pi} A)}$, and for every morphism of divided power thickenings of an isomorphism

$$T_{(\psi,\phi)} : B' \otimes_{A'} (M_{(A' \xrightarrow{\pi} A)}) \xrightarrow{\sim} M_{(B' \xrightarrow{\tilde{\pi}} B)},$$

these isomorphisms being required to satisfy the obvious transitivity condition.

Similarly we define a Witt-crystal on R as the giving for any divided power thickening of an R -algebra ($A' \xrightarrow{\pi} A$) of a $W(A')$ -module $K_{(A' \xrightarrow{\pi} A)}$ together with, for any diagram $(*)$, an isomorphism

$$T'_{(\psi,\phi)} : W(B') \otimes_{W(A')} K_{(A' \xrightarrow{\pi} A)} \longrightarrow K_{(B' \xrightarrow{\tilde{\pi}} B)}.$$

We want now to explain Zink’s functors

$$\text{Nilpotent Displays}/R \longrightarrow \text{Crystals}/R$$

$$\mathcal{P} \longmapsto \mathcal{D}_{\mathcal{P}}$$

$$\text{Nilpotent Displays}/R \longrightarrow \text{Witt crystals}$$

$$\mathcal{P} \longmapsto \mathcal{H}_{\mathcal{P}}.$$

Note first, and for later use as well, that if C is a $\mathbf{Z}_{(p)}$ -algebra commutative and associative, but not necessarily with an identity, $W(C)$ is defined.

Assume C has divided powers; then we can divide the n -th ghost component and write $w'_n = \frac{w_n}{p^n}$ via

$$w'_n(\gamma) = \frac{\sum_{i=0}^n p^i x_i^{p^{n-i}}}{p^n} = \sum_{i=0}^n (p^{n-i} - 1)! \gamma_{p^{n-i}}(x_i).$$

In particular if $J = \text{Ker}(\pi)$, $\pi : A' \twoheadrightarrow A$ has divided powers. The map $W(J) \xrightarrow{w'} J^{\mathbf{N}}$, $x \mapsto (w'_0(x_0), w'_1(x_0, x_1), \dots, w'_n(x_0, \dots, x_n), \dots)$ is an isomorphism of $W(A')$ -modules where the target is made into a $W(A')$ -module via

$$\xi \cdot \underline{y} = (w_0(\xi)y_0, w_1(\xi)y_1, \dots, w_n(\xi)y_n, \dots), \quad \underline{y} \in J^{\mathbf{N}}.$$

We also refer to the $y_i = w'_i(x)$ as the logarithmic coordinates and write this isomorphism as $\log : W(J) \xrightarrow{\sim} J^{\mathbf{N}}$. Then $\log^{-1}(J, 0, 0, \dots) \subset W(J)$ is an ideal in $W(A')$ which we abusively denote by J .

This embedding $J \subset W(A')$ depends of course on the divided power structure on J . We have $J \oplus I_{A'} \subset W(A')$ and $f(J) = (0)$. By using a normal decomposition, one easily obtains the basic lemma:

LEMMA 2.1. — *Assuming \mathcal{P}' is a display over A' , p is nilpotent in A' and $J \subset A'$ is an ideal with divided powers, F'_1 has a unique extension to $W(J) \cdot P + Q$ such that $F'_1(J \cdot P) = (0)$.*

COROLLARY 2.2. — *Assume p is nilpotent in A' , $J \subset A'$ is a nilpotent ideal and $\mathcal{P}'_1, \mathcal{P}'_2$ are nilpotent displays over A' . Let $\pi : A' \rightarrow A'/J$, $\mathcal{P}_i = \pi_*(\mathcal{P}'_i)$. Then*

$$\text{Hom}(\mathcal{P}'_1, \mathcal{P}'_2) \longrightarrow \text{Hom}(\mathcal{P}_1, \mathcal{P}_2)$$

is injective and its cokernel is p -torsion. Further, if the nilradical of A' is nilpotent, $\text{Hom}(\mathcal{P}'_1, \mathcal{P}'_2)$ is p -torsion free.

This result is the analogue of rigidity for p -divisible groups.

We now introduce the concept of a \mathcal{P} -triple where \mathcal{P} is a nilpotent display over R . Let $R' \xrightarrow{\pi} R$ be a surjection whose kernel J is equipped with divided powers and where p is nilpotent in R' .

DEFINITION 2.3. — *A \mathcal{P} -triple over R' is a triple (P', F', F'_1) such that P' is a finitely generated projective $W(R')$ -module which lifts P , and if $Q' =$ the inverse image of Q , $F' : P' \rightarrow P'$, $F'_1 : Q' \rightarrow P'$ are f -semilinear and satisfy*

- (i) F' (resp. F'_1) lifts F (resp. F_1).
- (ii) $F'_1(v(\xi) \cdot x) = \xi F'(x)$, $\xi \in W(R')$, $x \in P'$.
- (iii) $F'_1(J \cdot P') = (0)$.

If $\mathcal{P}_i, i = 1, 2$ are nilpotent displays and \mathcal{T}_i are \mathcal{P}_i -triples and $\alpha : \mathcal{P}_1 \rightarrow \mathcal{P}_2$, there is an obvious notion of an α -morphism $\tilde{\alpha} : \mathcal{T}_1 \rightarrow \mathcal{T}_2$.

There is also the notion of base change for \mathcal{P} -triples. If

$$\begin{array}{ccc} R' & \xrightarrow{u'} & S' \\ \downarrow & & \downarrow \\ R & \xrightarrow{u} & S \end{array}$$

is a morphism of divided power thickenings and \mathcal{T} is a \mathcal{P} -triple, $u'_*(\mathcal{T})$ is a $u_*(\mathcal{P})$ -triple with $P'_{S'} = W(S') \otimes_{W(R)} P'$ and $F'_{S'}, F'_{1S'}$ determined in the obvious way.

THEOREM 2.4. — *Let $\alpha : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ be a morphism of nilpotent displays over R , $R' \rightarrow R$ a divided power thickening and $\mathcal{T}_i, \mathcal{P}_i$ -triples over R' . Then there exists a unique α -morphism $\tilde{\alpha} : \mathcal{T}_1 \rightarrow \mathcal{T}_2$.*

Using this theorem, we construct functors:

$$\begin{aligned} \mathcal{D} &: \text{nilpotent displays}/R \longrightarrow \text{crystals}/R \\ \mathcal{K} &: \text{nilpotent displays}/R \longrightarrow \text{Witt crystals}/R \end{aligned}$$

If \mathcal{P} is a nilpotent display over R and $\mathcal{T} = (P', F, F_1)$ is a \mathcal{P} -triple over the divided power thickening $R' \rightarrow R$, then $\mathcal{D}_{\mathcal{P}}(R') = P'/I_{R'} P'$ and $\mathcal{K}_{\mathcal{P}}(R') = P'$.

Let $\mathcal{P} = (P, Q, F, F_1)$ be a nilpotent display over R . We consider its ‘‘Hodge filtration’’ $Q/I_R P \subset P/I_R P$. If $R' \rightarrow R$ is a divided power thickening and $\mathcal{T} = (P', F, F_1)$ is a \mathcal{P} -triple over R' , we call a lift of the Hodge filtration the giving of a direct summand $L \subset P'/I_{R'} P'$ such that $L \otimes_{R'} R = Q/I_R P$. If Q'_L is the inverse image of L in P' , then $\mathcal{P}' = (P', Q'_L, F, F_1)$ is a nilpotent display over R' which lifts \mathcal{P} .

Consider a divided power thickening $R' \rightarrow R$. Let \mathcal{C} be the category whose objects are pairs (\mathcal{P}, L) where \mathcal{P} is a nilpotent display over R and L is a direct summand of $\mathcal{D}_{\mathcal{P}}(R')$ lifting the Hodge filtration. A morphism $\phi : (\mathcal{P}_1, L_1) \rightarrow (\mathcal{P}_2, L_2)$ in \mathcal{C} is a morphism of nilpotent displays $\phi : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ such that $\mathcal{D}(\phi)(R')L_1 \subset L_2$.

The following result is the analogue of [M, V (1.6)].

PROPOSITION 2.5. — *The functor nilpotent display/ $R' \rightarrow \mathcal{C}$ given by $\mathcal{P}' = (P', Q', F, F_1) \mapsto (\mathcal{P}'_R, Q'/I_{R'} P')$ is an equivalence of categories.*

If R is a ring of characteristic p , then the morphism

$$\mathcal{F}_{\mathcal{P}} : \mathcal{P} \longrightarrow \mathcal{P}^{(p)}, \quad \mathcal{V}_{\mathcal{P}} : \mathcal{P}^{(p)} \longrightarrow \mathcal{P}$$

defines associated morphisms of the crystals. In particular $\mathcal{D}_{\mathcal{P}}$ is endowed with the structure of a Dieudonné crystal in the sense of [Gr1], [Gr2], [BBM].

Let $R' \xrightarrow{\pi} R$ be a divided power thickening where p is nilpotent in R' . If $J = \text{Ker}(\pi)$, then $J \subset W(R')$ has divided powers given by transport of structure via \log . $I_{R'}$ has divided powers and as $J + I_{R'} = J \oplus I_{R'}$, we obtain divided powers on this ideal. Hence $W(R') \xrightarrow{\pi \circ w_0} R$ is a (topological) divided power thickening, inducing divided power thickenings $W_n(R') \rightarrow R$.

Using the Cartier map $W(R) \xrightarrow{\delta} W(W(R))$, characterized by $w_n \circ \delta = f^n$, $n \in \mathbf{N}$, Zink proves:

PROPOSITION 2.6. — *Let \mathcal{P} be a nilpotent display. There is a canonical isomorphism*

$$\mathcal{K}_{\mathcal{P}}(R') \simeq \varprojlim \mathcal{D}_{\mathcal{P}}(W_n(R')).$$

If R has characteristic p , this isomorphism is compatible with Frobenius and Verschiebung.

PROPOSITION 2.7. — *Let R have characteristic p . Assume there is a topological divided power thickening $S \rightarrow R$ which is a flat \mathbf{Z}_p -algebra. Then the functor $\mathcal{P} \mapsto (\mathcal{D}_{\mathcal{P}}, Q/I_R P)$ from nilpotent displays to filtered Dieudonné crystals is fully-faithful in the following weak sense. Given a morphism ψ between the filtered Dieudonné crystals, there is a morphism of displays ϕ inducing ψ for every divided power thickening $R' \rightarrow R$ which receives a map from (some) S to R' .*

Remark 2.8. — The same result is true for the filtered Witt-crystals, the filtration being given by \widehat{Q} . Also one need only assume ψ is compatible with Frobenius. Finally, if $p \neq 2$, the same result holds for the nilpotent crystalline site.

3. THE FUNCTOR BT

Recall that an R -algebra N without identity is nilpotent provided $N^n = (0)$ for some positive integer n . There are many equivalent ways to define formal groups. For us we will regard a formal group G over R as a functor from nilpotent R -algebras to abelian groups satisfying:

- (i) if $0 \rightarrow N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow 0$ is exact (as a sequence of R -modules)

$$0 \longrightarrow G(N_1) \longrightarrow G(N_2) \longrightarrow G(N_3) \longrightarrow 0$$

is exact;

- (ii) if any R -module M is viewed as a nilpotent R -algebra ($M^2 = (0)$), then $\bigoplus_{i \in I} G(M_i) \xrightarrow{\sim} G(\bigoplus_{i \in I} M_i)$;

- (iii) $G(R)$ is a finitely generated projective module, R being viewed as having square zero. This is the tangent space, denoted $\text{Lie}(G)$.

We will define a functor $BT : \text{Displays}/R \rightarrow \text{Formal Groups}/R$.

If N is a nilpotent R -algebra, let $\widehat{W}(N) \subset W(N)$ be the $W(R)$ -subalgebra consisting of all Witt vectors almost all whose components are zero. Given a display $\mathcal{P} = (P, Q, F, F_1)$ over R , we set $G_{\mathcal{P}}^0(N) = \widehat{W}(N) \otimes_{W(R)} P$, $G_{\mathcal{P}}^{-1}(N) \subset G_{\mathcal{P}}^0(N)$ is the subgroup generated by $\{v(\xi) \otimes x, \xi \otimes y \mid \xi \in \widehat{W}(N), x \in P, y \in Q\}$. The functors $G_{\mathcal{P}}^{-1}(N)$ and $G_{\mathcal{P}}^0(N)$ are exact in the sense of condition (i) in our definition of a formal group.

LEMMA 3.1. — (i) *The map $F_1 : Q \rightarrow P$ extends to a map $G_{\mathcal{P}}^{-1}(N) \xrightarrow{F_1} G_{\mathcal{P}}^0(N)$ determined by $F_1(v(\xi) \otimes x) = \xi \otimes F(x)$, $F_1(\xi \otimes y) = f(\xi) \otimes F_1(y)$.*

(ii) *If N is equipped with nilpotent divided powers, the F_1 of (i) extends to a nilpotent endomorphism of $G_{\mathcal{P}}^0(N)$.*

(iii) *If i is the inclusion of $G_{\mathcal{P}}^{-1}(N)$ in $G_{\mathcal{P}}^0(N)$, $F_1 - i : G_{\mathcal{P}}^{-1}(N) \rightarrow G_{\mathcal{P}}^0(N)$ is injective.*

DEFINITION 3.2. — *Let $BT_{\mathcal{P}}$ be the functor on nilpotent R -algebras $BT_{\mathcal{P}}(N) = \text{coker}(F_1 - i)$.*

THEOREM 3.3. — (i) *The functor BT takes values in the category of formal groups and commutes with base change.*

(ii) *If N has nilpotent divided powers, there is a canonical isomorphism*

$$\exp : N \otimes_R \frac{P}{Q} \xrightarrow{\sim} BT_{\mathcal{P}}(N).$$

(iii) *If \mathcal{P} is a nilpotent display, $BT_{\mathcal{P}}$ is a p -divisible formal group.*

(iv) *If R has characteristic p , BT transforms $\mathcal{F}_{\mathcal{P}}$ and $\mathcal{V}_{\mathcal{P}}$ to Frobenius and Verschiebung.*

Recall that the Cartier ring C_R is by definition $\text{End}(\widehat{W})^0$ and that its elements may be uniquely expressed as $c = \sum_{n,m \geq 0} V^n [a_{n,m}] F^m$ where, for fixed n , $a_{nm} = 0$ for all but finitely many m . For $x \in \widehat{W}(N)$, $c \in C_R$ as above, we have $x \cdot c = \sum_{n,m \geq 0} v^m ([a_{n,m}] f^n(x))$. The Cartier module associated to a formal group G is the left C_R -module $\text{Hom}(\widehat{W}, G) = M(G)$.

PROPOSITION 3.4. — *Let \mathcal{P} be a display over R , $C_R \otimes_{W(R)} P \xrightarrow{\sim} \text{Hom}(\widehat{W}, G_{\mathcal{P}}^0)$ via $c \otimes z \mapsto (x \mapsto xc \otimes z)$. The map $G_{\mathcal{P}}^0 \rightarrow BT_{\mathcal{P}}$ induces a surjection $C_R \otimes_{W(R)} P \rightarrow M(BT_{\mathcal{P}})$ whose kernel is the C_R -submodule generated by*

$$\{F \otimes x - 1 \otimes Fx, V \otimes F_1 y - 1 \otimes y \mid x \in P, y \in Q\}.$$

Recall that if G/R is a p -divisible formal group, the crystalline Dieudonné theory associates to G a crystal as follows. For any divided power thickening $R' \rightarrow R$ with nilpotent divided powers and any p -divisible group G' lifting G , we consider the universal extension by a vector group

$$0 \longrightarrow \omega_{G^*} \longrightarrow E(G') \longrightarrow G' \longrightarrow 0.$$

By definition $\mathbf{D}(G)_{R'}$ is $\text{Lie}(E(G'))$ which is, up to canonical isomorphism, independent of G' . Because G is a formal group, this definition extends to the site consisting of divided power thickening $R' \xrightarrow{\pi} R$ where $\text{Ker}(\pi)$ is a nilpotent ideal [MM], [Z5].

The following is one of Zink’s main theorems.

THEOREM 3.5. — (i) *The functors $\mathcal{P} \mapsto \mathcal{D}_{\mathcal{P}}$ and $\mathcal{P} \mapsto \mathbf{D}(BT_{\mathcal{P}})$ from the category of nilpotent displays to the category of crystals (where the divided power ideals of thickenings are nilpotent) are canonically isomorphic. It transforms Hodge filtration to Hodge filtration.*

(ii) *If \mathcal{P} is a nilpotent display over R , $R' \xrightarrow{\pi} R$ has nilpotent kernel and G'/R' is a p -divisible group lifting $BT_{\mathcal{P}}$, then there is a nilpotent display \mathcal{P}' lifting \mathcal{P} and an isomorphism $BT_{\mathcal{P}'} \rightarrow G'$ lifting the identity.*

With the notation of (ii),

(iii) *if $\alpha : \mathcal{P} \rightarrow \widetilde{\mathcal{P}}$ is a morphism of nilpotent displays over R and \widetilde{G}' is a lift to R' of $BT_{\widetilde{\mathcal{P}}}$ with corresponding display $\widetilde{\mathcal{P}'}$, then α lifts to $\alpha' : \mathcal{P} \rightarrow \widetilde{\mathcal{P}'}$ if and only if $BT(\alpha)$ lifts to an homomorphism $G' \rightarrow \widetilde{G}'$.*

(iv) *The functor BT from nilpotent displays to p -divisible formal groups is faithful.*

Parts (ii), (iii) of Theorem 3.5 follow from Proposition 2.5 and [M] in the case where $R' \twoheadrightarrow R$ has kernel with nilpotent divided powers. As usual this extends inductively to the case where this kernel is a nilpotent ideal. For (iv), let $\alpha : \mathcal{P} \rightarrow \mathcal{P}'$ induce $a : BT_{\mathcal{P}} \rightarrow BT_{\mathcal{P}'}$. If $a = 0$, $\mathbf{D}(a) = 0$. But evaluating $\mathbf{D}(a)$ on the thickenings $W_n(R) \rightarrow R$ and passing to the projective limit, we find, by (i) and Proposition 2.6, that $\alpha = 0$.

4. PROPERTIES OF THE FUNCTOR BT FOR NILPOTENT DISPLAYS

All displays in this paragraph are nilpotent. Before stating the next result recall that a noetherian ring R is said to be a Nagata ring if, for every prime ideal \mathfrak{p} of R , the integral closure of R/\mathfrak{p} in a finite extension of $\text{Frac}(R/\mathfrak{p})$ is a finitely generated R/\mathfrak{p} -module. Examples of such rings include excellent rings. The concept of Nagata ring is discussed in [Bour, IX, § 4] and, of course, in E.G.A. IV, § 7.7, where these rings are called (noetherian and) universally Japanese⁽¹⁾.

PROPOSITION 4.1. — (i) *Let p be nilpotent in R and assume the nilradical of R is nilpotent. Then $BT : \text{nilpotent displays over } R \rightarrow p\text{-divisible formal groups}$ is fully faithful.*

(ii) *Assume R has characteristic p and admits a topological divided power thickening $S \xrightarrow{\pi} R$, i.e. $J = \text{Ker}(\pi)$ has divided powers and S is separated and complete for the linear topology defined by a sequence of sub-divided power ideals $J = J_1 \supset J_2 \supset \dots$ such that $\text{Ker}(S/J_n \rightarrow R)$ is a nilpotent ideal and S is flat as a \mathbf{Z}_p -algebra. Then BT is fully-faithful.*

THEOREM 4.2. — *Assume R is a Nagata ring which is separated and complete in the p -adic topology. Then BT is an equivalence of categories between nilpotent displays over R and p -divisible formal groups over R .*

Proof.— The theorem is proved in successive steps. As each category is the projective limit in the sense of \varprojlim of the categories relative to R/p^n , we may assume p is nilpotent in R . Then, by Proposition 4.1, we know BT is fully-faithful. Hence we need show that it is essentially surjective.

(i) Assume $R = k$ a field. Let $K = k^{p^{-\infty}}$. By classical Dieudonné theory the result is true over K . If G/k is a p -divisible formal group, let \mathcal{P}_K be a nilpotent display over K such that $BT(\mathcal{P}_K) = G_K$. If C is a Cohen ring for k we choose a map $C \rightarrow W(K)$ lifting the inclusion of k in K . Let $A = W(K) \otimes_C W(K)$, a p -torsion free ring such that $A/pA = K \otimes_k K$. If $S = \widehat{A}$, the p -adic completion, then Proposition 4.1 (ii) tells us that BT is fully-faithful over $K \otimes_k K$. Then \mathcal{P}_K is equipped by Proposition 4.1 (ii),

⁽¹⁾Gabber has told me that if V is a complete height 1 valuation ring with algebraically closed fraction field, then $V\{X_1, \dots, X_n\}[Y_1, \dots, Y_m]$ is universally Japanese, but of course not noetherian.

with descent data and, by Proposition 1.11, there is a nilpotent display \mathcal{P} over k which descends \mathcal{P}_K . Then $BT(\mathcal{P})$ is isomorphic to G .

(ii) R is an artin local ring. Let G/R be a p -divisible formal group. Let \mathcal{P}_k be a nilpotent display such that $BT(\mathcal{P}_k) = G_k$. By theorem 3.5 (ii), there is a nilpotent display \mathcal{P} lifting \mathcal{P}_k and an isomorphism $BT(\mathcal{P}) \rightarrow G$.

(iii) R is a complete local ring. By the result just used in case (ii) we may assume R is reduced. Let $G_n = G \otimes R/\mathfrak{m}^n$. Then $G_n = BT(\mathcal{P}_n)$ and let \mathcal{P} be the display corresponding to the system (\mathcal{P}_n) . We must show that \mathcal{P} is a nilpotent display. If $H = BT(\mathcal{P})$ we look at $M(H)$ described in Proposition 3.4, the Cartier module. We embed R in a finite product of algebraically closed fields and hence reduce to the case where $R = k$ is an algebraically closed field. Then $\mathcal{P} = \mathcal{P}_{\text{nil}} \oplus \mathcal{P}_{\text{ét}}$, where \mathcal{P}_{nil} is a nilpotent display and $P_{\text{ét}}$ has a basis $\{e_1, \dots, e_h\}$, $Q = P_{\text{ét}}$ and $F_1(e_i) = e_i$ for $i = 1, \dots, h$ and $M(BT(\mathcal{P}_{\text{ét}}))$ has a presentation $\bigoplus_{i=1}^h C_k \cdot \frac{e_i}{V e_i - e_i}$. But $V - 1$ is a unit in C_k so $M(BT(\mathcal{P}_{\text{ét}})) = (0)$ and $M(BT(\mathcal{P}_{\text{nil}})) = M(BT(\mathcal{P}))$. But $\text{height}(G) = \text{rank}(P_{\text{nil}})$ and $\text{height}(G) = \text{rank}(P)$. Then $\mathcal{P}_{\text{nil}} = \mathcal{P}$, finishing the case when R is a complete local ring.

(iv) R is a Nagata local ring. Let \widehat{R} be its completion. We may assume R is reduced. Then \widehat{R} and $\widehat{R} \otimes_R \widehat{R}$ are both reduced [Bour, IX, §4, théorème 3]. Given G/R , let $\widehat{\mathcal{P}}$ be a nilpotent display over \widehat{R} such that $BT(\widehat{\mathcal{P}}) = G_{\widehat{R}}$. Using Proposition 4.1 (i), Theorem 3.5 (iv), we see $\widehat{\mathcal{P}}$ is equipped with descent data and this gives the result just as in the case when R is a field.

(v) The general case. We may assume that the $h = \text{height}$ of G is constant on $\text{Spec}(R)$. Let $R' = \prod R_{\mathfrak{m}}$, the product running over all maximal ideals of R . Then to give a p -divisible formal group G of height h over R' is the same as giving for each \mathfrak{m} a p -divisible formal group $G_{\mathfrak{m}}$ of height h over each $R_{\mathfrak{m}}$. Similarly to give a display \mathcal{P} over R' , with $\text{rank}(P) = h$, is the same as giving displays $\mathcal{P}_{\mathfrak{m}}$ over each $R_{\mathfrak{m}}$, these each having $\text{rank}(P_{\mathfrak{m}}) = h$. Both statements follow because idempotents in $M_n(R')$ are given by families of idempotents in the $M_n(R_{\mathfrak{m}})$. If all the $\mathcal{P}_{\mathfrak{m}}$ are nilpotent displays, then \mathcal{P} is a nilpotent display also because the exponent of nilpotency for $V_{\mathfrak{m}}^{\sharp}$ is bounded above by h . As each $R_{\mathfrak{m}}$ is a Nagata local ring we conclude from Part (iv) that there is a nilpotent display \mathcal{P}' over R' such that $BT(\mathcal{P}') = G_{R'}$. We will be able to apply descent to finish the proof, provided we can show $R' \otimes_R R'$ is reduced (so to be able to invoke Proposition 4.1 (i), Theorem 3.5 (iv) again). For any ring A , let $Q(A)$ be its full ring of quotients. As R' is faithfully flat over R , $R' \otimes_R R' \hookrightarrow Q(R') \otimes_{Q(R)} Q(R')$. As R is noetherian and reduced, $Q(R) = \prod_{i=1}^n Q(R/\mathfrak{p}_i)$ if $\{\mathfrak{p}_1, \dots, \mathfrak{p}_n\}$ is the set of minimal prime ideals of R . The ring $Q(R') \otimes_{Q(R)} Q(R')$ then decomposes as the $\prod_{i=1}^n (Q(R') \otimes_{Q(R)} Q(R')) \otimes_{Q(R)} Q(R/\mathfrak{p}_i)$. As $Q(R') \hookrightarrow \prod Q(R_{\mathfrak{m}})$, we see that $R' \otimes_R R'$ embeds into the product of the rings

$$\prod_{\mathfrak{m}} Q\left(\frac{R'_{\mathfrak{m}}}{\mathfrak{p}_i}\right) \otimes_{Q(R/\mathfrak{p}_i)} \prod_{\mathfrak{m}} Q\left(\frac{R'_{\mathfrak{m}}}{\mathfrak{p}_i}\right), \quad i = 1, \dots, n.$$

This reduces us to showing that if K is a field and I any index set $K^I \otimes_K K^I$ is reduced. This is standard as products and tensor products of separable algebras are separable.

Remark 4.3. — In [Z5], Zink stated Theorem 4.2 only for p nilpotent in R and R an excellent local ring or R/pR of finite type over a field (which implies R is excellent). But his proof works, as we have seen, in general. Indeed the proof shows that all we need to assume is that if R is noetherian and reduced, for each localization $R_{\mathfrak{m}}$, $\widehat{R}_{\mathfrak{m}}$ and $\widehat{R}_{\mathfrak{m}} \otimes_{R_{\mathfrak{m}}} \widehat{R}_{\mathfrak{m}}$ are both reduced.

If R is a complete noetherian local ring which is normal, with fraction field K of characteristic zero and residue field of characteristic p , then for a nilpotent display \mathcal{P} over R , the Tate module $T_p(BT_{\mathcal{P}})$ can be described explicitly in terms of \mathcal{P} . Let \overline{K} be an algebraic closure of K , \overline{R} the integral closure of R in \overline{K} , $\overline{\mathfrak{m}}$ its maximal ideal. For $K \subset E \subset \overline{K}$ with $[E:K]$ finite, set $\widehat{W}(\mathfrak{m}_E) = \varprojlim \widehat{W}(\mathfrak{m}_E/\mathfrak{m}_E^n)$ and $\widehat{W}(\overline{\mathfrak{m}}) = \varinjlim \widehat{W}(\mathfrak{m}_E)$. Let $\overline{W}(\overline{\mathfrak{m}})$ be the p -adic completion of $\widehat{W}(\overline{\mathfrak{m}})$. Let $\widetilde{\mathfrak{m}}$ be the p -adic completion of $\overline{\mathfrak{m}}$. For \mathcal{P} a nilpotent display over R , let $G = BT_{\mathcal{P}}$ and define $\overline{G}_{\mathcal{P}}^0$ to be $\overline{W}(\overline{\mathfrak{m}}) \otimes_{W(R)} P$,

$$\overline{G}_{\mathcal{P}}^{-1} = \text{Ker} \left(\overline{W}(\overline{\mathfrak{m}}) \otimes_{W(R)} P \longrightarrow \widetilde{\mathfrak{m}} \otimes_R P/Q \right).$$

As G is p -divisible we may write $G = \varinjlim G_n$ where G_n is the kernel of p^n on G . The Tate module $T_p(G)$ is by definition $\text{Hom}(\mathbf{Q}_p/\mathbf{Z}_p, \varinjlim G_n(\overline{K}))$.

PROPOSITION 4.4. — *There is an exact sequence of $\text{Gal}(\overline{K}/K)$ -modules*

$$0 \longrightarrow T_p(G) \longrightarrow \overline{G}_{\mathcal{P}}^{-1} \xrightarrow{F_1-i} \overline{G}_{\mathcal{P}}^0 \longrightarrow 0.$$

5. DUALITY

We briefly sketch now the duality theory for nilpotent displays and corresponding p -divisible formal groups. From Zink’s perspective it is based upon a canonical homomorphism

$$\text{Bil}(\mathcal{P}, \mathcal{P}'; \mathcal{G}) \longrightarrow \text{Biext}(BT_{\mathcal{P}}, BT_{\mathcal{P}'}; \widehat{\mathbf{G}}_m).$$

We do not review here the formalism of biextensions referring the reader to Mumford’s original paper [Mu], Grothendieck’s geometric and homological version of the theory [Gr3] and Zink’s reformulation of these in his context [Z5]. Let us only say that we have, if A, B, C are abelian groups (in a topos) and we are given exact sequences

$$\begin{aligned} 0 &\longrightarrow K_1 \longrightarrow K_0 \longrightarrow B \longrightarrow 0 \\ 0 &\longrightarrow L_1 \longrightarrow L_0 \longrightarrow C \longrightarrow 0, \end{aligned}$$

an exact sequence

$$\begin{aligned} \mathrm{Hom}(K_0 \otimes L_0, A) &\longrightarrow \mathrm{Hom}(K_0 \otimes L_1, A) \times_{\mathrm{Hom}(K_1 \otimes L_1, A)} \mathrm{Hom}(K_1 \otimes L_0, A) \\ &\longrightarrow \mathrm{Biext}(B, C; A) \longrightarrow \mathrm{Biext}(K_0, L_0; A). \end{aligned}$$

We apply this formalism to the exact sequences

$$\begin{aligned} 0 &\longrightarrow G_{\mathcal{P}}^{-1} \xrightarrow{F_1-i} G_{\mathcal{P}}^0 \longrightarrow BT_{\mathcal{P}} \longrightarrow 0 \\ 0 &\longrightarrow G_{\mathcal{P}'}^{-1} \xrightarrow{F_1-i} G_{\mathcal{P}'}^0 \longrightarrow BT_{\mathcal{P}'} \longrightarrow 0 \end{aligned}$$

where $\mathcal{P}, \mathcal{P}'$ are displays. Mumford proved that $\mathrm{Biext}(G_{\mathcal{P}}^0, G_{\mathcal{P}'}^0, \widehat{\mathbf{G}}_m) = (0)$ so every biextension of $BT_{\mathcal{P}}, BT_{\mathcal{P}'}$, by $\widehat{\mathbf{G}}_m$ is defined by a pair of morphism

$$\begin{aligned} \alpha_1 &: G_{\mathcal{P}}^{-1} \otimes G_{\mathcal{P}'}^0 \longrightarrow \widehat{\mathbf{G}}_m \\ \alpha_2 &: G_{\mathcal{P}}^0 \otimes G_{\mathcal{P}'}^{-1} \longrightarrow \widehat{\mathbf{G}}_m \end{aligned}$$

agreeing on $G_{\mathcal{P}}^{-1} \otimes G_{\mathcal{P}'}^{-1}$.

Let $\alpha \in \mathrm{Bil}(\mathcal{P}, \mathcal{P}', \mathcal{G})$. We will associate to α a pair of such homomorphisms as follows. First, for N a nilpotent R -algebra we have

$$\alpha_N : G_{\mathcal{P}}^0(N) \otimes G_{\mathcal{P}'}^0(N) \longrightarrow \widehat{W}(N)$$

defined by $(\xi \otimes x) \otimes (\xi' \otimes x') \mapsto \alpha(x \otimes x')\xi\xi'$.

Next recall that the Artin-Hasse exponential defines an exact sequence

$$0 \longrightarrow \widehat{W}(N) \xrightarrow{v-\mathrm{id}} \widehat{W}(N) \xrightarrow{AH} \widehat{\mathbf{G}}_m \longrightarrow 0.$$

Here for $x \in \widehat{W}(N)$,

$$AH(x) = \left[\exp \left(\sum_{n=0}^{\infty} w_n(x) \frac{T^{p^n}}{p^n} \right) \right]_{T=1}.$$

Then with these notations we define

$$\begin{aligned} \alpha_1(y, x') &= AH(\alpha_N(F_1 y \otimes x')), \quad y \in G_{\mathcal{P}}^{-1}(N), \quad x' \in G_{\mathcal{P}'}^0(N) \\ \alpha_2(x, y') &= [AH(\alpha_N(x \otimes y'))]^{-1}, \quad x \in G_{\mathcal{P}}^0(N), \quad y' \in G_{\mathcal{P}'}^{-1}(N). \end{aligned}$$

The verification that α_1, α_2 agree on $G_{\mathcal{P}}^{-1} \otimes G_{\mathcal{P}'}^{-1}$ is easy. One can show that we obtain the same biextension if $F_1 y$ is replaced by y in the equation defining α_1 and, simultaneously, y' is replaced by $F_1 y'$ in the equation defining α_2 .

The category of nilpotent R -algebras is isomorphic to that of augmented R -algebras such that the augmentation ideal is nilpotent, Aug_R . We endow $(\mathrm{Aug}_R)^\circ$ with the fpqc topology and consider formal groups as abelian sheaves on this site. For any abelian sheaf F , we define a subsheaf F^+ by setting $F^+(X) = \mathrm{Ker}(F(X) \xrightarrow{F(\varepsilon)} F(\mathrm{Spec}(R)))$, where $\varepsilon : \mathrm{Spec}(R) \rightarrow X$ is the canonical section.

The homological formalism provides a canonical homomorphism $\mathrm{Biext}(B, C; A) \rightarrow \mathrm{Hom}(B, \underline{\mathrm{Ext}}^1(C, A))$. We apply this taking a nilpotent display \mathcal{P} , its dual display

\mathcal{P}^t and the tautological $\alpha \in \text{Bil}(\mathcal{P}^t, \mathcal{P}, \mathcal{G})$. This furnishes us with a homomorphism $BT_{\mathcal{P}^t} \xrightarrow{\chi} \underline{\text{Ext}}^1(BT_{\mathcal{P}}, \widehat{\mathbf{G}}_m)$.

PROPOSITION 5.1. — *The map χ defines an isomorphism of formal groups*

$$BT_{\mathcal{P}^t} \longrightarrow \underline{\text{Ext}}^1(BT_{\mathcal{P}}, \widehat{\mathbf{G}}_m)^+.$$

Remark. — In the case where \mathcal{P}^t is also a nilpotent display, this result was given in [MM], with a completely different proof using Illusie’s thesis.

PROPOSITION 5.2. — *Assume $\mathcal{P}, \mathcal{P}'$ are nilpotent displays and $(\mathcal{P}')^t$ is also a nilpotent display. Then the map $\text{Bil}(\mathcal{P}, \mathcal{P}', \mathcal{G}) \rightarrow \text{Biext}(BT_{\mathcal{P}}, BT_{\mathcal{P}'}; \widehat{\mathbf{G}}_m)$ is an isomorphism.*

6. DIEUDONNÉ DISPLAYS AND p -DIVISIBLE GROUPS

Assume R is a complete noetherian local ring with perfect residue field k of characteristic p . If $p = 2$, assume in addition that $2R = (0)$. We will modify the definition of a display so as to obtain an equivalence of categories

$$BT : \text{Dieudonné displays}/R \xrightarrow{\cong} p\text{-divisible groups}/R.$$

If R is artinian we will consider a subring $\widehat{W}(R) \subset W(R)$, stable under f and v , functorial in R and having $\widehat{W}(k) = W(k)$. Let \mathfrak{m} be the maximal ideal of R and let $x \in W(k)$. Let $y_n \in W(k)$ satisfy $f^n(y_n) = x$. Let $\tilde{y}_n \in W(R)$ be any lift of y_n . Then for $n \gg 0$, $f^n(\tilde{y}_n)$ is independent of the choice and the map $\delta : W(k) \rightarrow W(R)$ defined by $\delta(x) = f^n(\tilde{y}_n)$, n large, is a ring homomorphism which is a section of $W(R) \xrightarrow{\pi} W(k)$.

We consider the sequence

$$0 \longrightarrow W(\mathfrak{m}) \longrightarrow W(R) \xrightarrow{\pi} W(k) \longrightarrow 0.$$

Let $\widehat{W}(\mathfrak{m}) = \{x_0, \dots, x_i, \dots\} \in W(\mathfrak{m}) \mid \text{almost all } x_i = 0\}$. Then $\widehat{W}(\mathfrak{m})$ is stable under f and v and we define $\widehat{W}(R) = \{x \in W(R) \mid x - \delta\pi x \in \widehat{W}(\mathfrak{m})\}$. $\widehat{W}(\mathfrak{m})$ is an ideal in $W(R)$ and consequently $\widehat{W}(R)$ is a ring. This ring is stable under f and v . It is easy to see that $\widehat{W}(R)$ is a (non-noetherian) local ring which is separated and complete in the topology defined by its maximal ideal $\widehat{W}(\mathfrak{m}) + \delta(p \cdot W(k))$.

Now we define a Dieudonné display \mathcal{P} over R in exactly the same manner as we defined a display in Definition 1.2, replacing $W(R)$ by $\widehat{W}(R)$. Just as before there is a map $V^\sharp : P \rightarrow P^{(1)}$ and we have

$$F^\sharp \circ V^\sharp = p \cdot \text{id}_P, \quad V^\sharp \circ F^\sharp = p \cdot \text{id}_{P^{(1)}}.$$

We consider divided power thickenings $S \xrightarrow{\pi} R$ such that for each $a \in \text{Ker}(\pi) = J$, we have $\gamma_n(a) = 0$ for $n \gg 0$ (depending on a). We have an exact sequence

$$0 \longrightarrow \widehat{W}(J) \longrightarrow \widehat{W}(S) \longrightarrow \widehat{W}(R) \longrightarrow 0.$$

The condition imposed on the divided powers implies that \log induces an isomorphism

$$\log \widehat{W}(J) \xrightarrow{\sim} J^{(\mathbf{N})}.$$

With this nilpotency assumption on the divided powers, Lemma 2.1 and Theorem 2.4 extend to Dieudonné displays. Just as before this enables us to define the Witt and Dieudonné crystals associated to a Dieudonné display.

We have a “forgetful” functor Dieudonné displays $\xrightarrow{\mathcal{F}}$ displays, $P \mapsto W(R) \otimes_{\widehat{W}(R)} P$.

We say that a Dieudonné display \mathcal{P} is nilpotent if $\mathcal{F}(\mathcal{P})$ is nilpotent. The functor \mathcal{F} establishes an equivalence of categories between nilpotent Dieudonné displays and nilpotent displays. This follows easily from the fact that, for a divided power thickening $S \xrightarrow{\pi} R$, the functor

Dieudonné displays/ $S \longrightarrow$ Dieudonné displays/ $R +$ lifts of the Hodge filtration is an equivalence of categories.

PROPOSITION 6.1. — *The category of nilpotent Dieudonné displays is equivalent via a functor BT to the category of p -divisible formal group/ R . If A is a finite R -algebra (so a product of artinian local rings) and \mathcal{P} is such a nilpotent display with base change to A , $\mathcal{P}_A = (P', Q', F, F_1)$, then we have an exact sequence*

$$0 \longrightarrow Q' \xrightarrow{F_1-i} P' \longrightarrow BT_{\mathcal{P}}(A) \longrightarrow 0.$$

PROPOSITION 6.2. — *Let P be a finitely generated projective $\widehat{W}(R)$ -module and $\phi : P^{(1)} \rightarrow P$ (resp. $\phi : P \rightarrow P^{(1)}$). Then there is a direct summand P^{tm} of P (resp. a projective quotient $P^{\acute{e}t}$ of P) such that ϕ induces on P^{tm} (resp. on $P^{\acute{e}t}$) an isomorphism. Further, if M is any $\widehat{W}(R)$ -module equipped with an isomorphism $\psi : M^{(1)} \rightarrow M$ (resp. an isomorphism $\psi : M \rightarrow M^{(1)}$) and $\alpha : M \rightarrow P$ (resp. $\alpha : P \rightarrow M$) and $\phi \circ \alpha^{(1)} = \alpha \circ \psi$ (resp. $\alpha^{(1)} \circ \phi = \psi \circ \alpha$), then α factors uniquely through P^{tm} (resp. $P^{\acute{e}t}$).*

PROPOSITION 6.3. — *Let $\mathcal{P} = (P, Q, F, F_1)$ be a Dieudonné display.*

- a) $P = Q \iff V^\sharp$ is an isomorphism.
- b) $I_R P = Q \iff F^\sharp$ is an isomorphism.

We say \mathcal{P} is étale if a) holds and \mathcal{P} is of multiplicative type if b) holds.

PROPOSITION 6.4. — *Let \mathcal{P}/R be a Dieudonné display. There is a map $\mathcal{P} \rightarrow \mathcal{P}^{\acute{e}t}$ to an étale Dieudonné display which is universal with respect to morphisms from \mathcal{P} to étale Dieudonné displays. The map $P \rightarrow P^{\acute{e}t}$ is surjective and if P^{nil} is its kernel, then $\mathcal{P}^{\text{nil}} = (P^{\text{nil}}, P^{\text{nil}} \cap Q, F, F_1)$ is a nilpotent Dieudonné display.*

This proposition has a dual proposition which we do not state explicitly.

Our artin ring R is canonically a $W(k)$ -algebra. Let \bar{k} be an algebraic closure of k , $\Gamma = \text{Gal}(\bar{k}/k)$ and $\bar{R} = R \otimes_{W(k)} W(\bar{k})$, equipped with its continuous action of Γ . If H is a finitely generated free \mathbf{Z}_p -module endowed with a continuous, for its p -adic topology, action of Γ we set

$$P(H) = \left(\widehat{W}(\bar{R}) \otimes_{\mathbf{Z}_p} H \right)^\Gamma.$$

The natural map $\widehat{W}(\overline{R}) \otimes_{\widehat{W}(R)} P(H) \rightarrow \widehat{W}(\overline{R}) \otimes_{\mathbf{Z}_p} H$ is an isomorphism. Let $\mathcal{P}(H)$ be the étale Dieudonné display $(P(H), Q(H), F, F_1)$ where $P(H) = Q(H)$ and F_1 is induced by $f \otimes_{\mathbf{Z}_p} \text{id}_H$ on $\widehat{W}(\overline{R}) \otimes_{\mathbf{Z}_p} H$, $F = p \cdot F_1$.

Conversely, if \mathcal{P} is an étale Dieudonné display over R define $H(\mathcal{P})$ to be the kernel of the \mathbf{Z}_p -linear homomorphism.

$$F_1 - \text{id} : \widehat{W}(\overline{R}) \otimes_{\widehat{W}(R)} P \longrightarrow \widehat{W}(\overline{R}) \otimes_{\widehat{W}(R)} P.$$

PROPOSITION 6.5. — *These functors establish equivalences of categories between étale Dieudonné displays over R and continuous Γ -modules, finitely generated and free as \mathbf{Z}_p -modules.*

PROPOSITION 6.6. — *Let \mathcal{P} be a nilpotent Dieudonné display over R , $\overline{\mathcal{P}}$ the corresponding nilpotent Dieudonné display over \overline{R} and $C_{\overline{R}}$ the cokernel of $F_1 - i : \overline{Q} \rightarrow \overline{P}$. Then there is an isomorphism $\text{Hom}_{\Gamma}(H, C_{\overline{R}}) \xrightarrow{\sim} \text{Ext}(\mathcal{P}(H), \mathcal{P})$.*

This is proved using Galois cohomology to establish that $H^1(\Gamma, \text{Hom}_{\mathbf{Z}_p}(H, \overline{Q})) = (0)$.

Let G be a p -divisible group over R ; we wish to associate to it a Dieudonné display. As R is an artin local ring we have an exact sequence

$$0 \longrightarrow \widehat{G} \longrightarrow G \longrightarrow G^{\text{ét}} \longrightarrow 0$$

where $G^{\text{ét}}$ is étale and \widehat{G} is a p -divisible formal group.

Write $G^{\text{ét}} = BT(H) = \varinjlim p^{-n}H/H$ and $\widehat{G} = BT_{\widehat{\mathcal{P}}}$, where $\widehat{\mathcal{P}}$ is a nilpotent Dieudonné display.

PROPOSITION 6.7. — *There is a canonical isomorphism*

$$\text{Hom}_{\Gamma}(H, \widehat{G}(\overline{R})) \xrightarrow{\sim} \text{Ext}(BT(H), \widehat{G}).$$

THEOREM 6.8. — *There is an equivalence of categories*

$$\text{Dieudonné displays}/R \xrightarrow{\sim} p\text{-divisible groups}/R.$$

This follows from Propositions 6.1, 6.6, 6.7 and the fact that there are no non-trivial homomorphisms in either direction between étale and nilpotent Dieudonné displays (resp. étale and p -divisible formal groups over R).

Remark 6.9. — The following questions raised by Zink in [Z3] remain open.

(i) If G/R is a p -divisible group with associated Dieudonné displays \mathcal{P} , is the Dieudonné crystal associated to \mathcal{P} (canonically) isomorphic to that associated to G by crystalline Dieudonné theory? The answer is of course yes by Theorem 3.5 if G is a p -divisible formal group.

(ii) Does the equivalence between p -divisible groups and Dieudonné displays respect duality? Again this is true if G and G^* are both p -divisible formal groups.

(iii) The classification when R is a complete noetherian local ring is obtained by defining $\widehat{W}(R) = \varprojlim \widehat{W}(R/\mathfrak{m}^n)$ and Dieudonné displays in the obvious manner since p -divisible groups/ $R \xrightarrow{\sim} \varprojlim p$ -divisible groups R/\mathfrak{m}^n .

7. WINDOWS AND DIEUDONNÉ DISPLAYS

If K is a local field of characteristic zero with perfect residue field of characteristic p , then one knows how to classify p -divisible groups over the ring of integer \mathcal{O}_K . If the absolute ramification index $e \leq p - 2$, the maximal ideal \mathfrak{m}_K has nilpotent divided powers and [Gr1] explains how to do this using the filtered Dieudonné module. A more direct approach to this classification is given in [Fon1]. Breuil has extended this result if $p \neq 2$ to the case where e is arbitrary, [Br], see also the appendix of [K]. Zink has generalized Breuil’s result to give a classification of p -divisible groups over a local finite flat $W(k)$ -algebra S . When $S = \mathcal{O}_K$, Zink further shows that the category Breuil uses, strongly divisible modules, is naturally equivalent to the category of Dieudonné displays over \mathcal{O}_K . We indicate now Zink’s generalization.

Let R be a local ring with perfect residue field of characteristic $p \geq 3$. Assume there is a n such that $x^n = 0$ for all $x \in \mathfrak{m}$. If $S \xrightarrow{\pi} R$ is a divided power thickening and S satisfies the same hypotheses we have the $\log : W(J) \xrightarrow{\sim} J^{\mathbf{N}}$, $J = \text{Ker}(\pi)$. Let $\widetilde{W}(J) = \log^{-1}(J^{\mathbf{N}})$. Then $\widehat{W}(J) \subset \widetilde{W}(J)$ and they are equal if the $\gamma_m(x)$ of each $x \in J$ are zero, for $m \gg 0$, depending on x . Denote by $\widehat{W}(S)$ the subring of $W(S)$ generated by $\widehat{W}(S)$ (cf. §6) and $\widetilde{W}(J)$. If A is a p -adic local ring equipped with an homomorphism $A \xrightarrow{\pi} R$ and $\text{Ker}(\pi)$ has divided powers compatible with those on pA , we define $\widetilde{W}(A) = \varprojlim \widetilde{W}(A/p^n)$, the limit over those n such that $p^n = 0$ in R . Similarly we define $\widehat{W}(A)$.

DEFINITION 7.1. — (i) *A frame for R is a flat \mathbf{Z}_p -algebra A which is a p -adic ring equipped with a surjection $\pi : A \rightarrow R$ whose kernel has divided powers and an endomorphism $\sigma : A \rightarrow A$ which lifts Frobenius.*

(ii) *A Dieudonné frame for R is a frame for R which, in addition, satisfies $A/p^n A$, is for all n a local ring whose maximal ideal satisfies the nilpotence condition imposed on R and the Cartier map $A \xrightarrow{\delta} W(A)$ factors through $\widetilde{W}(A)$.*

(iii) *A Dieudonné window for a frame (A, σ) (resp. a Dieudonné frame) is a finitely generated projective A -module M a submodule M_1 which contains JM , $J = \text{Ker}(\pi)$, a σ -linear map $\Phi : M \rightarrow M$ such that M/M_1 is a projective R -module, $\Phi M_1 \subset pM$ and M is generated by $\Phi M \cup \frac{1}{p} \Phi M_1$.*

The hypothesis that $A \xrightarrow{\delta} W(A)$ factors through $\widetilde{W}(A)$ implies that the composite

$$A \xrightarrow{\delta} W(A) \xrightarrow{W(\pi)} W(R)$$

factors through $\widehat{W}(R)$. Hence we may associate to a Dieudonné window a Dieudonné display over R as follows

$$\begin{aligned} P &= \widehat{W}(R) \otimes_A M \\ Q &= \ker\left(\widehat{W}(R) \otimes_A M \longrightarrow M/M_1\right) \\ F(\xi \otimes x) &= f\xi \otimes \Phi x, \quad \xi \in \widehat{W}(R), \quad x \in M \\ F_1(\xi \otimes y) &= f\xi \otimes \frac{1}{p}\Phi y, \quad \xi \in \widehat{W}(R), \quad y \in M_1 \\ F_1(v\xi \otimes x) &= \xi \otimes \Phi x, \quad \xi \in \widehat{W}(R), \quad x \in M. \end{aligned}$$

THEOREM 7.2. — *Let R satisfy our hypotheses and (A, σ) be a Dieudonné frame for R . Then the functor Dieudonné A -windows \rightarrow Dieudonné displays/ R is an equivalence of categories.*

This theorem is established by constructing a quasi-inverse functor. To do this one associates to a Dieudonné display a crystal generalizing the discussion of § 6 in that we no longer require any nilpotence condition on divided power thickenings. Lemma 2.1 and Theorem 2.4 extend to this situation. Note if the divided powers on $J = \ker(S \xrightarrow{\pi} R)$ are not nilpotent it is essential that one works with $\widetilde{W}(S)$ and not $\widehat{W}(S)$ for proving Lemma 2.1 in this context. Then, if \mathcal{P} is a Dieudonné display over R and $\widetilde{\mathcal{P}}$ is a triple over A which lifts \mathcal{P} in the sense that $\widetilde{\mathcal{P}} = (\widetilde{P}, F, F_1)$ where \widetilde{P} is a finitely generated free $\widetilde{W}(A)$ -module, the window associated to \mathcal{P} is

$$M = A \otimes_{\widetilde{W}(A)} \widetilde{P}, \quad M_1 = \text{Ker}\left(M \longrightarrow R \otimes_{\widetilde{W}(R)} P \longrightarrow P/Q\right).$$

The crystal associated to \mathcal{P}/R depends only on $\overline{\mathcal{P}} = \mathcal{P}_{R/p}$. For this Dieudonné display we have

$$\mathcal{V}_{\overline{\mathcal{P}}} : \overline{\mathcal{P}}^{(p)} \longrightarrow \overline{\mathcal{P}}$$

and this induces $\Phi : M \rightarrow M$, σ -similinear.

Let S be a finite flat local $W(k)$ -algebra. Consider a presentation

$$0 \longrightarrow I \longrightarrow W(k)[T_1, \dots, T_d] \longrightarrow S \longrightarrow 0,$$

where each T_i has image in the maximal ideal of S .

Let $A_0 =$ the divided power envelope of I modulo p -torsion (*i.e.* if $K = \text{Frac}(W(k))$), the subring of $K[T_1, \dots, T_d]$ generated over $W(k)$ by $\{\frac{x^n}{n!}, x \in I\}$. We have a surjection $A_0 \twoheadrightarrow S$. Let $\sigma : W(k)[T_1, \dots, T_d] \rightarrow W(k)[T_1, \dots, T_d]$ be f semi-linear with $\sigma(T_i) = T_i^p$. Then σ leaves $I + (p)$ stable. If $J_0 = \text{Ker}(A_0 \rightarrow S)$, that is the divided power ideal, $J_0 + pA_0$ has divided powers and σ extends to A_0 , σ lifts Frobenius. Let A be the p -adic completion of A_0 . If $R = S/p^n S$ for some fixed n , then (A, σ) is a Dieudonné frame for R since one shows that $A \xrightarrow{\delta} W(A)$ factors through $\widetilde{W}(A)$. By Theorem 6.8 and 7.2, Dieudonné windows for (A, σ) classify p -divisible groups over R .

Passing to the limit over n we obtain:

THEOREM 7.3. — *The category of Dieudonné windows for the frame (A, σ) is equivalent to the category of p -divisible groups over S .*

8. FURTHER RESULTS

Zink has applied techniques from his theory of displays to give new and simpler proofs of a purity result for p -divisible groups. This result [dJO] states the following:

PROPOSITION 8.1 (de Jong–Oort). — *Let R be a noetherian local ring of dimension ≥ 2 , U the complement of the closed point in $\text{Spec}(R)$. If G/R is a p -divisible group which has constant Newton polygon over U , then G has constant Newton polygon over $\text{Spec}(R)$.*

Zink’s proof of this result uses, in part, ideas of Vasiu, who proved a more general result of [V].

Langer and Zink have developed a theory of the de Rham–Witt complex valid for X an arbitrary scheme over a $\mathbf{Z}_{(p)}$ -algebra R . This theory is related to, but not the same as the “absolute theory” of [HM]. It generalizes the classical theory of Bloch–Illusie, [Ill1]. If R is a ring with p nilpotent in R and X is an R -scheme there are complexes $W_n \Omega_{X/R}^\bullet$ of $W_n(\mathcal{O}_X)/W_n(R)$ differential graded algebras such that

$$W_n(\mathcal{O}_X) = W_n \Omega_{X/R}^0, \quad d : W_n(\mathcal{O}_X) \rightarrow W_n \Omega_{X/R}^1$$

satisfies $d(\gamma_n(v\xi)) = \gamma_{n-1}(v\xi)d(v\xi)$. There are algebra homomorphisms $F : W_{n+1} \Omega_{X/R}^\bullet \rightarrow W_n \Omega_{X/R}^\bullet$ and additive maps $V : W_n \Omega_{X/R}^\bullet \rightarrow W_{n+1} \Omega_{X/R}^\bullet$. If X/R is smooth, then $\mathbb{H}^*(X, W_n \Omega_{X/R}^\bullet)$ is canonically isomorphic to $H_{\text{crys}}^*(X/W_n(R))$. Passing to the inverse limit one defines $W\Omega_{X/R}^\bullet$. It has operators F, V, d satisfying the standard relations $FdV = d, Vd = dVp, dF = pFd$.

If X/R is proper and smooth, we have

$$\mathbb{H}^*(X, W\Omega_{X/R}^\bullet) \simeq H_{\text{crys}}^*(X/W(R)).$$

As an application we give the Zink–Langer construction of the display associated to an abelian scheme. Let X/R be an abelian scheme of relative dimension g . Let $P = H_{\text{crys}}^1(X/W(R))$, a projective $W(R)$ -module of rank $2g$. Consider the subcomplex of $W\Omega_{X/R}^\bullet$ obtained by replacing $W(\mathcal{O}_X)$ by $I_X = v((W(\mathcal{O}_X)))$. Denote this complex by $IW\Omega_{X/R}^\bullet$.

We have a commutative diagram

$$(*) \quad \begin{array}{ccccccc} I_X & \xrightarrow{d} & W\Omega_{X/R}^1 & \xrightarrow{d} & W\Omega_{X/R}^2 & \longrightarrow & \cdots \\ \downarrow V^{-1} & & \downarrow F & & \downarrow pF & & \\ W(\mathcal{O}_X) & \xrightarrow{d} & W\Omega_{X/R}^1 & \xrightarrow{d} & W\Omega_{X/R}^2 & \longrightarrow & \cdots \end{array}$$

Let Q be the $W(R)$ -module obtained as the $\mathbb{H}^1(X, IW\Omega_{X/R}^\bullet)$.

We have an exact sequence

$$0 \longrightarrow IW\Omega_{X/R}^\bullet \longrightarrow W\Omega_{X/R}^\bullet \longrightarrow \mathcal{O}_X \longrightarrow 0$$

where \mathcal{O}_X is viewed as a complex concentrated in degree zero.

Taking \mathbb{H}^1 of this sequence we find

$$0 \longrightarrow Q \longrightarrow P \longrightarrow \text{Lie}(X) \longrightarrow 0.$$

PROPOSITION 8.2. — *Let $F_1 : Q \rightarrow P$ be the f -linear map induced by the diagram (*). Then $\mathcal{P} = (P, Q, F, F_1)$ is a display over R .*

We end with a question and two comments:

1) Can the de Rham–Witt complex be used to directly construct the nilpotent display associated to a p -divisible formal group over a Nagata ring R . More generally we ask whether using crystalline techniques will allow us to associate a Dieudonné display to a p -divisible group over a complete noetherian local ring whose residue field is perfect of characteristic p .

2) Breuil has, extending work of Fontaine and Conrad, classified finite flat commutative p -group schemes over \mathcal{O}_K (notation as in §7) for $p \neq 2$. Crystalline Dieudonné theory is defined for such group schemes over any base where p is nilpotent and has good faithfulness properties if the base is of characteristic p and is a reasonably nice scheme, [BM2], [dJ1], [dJM]. Thus it seems reasonable to hope that a good theory of displays which will classify finite flat p -group schemes can be developed.

3) In [Z5], page 132, Zink says he “would expect” that BT defines an equivalence of categories between nilpotent displays and p -divisible formal groups over any noetherian R . Recently Lau, [L], has proven this conjecture of Zink without the noetherian hypothesis on R , requiring only that R is separated and complete for the p -adic topology. His proof uses the Grothendieck-Illusie theory, [Ill2], of deformations of truncated Barsotti-Tate groups.

REFERENCES

- [Ber] P. BERTHELOT – *Cohomologie cristalline des schémas de caractéristique $p > 0$* , Lecture Notes in Math. **407**, Springer-Verlag (1974).
- [BBM] P. BERTHELOT, L. BREEN, W. MESSING – *Théorie de Dieudonné cristalline II*, Lecture Notes in Math. **930**, Springer-Verlag (1982).
- [BM1] P. BERTHELOT, W. MESSING – *Théorie de Dieudonné cristalline I*, in Journées de géométrie algébrique de Rennes I, Astérisque **63** (1979), 17–37.
- [BM2] P. BERTHELOT, W. MESSING – *Théorie de Dieudonné cristalline III*, in The Grothendieck Festschrift I, Birkhäuser (1990), 173–247.
- [Bour] N. BOURBAKI – *Algèbre commutative*, Chapitres VIII et IX, Masson (1983).

- [Br] C. BREUIL – *Groupes p -divisibles, groupes finis et modules filtrés*, Annals of Math. **152** (2000), 489–549.
- [Car1] P. CARTIER – *Groupes formels associés aux anneaux de Witt généralisés*, C. R. Acad. Sci. Paris **265** (1967), 50–52.
- [Car2] P. CARTIER – *Modules associés à un groupe formel commutatif, courbes typiques*, C. R. Acad. Sci. Paris **265** (1967), 129–132.
- [Car3] P. CARTIER – *Seminar at the IHÉS*, Spring 1972.
- [dJ1] A.J. de JONG – *Finite locally free group schemes in characteristic p and Dieudonné modules*, Invent. Math. **114** (1993), 89–137.
- [dJ2] A.J. de JONG – *Crystalline Dieudonné module theory via formal and rigid geometry*, Publ. Math. IHÉS **82** (1995), 5–96.
- [dJM] A.J. de JONG, W. MESSING – *Crystalline Dieudonné theory over excellent schemes*, Bull. Soc. Math. France **127** (1999), 333–348.
- [dJO] A.J. de JONG, F. OORT – *Purity of the stratification by Newton polygons*, J. Amer. Math. Soc **13** (2000), 209–241.
- [Dem] M. DEMAZURE – *Lectures on p -divisible groups*, Lecture Notes in Math. **302**, Springer-Verlag (1972).
- [Fon1] J.-M. FONTAINE – *Groupes p -divisibles sur les corps locaux*, Astérisque **47–48** (1977).
- [Fon2] J.-M. FONTAINE – *Groupes finis commutatifs sur les vecteurs de Witt*, C. R. Acad. Sci. Paris **280** (1975), A 1423–1425.
- [Gr1] A. GROTHENDIECK – *Groupes de Barsotti–Tate et cristaux*, in Actes Congrès Intern. Math. Nice 1970, I, Gauthier-Villars, 431–436.
- [Gr2] A. GROTHENDIECK – *Groupes de Barsotti–Tate et cristaux de Dieudonné*, Sémin. Math. Sup. **45**, Presse de l’Université de Montréal (1974).
- [Gr3] A. GROTHENDIECK – *Exposé VII*, in Séminaire de géométrie algébrique 7, Groupes de monodromie et géométrie algébrique, Lecture Notes in Math. **288**, Springer-Verlag (1972).
- [Haz] M. HAZEWINDEL – *Formal groups and applications*, Academic Press (1978).
- [HM] L. HESSELHOLT, I. MADSEN – *On the de Rham–Witt complex in mixed characteristic*, Ann. Sci. Éc. Norm. Sup. **37** (2004), 1–43.
- [Ill1] L. ILLUSIE – *Complexe de de Rham–Witt et cohomologie cristalline*, Ann. Sci. Éc. Norm. Sup. **12** (1979), 501–661.
- [Ill2] L. ILLUSIE – *Déformations de groupes de Barsotti–Tate (d’après A. Grothendieck)*, in Séminaire sur les pinceaux arithmétiques : la conjecture de Mordell, Société Mathématique de France, Paris, Astérisque **127** (1985), 151–198.
- [K] M. KISIN – *Crystalline representations and F -crystals*, Algebraic geometry and number theory, Birkhäuser, Progress in mathematics **253** (2006), 459–496.
- [LZ1] A. LANGER, T. ZINK – *De Rham–Witt cohomology for a proper and smooth morphism*, Journal Inst. Math. Jussieu **3** (2004), 231–314.

- [LZ2] A. LANGER, T. ZINK – *De Rham-Witt cohomology and displays*, available on Zink’s web page.
- [L] E. LAU – *Displays and formal p -divisible groups*, submitted manuscript.
- [Laz] M. LAZARD – *Commutative formal groups*, Lecture Notes in Math. **443**, Springer-Verlag (1975).
- [MM] B. MAZUR, W. MESSING – *Universal extensions and one dimensional crystalline cohomology*, Lecture Notes in Math. **370**, Springer-Verlag (1974).
- [M] W. MESSING – *The crystals associated to Barsotti–Tate groups*, Lecture Notes in Math. **264**, Springer-Verlag (1972).
- [Mu] D. MUMFORD – *Bi-extensions of formal groups*, in *Algebraic Geometry*, Oxford University Press (1969), 307–322.
- [N] P. NORMAN – *An algorithm for computing local moduli of abelian varieties*, Ann. Math. **101** (1975), 499–509.
- [Ser] J-P. SERRE – *Groupes p -divisibles (d’après J. Tate)*, Sémin. Bourbaki 1966/67, exp. n° 318, Société Mathématique de France, Collection Hors Série **10** (1995), 73–86.
- [Ta] J. TATE – *p -divisible groups*, in *Proceedings of a Conference on local fields*, Driebergen 1966, Springer-Verlag (1967), 158–183.
- [V] A. VASIU – *Crystalline boundedness principle*, math. NT/0205199, to appear in Ann. Sci. Éc. Norm. Sup. (2) **39** (2006).
- [Z1] T. ZINK – *Cartiertheorie kommutativer formaler Gruppen*, Teubner-Texte zur Math. **68** (1984).
- [Z2] T. ZINK – *Cartiertheorie über perfecten Ringen I, II*, preprints Akad. Wissenschaften Berlin (1986).
- [Z3] T. ZINK – *A Dieudonné theory for p -divisible groups* in *Class field theory - its centenary and prospect*, Act. St. in Pure Math. **30** (2001), 139–160.
- [Z4] T. ZINK – *Windows for displays of p -divisible groups*, in *Moduli of abelian varieties*, Prog. in Math. **195**, Birkhäuser (2001), 491–518.
- [Z5] T. ZINK – *The display of a formal p -divisible group*, in *Cohomologies p -adiques et applications arithmétiques I*, Astérisque **278** (2002), 127–248.
- [Z6] T. ZINK – *Lectures on displays, formal groups, ...*, Lectures at Université de Paris XIII, March 2006.

William MESSING

School of Mathematics
 University of Minnesota
 206 Church St. S.E.
 Minneapolis, MN 55455 – U.S.A.
E-mail : messing@math.umn.edu

**THE WAVE MAP PROBLEM
SMALL DATA CRITICAL REGULARITY**
[after T. Tao]

by Igor RODNIANSKI

1. INTRODUCTION

The purpose of this paper is to describe the wave map problem

$$(1) \quad \begin{aligned} \square\phi &= -\phi(\partial_\alpha\phi \cdot \partial^\alpha\phi), \\ \phi|_{t=0} &= \phi_0, \quad \partial_t\phi|_{t=0} = \phi_1 \end{aligned}$$

where ϕ is a map $\phi : \mathbb{R}^{n+1} \rightarrow \mathbb{S}^{m-1} \subset \mathbb{R}^m$, and its analogs for other target manifolds, with a specific focus on the small data critical regularity results of T. Tao, contained in the following

THEOREM 1 ([30], [31]). — *Let $n \geq 2$ and $s > \frac{n}{2}$. The solution of the Cauchy problem (1) with initial data $(\phi_0, \phi_1) \in (\mathbb{S}^{m-1}, T\mathbb{S}^{m-1})$ in $(\dot{H}^s, \dot{H}^{s-1})$ can be extended uniquely to a global solution $(\phi(t), \partial_t\phi(t)) \in (\dot{H}^s, \dot{H}^{s-1})$ on \mathbb{R}^{n+1} provided that the initial data (ϕ_0, ϕ_1) has a sufficiently small $(\dot{H}^{\frac{n}{2}}, \dot{H}^{\frac{n}{2}-1})$ norm:*

$$\|\phi_0\|_{\dot{H}^{\frac{n}{2}}(\mathbb{R}^n)} + \|\phi_1\|_{\dot{H}^{\frac{n}{2}-1}(\mathbb{R}^n)} < \epsilon.$$

These results imply that in dimensions $n \geq 3$, despite the fact that the wave map problem is *supercritical* relative to a conserved energy and there exist solutions blowing up in finite time, its classical solutions with \mathbb{S}^{m-1} target can be extended globally in time as long as the initial data has a small scale-invariant $\dot{H}^{\frac{n}{2}}$ norm⁽¹⁾. In the critical dimension $n = 2$ the result is particularly exciting as it implies that a solution exists globally as long as it has a small energy.

⁽¹⁾Here and in what follows we will denote the initial data by $\phi[0] = (\phi_0, \phi_1)$ and will say that $\phi[0] \in H^s$ meaning $(\phi_0, \phi_1) \in H^s \times H^{s-1}$.

The problem (1) arises as an Euler-Lagrange equation corresponding (formally) to the critical points of the Lagrangian density:

$$(2) \quad \mathcal{L}[\phi] = \frac{1}{2}(\partial_\alpha\phi \cdot \partial_\beta\phi) m^{\alpha\beta},$$

where $m_{\alpha\beta}$ is the Minkowski metric on \mathbb{R}^{n+1} . The density $\mathcal{L}[\phi]$ gives rise to the Minkowski analog of the harmonic map problem on \mathbb{R}^n , in which the energy density is given by $\frac{1}{2}\nabla\phi \cdot \nabla\phi$ and the critical points, harmonic maps $\phi : \mathbb{R}^n \rightarrow \mathbb{S}^{m-1}$, satisfy the equation

$$\Delta\phi = -\phi(\nabla\phi \cdot \nabla\phi).$$

The equation (1) belongs to the more general class of wave map problems, in which ϕ is a map from an $(n + 1)$ -dimensional Lorentzian manifold (\mathcal{M}, g) to a Riemannian manifold (\mathcal{N}, h) . The map ϕ is a solution of the Euler-Lagrange equations:

$$(3) \quad D^\alpha\partial_\alpha\phi = 0,$$

corresponding to the Lagrangian density:

$$(4) \quad \mathcal{L}[\phi] = \frac{1}{2} h_{ij} (\partial_\alpha\phi^i\partial_\beta\phi^j) g^{\alpha\beta}.$$

Here $\{\phi^i\}$ denote local coordinates on \mathcal{N} . D is the pull-back of the Levi-Civita connection on $T\mathcal{N}$ to the bundle $\phi^*(T\mathcal{N})$. In terms of the local coordinates $\{\phi^i\}$ this pull-back connection acting on sections of $\phi^*(T\mathcal{N})$ reads:

$$(5) \quad D_\alpha = \nabla_\alpha + \bar{\Gamma}_{\alpha j}^k, \quad \bar{\Gamma}_{\alpha j}^k = \Gamma_{ij}^k(\phi)\partial_\alpha\phi^i,$$

where Γ_{ij}^k is the Christoffel symbol in the coordinates $\{\phi^i\}$ and ∇ is a covariant derivative on $T\mathcal{M}$. The wave-map equation (3) has the form:

$$(6) \quad \square_g\phi^k = -\Gamma_{ij}^k(\phi) g^{\alpha\beta}(\partial_\alpha\phi^i\partial_\beta\phi^j).$$

In particular in the case of a wave map problem from Minkowski space (\mathbb{R}^{n+1}, m) the map ϕ verifies the equation

$$(7) \quad \square\phi = -\Gamma(\phi)(\partial_\alpha\phi, \partial^\alpha\phi).$$

The wave map problem appears naturally in solid-state physics, theory of topological solitons, Quantum Field Theory and General Relativity:

Topological solitons: One of the simplest non-trivial models with *topological soliton* solutions is the $(2 + 1)$ dimensional Lorentz invariant $O(3)$ classical σ -model which is nothing else but a $(2+1)$ -dimensional wave map problem with \mathbb{S}^2 target. It arises in the study of a continuum limit of an isotropic anti-ferromagnet, [8]. Topological solitons in this model are the *static* solutions (harmonic maps from $\mathbb{R}^2 \rightarrow \mathbb{S}^2$ of the equation

$$\square\phi = -\phi(\partial_\alpha\phi \cdot \partial^\alpha\phi)$$

which minimize the energy (conserved under evolution)

$$E[\phi] = \int_{\mathbb{R}^2} (|\partial_t \phi|^2 + |\nabla_x \phi|^2) dx$$

in a given homotopy class. Such maps satisfy the Bogomol’nyi equation

$$\partial_i \phi = \pm \epsilon_{ij} \phi \times \partial_j \phi$$

and are thought to represent meta-stable particles, [1]. Here ϵ_{ij} is an anti-symmetric tensor in two dimensions. The important feature of this model, common to all $(2 + 1)$ -dimensional wave map problems, is its *criticality*. Both the equation and the conserved energy $E[\phi]$ are invariant under scaling transformations $\phi(t, x) \rightarrow \phi(\lambda t, \lambda x)$. The problem displays a fascinating interplay between the infinite dimensional wave map dynamics defined by (1) and a finite dimensional dynamics generated by restricting the full dynamics to the moduli space of static solutions (e.g. self-shrinking $(\lambda \rightarrow 0)$ of harmonic maps), see e.g. [20], ultimately leading to the existence of large data solutions of (1) blowing up in finite time, [23].

General Relativity: The wave map problem on a curved $(2 + 1)$ -dimensional background with an \mathbb{H}^2 target arises in the $U(1)$ symmetry reduction of the Einstein vacuum equations. In this case one starts with a (\mathbf{M}, \mathbf{g}) Lorentzian $(3 + 1)$ -dimensional manifold with Ricci curvature

$$\mathbf{R}_{\alpha\beta} = 0.$$

Under the assumption that (\mathbf{M}, \mathbf{g}) is invariant under the group action of $U(1)$ which orbits are space-like the metric \mathbf{g} can be decomposed

$$\mathbf{g} = e^{-2\gamma} g + e^{2\gamma} (\theta)^2$$

where g is a Lorentzian metric on a $(2 + 1)$ -dimensional manifold $\mathcal{N} = (\Sigma \times \mathbb{R})$ and $\theta = dx^3 + A_a dx^a$ with $a = 0, 1, 2$ local coordinates on \mathcal{N} and x^3 a coordinate along the orbit. The equations $\mathbf{R}_{a3} = 0$ (and the assumption of triviality of the first cohomology class of Σ) imply that

$$dA = \frac{1}{2} e^{-4\gamma} \star d\omega,$$

where \star is the Hodge dual relative to the metric g and a scalar function ω is called a twist potential. The equation $\mathbf{R}_{33} = 0$ implies that

$$\begin{aligned} \square_g \gamma + \frac{1}{2} e^{-4\gamma} g^{ab} \partial_a \omega \partial_b \omega &= 0, \\ \square_g \omega - 4g^{ab} \partial_a \omega \partial_b \gamma &= 0 \end{aligned}$$

which can be recognized as a wave map equation from $(\Sigma \times \mathbb{R}, g)$ into the hyperbolic space \mathbb{H}^2 with the metric

$$2(d\gamma)^2 + \frac{1}{2} e^{-4\gamma} (d\omega)^2.$$

Note that the wave map evolves on a dynamic background with the metric g , which itself depends on the wave map. This coupling is determined by satisfying the remaining

equations $\mathbf{R}_{ab} = 0$. The only result available in this fully nonlinear context is a small data global stability in the expanding direction statement for solutions with Σ a compact surface with genus greater than one, metric $g = -dt^2 + t^2\sigma$ with σ a metric of scalar curvature -1 on Σ and the wave map $\phi = 0$, see [6].

Acknowledgments: The author would like to thank Terry Tao for valuable comments on this paper.

2. SUMMARY OF QUESTIONS AND RESULTS FOR THE WAVE MAP PROBLEM FROM MINKOWSKI SPACE

Traditionally⁽²⁾, as the wave map equation is a hyperbolic evolution problem, one is interested in the questions of local and global in time existence and uniqueness of solutions, existence of solutions blowing up in finite time and stability of static or other “preferred”⁽³⁾ solutions. The wave map equation from Minkowski space is invariant under the scaling transformation $\phi(t, x) \rightarrow \phi(\lambda t, \lambda x)$, which also preserves the $\dot{H}^{\frac{n}{2}}$ Sobolev norm. On this basis and in view of a geometric nature of the problem, our experience suggests that we could expect⁽⁴⁾ that:

Local in time solutions exist and unique for any initial data $\phi[0] \in H^s$ with $s > n/2$.

Solutions with data with a small $\dot{H}^{\frac{n}{2}}$ -norm can be extended globally in time.

Large data classical solutions can be extended globally in time for the (2+1)-dimensional (*critical*) wave map problem, where the scale invariant space \dot{H}^1 coincides with a conserved energy space, at least in the case of a target manifold of negative curvature, in analogy with the harmonic map heat flow.

Large data classical solutions can be extended globally in time for the (1+1)-dimensional wave map problem, where the scale invariant space $\dot{H}^{\frac{1}{2}}$ is larger (subcritical) than the energy space.

Below we briefly (and incompletely, sometimes referring to just the final result) summarize known results (a good survey of the wave map problem is given in [34]):

⁽²⁾The connection of the wave problem to QFT and GR may present an additional set of questions.

⁽³⁾An example of such a solution is $\phi = \gamma(u)$ where γ is a geodesic on (\mathcal{N}, h) and u verifies the wave equation $\square u = 0$.

⁽⁴⁾Just on the basis of presented here “evidence” perhaps a more appropriate term here would be “hope” as in some other problems these expectations have not been yet fulfilled or simply turned out to be wrong. For the wave map problem these expectations are more grounded due to the referred to above geometric origin of the problem, which makes available various cancellation properties (e.g. the expression $\partial^\alpha \phi \cdot \partial_\alpha \phi$ is an example of a *null form* eliminating parallel interactions of free waves).

Existence and uniqueness of local in time solutions in H^s with $s > n/2$ is in [13], [15] and [11] in dimension $n = 1$.

Small data global existence in $\dot{H}^{\frac{n}{2}} \times \dot{H}^{\frac{n}{2}}$ in dimensions $n \geq 2$ is shown in [30], [31]. Extensions to other targets are in [14], [25], [21], [16], [17], [35].

Large data global existence for the $(1 + 1)$ -dimensional wave map is established in [9], [19].

Existence of large data solutions blowing up in finite time in dimensions $n \geq 3$ is shown in [24], [4].

Stability of a trivial constant wave map and geodesic wave maps is in [27] and stability of certain $(2 + 1)$ -dimensional spherically symmetric solutions is in [18].

For the critical $(2 + 1)$ -dimensional wave map problem existence of solutions blowing up in finite time was proved in [23] for the \mathbb{S}^2 target. The large data global existence result is conjectured for the \mathbb{H}^2 target.

We should also mention that good results have been obtained for the large data critical $(2 + 1)$ -dimensional wave map problem for solutions with additional *spherical* or *equivariant* symmetry assumptions. It was shown in [7] (for geodesically convex targets), [29] that large data global spherically symmetric solutions can be extended globally and uniquely in time.

The k -equivariant (co-rotational) solutions of the wave map problem are considered in the case when a target manifold is a surface of revolution. The results in [26] and [28] imply that a solution blows up in finite time only if the energy concentrates (in particular small energy implies regularity), blow-up can not occur at a self-similar rate and at the blow-up a harmonic map can be “bubbled off”. We note that in the case of the \mathbb{S}^2 target the equation for a k -equivariant wave map takes the form

$$\partial_t^2 u - \left(\partial_r^2 + \frac{1}{r}\partial_r\right)u + k^2 \frac{\sin(2u)}{2r^2} = 0$$

for a single scalar function u satisfying the boundary conditions $u(0) = 0$ and $u(\infty) = \pi$.

3. LOCAL SUBCRITICAL THEORY

We begin by discussing the framework for proving (subcritical) local existence and uniqueness results for the wave map problem (7) as it already contains major elements required for the (critical) small data global existence problem.

A general scheme for proving local existence and uniqueness in a Sobolev space H^s can be loosely described as follows:

Identify a space X with the property that X contains solutions of the homogeneous wave equation $\square w = 0$ with initial data in $w[0] \in H^s$.

Assume a priori that a solution ϕ belongs to X and

$$(8) \quad \|\phi\|_X \leq 2C_0$$

for some constant C_0 .

Express ϕ via a representation

$$(9) \quad \phi = W(t)\phi[0] - \square^{-1}(\Gamma(\phi)\partial^\alpha\phi \cdot \partial_\alpha\phi),$$

where $w = W(t)\phi[0]$ is a solution of the homogeneous wave equation $\square w = 0$ with initial $\phi[0]$ and $v = \square^{-1}F$ denotes a solution of the inhomogeneous problem $\square v = F$ with zero initial data at $t = 0$.

Show that

$$(10) \quad \|\square^{-1}(\Gamma(\phi_1)(\partial^\alpha\phi_2 \cdot \partial_\alpha\phi_3))\|_X \leq C$$

for arbitrary functions $\|\phi_i\|_X \leq 2C_0$.

In the energy method, in which space X is chosen to be $L_t^\infty H^s$, the representation (9) reads

$$(11) \quad \phi(t) = W(t)\phi[0] - \int_0^t W(t-s)(\Gamma(\phi)\partial^\alpha\phi \cdot \partial_\alpha\phi)(s) ds,$$

and the estimate (10) follows from the standard energy estimates for the wave equation and a choice of a small time interval $[0, T]$ gives local well-posedness in Sobolev spaces H^s with $s > \frac{n}{2} + 1$.

The Strichartz method is based on (11) and combines the energy estimates for the solution of the wave equation with the Strichartz estimates:

$$\|\phi\|_{L_t^q L_x^p} \lesssim \|\phi[0]\|_{\dot{H}^s} + \|\square\phi\|_{L_t^1 \dot{H}^{s-1}},$$

where

$$(12) \quad \begin{aligned} \frac{2}{q} &\leq (n-1)\left(\frac{1}{2} - \frac{1}{p}\right), \quad q \geq 2, \quad (n, q, p) \neq (3, 2, \infty), \\ \frac{1}{q} + \frac{n}{p} &= \frac{n}{2} - s. \end{aligned}$$

The space $X = L_t^\infty H^s \cap L_t^2 W_\infty^1$ for $n \geq 3$ and $X = L_t^\infty H^s \cap L_t^4 W_\infty^1$ for $n = 2$. Local existence and uniqueness of solutions with initial data in H^s with $s > \frac{n}{2} + \frac{1}{2}$ for $n \geq 3$ and $s > \frac{n}{2} + \frac{3}{4}$ for $n = 2$ follow by iterating the wave map equation in the spaces where the energy norm is complemented by the Strichartz norm $\|\partial\phi\|_{L_t^2 L_x^\infty}$ for $n \geq 3$ and $\|\partial\phi\|_{L_t^4 L_x^\infty}$ for $n = 2$, see [22].

The $H^{s,\delta}$ method⁽⁵⁾ is based on the iteration in the space $X = H^{s,\delta}$ with the norm

$$\|F\|_{H^{s,\delta}} = \|(1 + |\tau| + |\xi|)^s(1 + ||\tau| - |\xi||)^\delta \tilde{F}(\tau, \xi)\|_{L^2_{\tau,\xi}}$$

adapted to the symbol of \square (i.e. $\square H^{s,\delta} = H^{s-1,\delta-1}$). For $\delta > 1/2$ the space $H^{s,\delta}$ is smaller than the intersection of energy space $L_t^\infty H^s$ and the $L_t^q L_x^p$ Strichartz spaces consistent with H^s regularity. Iteration in $H^{s,\delta}$ leads to the local existence and uniqueness results in H^s with $s > n/2$, see [13], [15]. The key to this result is the algebra property

$$H^{s,\frac{1}{2}+} \cdot H^{s-1,-\frac{1}{2}+} \subset H^{s-1,-\frac{1}{2}+},$$

and the null form estimate

$$Q_0(H^{s,\frac{1}{2}+}, H^{s,\frac{1}{2}+}) \subset H^{s-1,-\frac{1}{2}+},$$

which both hold with $s > \frac{n}{2}$. Here $H^{s,\frac{1}{2}+}$ stands for the space $H^{s,\delta}$ with $1/2 < \delta < 1/2 + s - n/2$. The null form $Q_0(\phi, \phi) = \partial_\partial \phi \partial^\alpha \phi$ is precisely the expression arising in the nonlinear term of the wave map problem. It has the property that for two solutions of the homogeneous wave equation $\square \phi = \square \psi = 0$ we have

$$2Q_0(\phi, \psi) = \square(\phi\psi).$$

In particular Q_0 eliminates parallel interactions, between ϕ and ψ whose space-time Fourier transform lies on the cone $|\tau|^2 = |\xi|^2$. This special structure of the nonlinearity in the wave map problem is crucial for both the local existence and the small data global existence results in low dimensions.

The challenge in strengthening these results to obtain a global scale-invariant critical statement lied in the fact that it would require a scale-invariant homogeneous version of the space $H^{\frac{n}{2},\frac{1}{2}}$

$$\|F\|_{\dot{H}^{s,\delta}} = \|(|\tau| + |\xi|)^s (||\tau| - |\xi||)^\delta \tilde{F}(\tau, \xi)\|_{L^2_{\tau,\xi}}.$$

The space $\dot{H}^{\frac{n}{2},\frac{1}{2}}$ is not suitable (in fact it is not even well defined) for the solution of the critical wave map problem in particular in view of the *division* and *summation* problems one faces in the process of constructing (i.e. proving a priori estimates) the wave map via a representation (9)

$$\phi = W(t)\phi[0] - \square^{-1}(\Gamma(\phi)(\partial^\alpha \phi \cdot \partial_\alpha \phi)).$$

First we can modify the space $\dot{H}^{\frac{n}{2},\frac{1}{2}}$ by improving its summability properties relative to the distance to the cone $||\tau| - |\xi||$ while keeping the space scale-invariant. We assume

⁽⁵⁾In the context of the well-posedness theory for hyperbolic equations the $H^{s,\delta}$ spaces were first used in the work of Klainerman-Machedon [12]. They are closely connected with the $X^{s,b}$ spaces introduced and used by Bourgain for the Schrödinger and KdV equations in [2], [3].

that the space-time Fourier transform $\tilde{F}(\tau, \xi)$ has support in the region $|\xi| \approx 2^k$ and define

$$(13) \quad \|F\|_{\dot{H}_k^{s, \frac{1}{2}, 1}} = 2^{ks} \sum_{j \in \mathbb{Z}} 2^{\frac{j}{2}} \|m(2^{-j} \|\tau\| - |\xi|) \tilde{F}(\tau, \xi)\|_{L_{\tau, \xi}^2},$$

where m is a smooth bump function supported on the interval $[1/4, 4]$ and equal to one on $[1/2, 2]$. The division problem arises already at the level of attempting to implement the general existence scheme described above assuming that ϕ is a linear combination of *finitely* many $H_k^{\frac{n}{2}, \frac{1}{2}, 1}$ atoms. The summation problem requires handling square summable combinations of $H_k^{\frac{n}{2}, \frac{1}{2}, 1}$ pieces.

4. TATARU’S RESULT

In [32], [33] Tataru solved the division problem under the assumption that the initial has a small critical Besov norm $\dot{B}_{2,1}^{\frac{n}{2}}$ instead of a larger Sobolev space $\dot{H}^{\frac{n}{2}}$. The summation problem was avoided by putting together the $H_k^{\frac{n}{2}, \frac{1}{2}, 1}$ pieces in ℓ^1 . Such a space is only consistent with free waves with initial data in the smaller Besov space $\dot{B}_{2,1}^{\frac{n}{2}}$. The solution of the division problem requires enlarging the atomic space built on $H_k^{\frac{n}{2}, \frac{1}{2}, 1}$ atoms by adding a component which lies in a *null frame* space. Without going into much details, the null frame spaces are motivated by the following observation. Let w be a unit frequency solution of the homogeneous wave equation $\square w = 0$ whose spatial Fourier support lies in an angular sector Ω . For simplicity we may assume that w is a + wave and thus it can be represented as a superposition of the traveling waves

$$w = \int_{\mathbb{R}^n} e^{it|\xi| + ix \cdot \xi} w_0(\xi) d\xi = \int_{\Omega} \int_{\lambda \sim 1} e^{i\lambda(t + ix \cdot \omega)} w_0(\lambda\omega) \lambda^{n-1} d\lambda d\omega = \int_{\Omega} w_{\omega}(t + x \cdot \omega) d\omega,$$

where

$$w_{\omega}(s) = \int_{\lambda \sim 1} e^{is\lambda} w_0(\lambda\omega) \lambda^{n-1} d\lambda.$$

For a fixed ω let $u_{\omega} = t + x \cdot \omega$ denote a variable parametrizing the corresponding null (i.e. the length of the tangent vector $(1, \omega)$ with respect to the Minkowski metric is zero) direction and let x_{ω} denote the variables $(t - x \cdot \omega, x - x \cdot \omega)$. Then the traveling wave decomposition has the property that

$$(14) \quad \int_{\Omega} \|u_{\omega}\|_{L_{u_{\omega}}^2 L_{x_{\omega}}^{\infty}} \lesssim |\Omega|^{\frac{1}{2}} \|w_0\|_{L_x^2}$$

In addition for any $\theta \notin \Omega$

$$(15) \quad \|w\|_{L_{u_{\theta}}^{\infty} L_{x_{\theta}}^2} \lesssim \text{dist}(\theta, \Omega)^{-1} \|w_0\|_{L_x^2}$$

The iteration space⁽⁶⁾ X is composed from the X_k atoms, where each space X_k contains functions F with spatial Fourier support in the region $|\xi| \sim 2^k$ and

$$\|F\|_{X_k} = \inf \left(\|F_0\|_{H_k^{\frac{n}{2}, \frac{1}{2}, 1}} + \sum_{\ell \in \mathbb{Z}_+} \|F_\ell\|_{Y_{k,\ell}} \right),$$

where the infimum is taken with respect to all possible decompositions $F = F_0 + \sum_\ell F_\ell$, the functions F_ℓ are supported at distance $\leq 2^{k-2\ell}$ from the cone $\tau^2 = |\xi|^2$ in Fourier space. Each of the spaces $Y_{k,\ell}$ is also an atomic space, where

$$\|G\|_{Y_{k,\ell}} = 2^{\frac{n}{2}k} \inf \sum_{m=1}^{\infty} |a_m|,$$

where the infimum is taken with respect to the decompositions $G = \sum_{m=1}^{\infty} a_m G_m$ and each atom G_m , in addition to the above requirements on its Fourier support, is assumed to verify the following conditions. Let K_ℓ denote a collection of $(2^{2\ell(n-1)})$ spherical caps of size $2^{-\ell}$ covering the unit sphere of directions in ξ space and P_κ denote the associated projection in Fourier space on the cap $\kappa \in K_\ell$. Then we require that for each κ there exists an angle $\omega_\kappa \notin 2\kappa$ such that

$$\sum_{\kappa \in K_\ell} \left(2^{-2k} (\text{dist}(\omega_\kappa, \kappa))^{-2} \|\square P_\kappa G_m\|_{L_{u\omega_\kappa}^1 L_{x\omega_\kappa}^2}^2 + (\text{dist}(\omega_\kappa, \kappa))^2 \|P_\kappa G_m\|_{L_{u\omega_\kappa}^\infty L_{x\omega_\kappa}^2}^2 \right) \leq 1.$$

Thus constructed space X is shown to satisfy the following properties:

1. $X \subset C_t^0 \dot{B}_{2,1}^{\frac{n}{2}} \cap C_t^1 \dot{B}_{2,1}^{\frac{n}{2}-1}$ and X contains solutions of the homogeneous wave equation with $(\dot{B}_{2,1}^{\frac{n}{2}} \times \dot{B}_{2,1}^{\frac{n}{2}-1})$ initial data.
2. $X \cdot X \subset X$ and $X \cdot \square X \subset \square X$ – bilinear estimates.

5. TAO’S RESULT IN HIGHER DIMENSIONS

We now discuss the result contained in [30], in which the $(\dot{H}^{\frac{n}{2}}$ critical) small data global existence problem was completely solved for the wave map equation (1) on Minkowski space \mathbb{R}^{n+1} for the \mathbb{S}^{m-1} target manifold in dimensions $n \geq 5$. The solution of the problem is particularly elegant in this case as the iteration spaces are essentially as in the Strichartz method.

5.1. Tools

A consistent theme in [30] is micro-localization. In this context it simply means that all involved functions⁽⁷⁾ are decomposed into parts each oscillating with frequencies in a given dyadic interval and all interactions are viewed in terms of interactions between such individual pieces. The framework for this is provided by the Littlewood-Paley

⁽⁶⁾The description given below is somewhat imprecise as some of the spatial Fourier localizations should be in fact space-time Fourier localizations.

⁽⁷⁾A wave map $\phi : \mathbb{R}^{n+1} \rightarrow \mathbb{S}^{m-1}$ is viewed as an \mathbb{R}^m -valued function with $|\phi| = 1$.

theory and paradifferential calculus. According to it an arbitrary function $f(x)$ can be decomposed with the help of the Littlewood-Paley projections P_k :

$$f = \sum_{k \in \mathbb{Z}} f_k, \quad f_k = P_k f = \int_{\mathbb{R}^n} e^{ix \cdot \xi} \chi(2^{-k}|\xi|) \hat{f}(\xi) d\xi$$

where

$$\sum_{k \in \mathbb{Z}} \chi(2^{-k}r) = 1, \quad \forall r \neq 0$$

and χ is a smooth non-negative cut-off function supported on $[1/2, 2]$. We set

$$f_{<k} = \sum_{m < k} f_m, \quad f_{>k} = \sum_{m > k} f_m.$$

The following capture usefulness of such decompositions:

$$\begin{aligned} \|\ |\nabla|^s f_k \|_{L^p} &\sim 2^{ks} \|f_k\|_{L^p}, & \forall 1 \leq p \leq \infty, \\ \|f_k\|_{L^p} &\lesssim 2^{k(\frac{n}{q} - \frac{n}{p})} \|f_k\|_{L^q}, & \forall 1 \leq q \leq p \leq \infty. \end{aligned}$$

The first relation reflects the fact that f_k oscillates at the frequency 2^k , while the second is called the Bernstein inequality and is simply a prototype of the Sobolev inequality.

The product of two different Littlewood-Paley pieces $f_{k_1}^1$ and $f_{k_2}^2$ can be characterized as follows⁽⁸⁾:

If $k_1 > k_2$ then the Fourier support $f_{k_1}^1 f_{k_2}^2$ is essentially still contained in the dyadic region $|\xi| \sim 2^k$.

If $k_1 = k_2$ then the Fourier support $f_{k_1}^1 f_{k_2}^2$ is essentially still contained in the region $|\xi| \leq 2^k$.

The analysis requires a micro-local version of the Strichartz spaces S_k adapted to frequency localized functions and consistent with the critical $\dot{H}^{\frac{n}{2}}$ regularity. Define the norm in S_k by

$$(16) \quad \|\psi\|_{S_k} = \sup_{q,p} 2^{(\frac{1}{q} + \frac{n}{p})k} \left(\|\psi\|_{L_t^q L_x^p} + 2^{-k} \|\partial_t \psi\|_{L_t^q L_x^p} \right),$$

where the sup is taken over all Strichartz *admissible* exponents described in (12). The Strichartz estimates applied to a function ψ_k localized at the frequency 2^k imply:

$$(17) \quad \|\psi_k\|_{S_k} \lesssim \|\psi_k[0]\|_{\dot{H}^{\frac{n}{2}}} + 2^{\frac{n-2}{2}k} \|\square \psi_k\|_{L_t^1 L_x^2}$$

⁽⁸⁾For simplicity of exposition we will make no distinction between the relations $k_1 > k_2$ and $k_1 \gg k_2$.

In particular, one easily obtains the following strengthened version of the Strichartz estimates:

$$\left(\sum_{k \in \mathbb{Z}} \|\psi_k\|_{S_k}^2 \right)^{\frac{1}{2}} \lesssim \|\psi[0]\|_{\dot{H}^{\frac{n}{2}}} + \|\square\psi\|_{L_t^1 \dot{H}^{\frac{n}{2}-1}}.$$

A very useful notion introduced in [30] is that of a σ -envelope: $c = \{c_k\}_{k \in \mathbb{Z}} \in \ell_2$, which has the property that

$$c_k \leq 2^{\sigma|k-k'|} c_{k'}, \quad \forall k, k' \in \mathbb{Z}.$$

Any ℓ^2 sequence a lies under a σ -envelope c (for instance)

$$c_k = \sum_{k' \in \mathbb{Z}} 2^{-\sigma|k-k'|} a_{k'}.$$

In particular the sequence $\|\phi[0]_k\|_{\dot{H}^{\frac{n}{2}}}$ associated with initial data for the wave map ϕ lies under a σ -envelope c with the property that

$$\sum_{k \in \mathbb{Z}} c_k^2 < \epsilon^2.$$

5.2. The setup

Existence, uniqueness and propagation of regularity⁽⁹⁾ follow from proving estimates in the space $X = S(c)$ defined by the norm

$$\|\phi\|_{S(c)} = \sup_k c_k^{-1} \|\phi_k\|_{S_k}.$$

This means that assuming that

$$\|\phi_k\|_{S_k} \leq 2C_0 c_k$$

for some sufficiently large constant C_0 the result will follow if one can prove the stronger estimate

$$\|\phi_k\|_{S_k} \leq C_0 c_k.$$

As opposed to the local existence scheme one is not able to prove the estimate

$$\|\square^{-1} (\Gamma(\phi^1)(\partial^\alpha \phi^2 \cdot \partial_\alpha \phi^3))\|_{S(c)} \leq C_0$$

for functions $\|\phi^i\|_{S(c)} \leq 2C_0$. The problem requires renormalization!

⁽⁹⁾Restriction on the parameter σ limits propagation of regularity to the values of Sobolev exponent $n/2 < s < n/2 + \sigma$. It is not however standard that a propagation of regularity in a Sobolev space H^s with $s > n/2$ immediately implies the same statement in all higher Sobolev spaces.

5.3. Micro-linearization of the wave-map equation

The next step is to project the wave map equation

$$\square\phi = -\phi(\partial_\alpha\phi \cdot \partial^\alpha\phi)$$

on the frequency 2^k with the help of Littlewood-Paley projection P_k thus deriving an equation for ϕ_k

$$\square\phi_k = - \sum_{k_1, k_2, k_3} P_k(\phi_{k_1}(\partial_\alpha\phi_{k_2} \cdot \partial^\alpha\phi_{k_3})).$$

Using scale invariance each ϕ_k can be seen to satisfy the same equation as ϕ_0

$$(18) \quad \square\phi_0 = - \sum_{k_1, k_2, k_3 \sim 0} P_0(\phi_{k_1}(\partial_\alpha\phi_{k_2} \cdot \partial^\alpha\phi_{k_3}))$$

and it suffices to prove that

$$\|\phi_0\|_{S_0} \leq C_0 c_0.$$

(This is where the envelope idea becomes very helpful.)

Next one says that F is an acceptable error term if $\|F\|_{L_t^1 L_x^2} \leq \epsilon C_0^3 c_0$ and thus by Strichartz estimates (17) its contribution to the $\|\phi_0\|_{S_0}$ is less than the allowed $C_0 c_0$ and thus can be discarded.

The main contribution of the nonlinear term in (18) is identified by rewriting (18) in the form

$$(19) \quad \square\phi_0 = -2\phi_{<0}(\partial_\alpha\phi_{<0} \cdot \partial^\alpha\phi_0) + F,$$

and claiming that the term F is an acceptable error. Before tackling F one sets the following

Rules of the game: place the terms with a higher frequency in the norm requiring fewer⁽¹⁰⁾ derivatives (e.g. $L_t^2 L_x^4$) and terms with a lower frequency in the norm requiring more (e.g. $L_t^\infty L_x^\infty$). Almost all estimates will appear as if they have “extra room” and thus are not scale invariant. This is merely an illusion due to the fact that we measure low and high frequency relative to a fixed benchmark frequency. That is, a low frequency will mean $k < 0$ while the high frequency will mean that $k > 0$. The actual gain occurs only through relative ratios of frequencies of functions involved in the products and is entirely consistent with an overall scale invariance.

Using the paradifferential calculus rules for products described above, one sees that the error term essentially contains⁽¹¹⁾ low-high-high, high-high-high, high-low-high and

⁽¹⁰⁾The number of derivatives is reflected in the exponential factors placed in front of the Strichartz norms in the definition of S_k in (16). For example the $L_t^2 L_x^4$ norm “costs” only $1/2 + n/4$ derivatives while the $L_t^\infty L_x^\infty$ norm requires $n/2$ derivatives.

⁽¹¹⁾We will consistently ignore all the terms arising from the commutators with the Littlewood-Paley projections.

0-low-low interactions. For example, the low-high-high interaction can be handled as follows (with the help of the envelope properties):

$$\begin{aligned} \left\| \sum_{k_2 \sim k_3 > 0} \phi_{<0}(\partial_\alpha \phi_{k_2} \cdot \partial^\alpha \phi_{k_3}) \right\|_{L_t^1 L_x^2} &\lesssim \sum_{k_2 \sim k_3 > 0} \|\phi_{<0}\|_{L_t^\infty L_x^\infty} \|\partial_\alpha \phi_{k_2}\|_{L_t^2 L_x^4} \|\partial^\alpha \phi_{k_3}\|_{L_t^2 L_x^4} \\ &\lesssim \sum_{k_2 \sim k_3 > 0} 2^{-(k_2+k_3)\frac{n-2}{4}} \|\phi_{k_2}\|_{S_{k_2}} \|\phi_{k_3}\|_{S_{k_3}} \\ &\lesssim \epsilon C^2 c_0 \sum_{k_2 \sim k_3 > 0} 2^{-(k_2+k_3)\frac{n-2}{4} + \sigma k_2} \lesssim \epsilon C^2 c_0 \end{aligned}$$

provided that σ is sufficiently small. Similarly for the high-low-high interactions

$$\begin{aligned} \left\| \sum_{k_1 \sim k_2 > 0, k_3 < 0} \phi_{k_1}(\partial_\alpha \phi_{k_2} \cdot \partial^\alpha \phi_{k_3}) \right\|_{L_t^1 L_x^2} &\lesssim \sum_{k_1 \sim k_2, k_3 < 0} \|\partial_\alpha \phi_{k_3}\|_{L_t^\infty L_x^\infty} \|\phi_{k_1}\|_{L_t^2 L_x^4} \|\partial^\alpha \phi_{k_3}\|_{L_t^2 L_x^4} \\ &\lesssim \sum_{k_1 \sim k_2 > 0, k_3 < 0} 2^{k_3 - k_1} 2^{-(k_1+k_3)\frac{n-2}{4}} \\ &\quad \times \|\phi_{k_1}\|_{S_{k_1}} \|\phi_{k_2}\|_{S_{k_2}} \|\phi_{k_3}\|_{S_{k_3}} \lesssim \epsilon^2 C^3 c_0. \end{aligned}$$

The remaining cases can be treated in a similar fashion.

The principal term $\phi_{<0}(\partial_\alpha \phi_{<0} \cdot \partial^\alpha \phi_0)$ cannot be iterated away by means of Strichartz estimates (or even with the help of more sophisticated spaces and estimates) as can be seen from the following argument. The term $\phi_{<0}$ has to be placed in $L_t^\infty L_x^\infty$, which means that one needs an estimate

$$\|\partial_\alpha \phi_{<0} \cdot \partial^\alpha \phi_0\|_{L_t^1 L_x^2} \lesssim \|\phi_{<0}\|_{L_t^2 L_x^p} \|\phi_0\|_{L_t^2 L_x^r} \lesssim \left(\sum_{k < 0} \|\phi_k\|_{S_k}^2 \right)^{\frac{1}{2}} \|\phi_0\|_{S_0},$$

where $1/2 = 1/p + 1/r$. By a scaling argument the term with $\phi_{<0}$ requires $p \geq 2n$. The condition $1/2 = 1/p + 1/r$ then implies that $r \leq 2n/(n-1)$. However the admissible Strichartz exponents (q, r) with $q = 2$ lie in the range $[2\frac{n-1}{n-3}, \infty]$ inconsistent with $r \leq 2n/(n-1)$.

5.4. The renormalization procedure

One begins with an anti-symmetrization trick, used previously in the context of regularity theory for harmonic maps in [10], [5]. Recall that ϕ takes values in a unit sphere $\mathbb{S}^{m-1} \subset \mathbb{R}^m$ and write relative to the standard coordinates on \mathbb{R}^m

$$\phi_{<0}^i(\partial_\alpha \phi_{<0}^j \partial^\alpha \phi_0^j) = B_\alpha^{ij} \partial^\alpha \phi_0^j.$$

The $m \times m$ matrices B_α can be anti-symmetrized as follows. The expression

$$\partial_\alpha \phi_{<0}^i(\phi_{<0}^j \partial^\alpha \phi_0^j) = \partial_\alpha \phi_{<0}^i(\partial^\alpha \phi_{<0}^j \phi_0^j) - \partial_\alpha \phi_{<0}^i \partial^\alpha(\phi_{<0}^j \phi_0^j).$$

Both of the terms on the right hand side above are error terms similar to F_0 satisfying a good $L_t^1 L_x^2$ estimate. In the first term a derivative has been successfully transferred

to a lower frequency making in a 0-low-low term. On the other hand the expression

$$(\phi_{<0}^j \phi_0^j) = P_0(|\phi|^2) + P_0(\text{high-high})$$

can be identified, modulo terms contributing to an acceptable error term, with the projection of $|\phi|^2$ on the unit frequencies. Since $\phi \in \mathbb{S}^{m-1}$ we have $|\phi| = 1$ and $P_0(|\phi|^2) = 0$.

This argument allows one to replace

$$\phi_{<0}^i (\partial_\alpha \phi_{<0}^j \partial^\alpha \phi_0^j) = B_\alpha^{ij} \partial^\alpha \phi_0^j = A_\alpha^{ij} \partial^\alpha \phi_0^j,$$

where A_α are the anti-symmetric matrices

$$A_\alpha = \phi_{<0}(\partial_\alpha \phi_{<0}, \cdot) - \partial_\alpha \phi_{<0}(\phi_{<0}, \cdot)$$

and (\cdot, \cdot) is the standard scalar product on \mathbb{R}^m . This anti-symmetry property is crucial for the following renormalization procedure. The unit frequency part of the wave-map ϕ is replaced by a new dynamic variable

$$w = U\phi_0$$

with U an almost orthogonal matrix nonlinearly dependent on ϕ . The map w verifies the equation

$$\square w = -2UA_\alpha \partial^\alpha \phi_0 + 2\partial_\alpha U \partial^\alpha \phi_0 + \square U\phi_0$$

and this change of variables is motivated by the attempt to eliminate the troublesome term $A_\alpha \partial^\alpha \phi_0$ by setting

$$(20) \quad \partial_\alpha U = UA_\alpha.$$

Solubility of the above transport equations depends on the Frobenius condition⁽¹²⁾

$$(21) \quad \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta] = 0.$$

Given the explicit form of A_α the condition (21) is not satisfied. However,

$$\partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta] = [A_\alpha, A_\beta] \approx \partial\phi_{<0} \cdot \partial\phi_{<0}.$$

is quadratic in the derivatives of a low frequency part of ϕ (and thus in particular only of size ϵ^2). The equation (20) is then solved approximately with the defect giving rise to an acceptable error term. The construction is recursive with⁽¹³⁾

$$U_k := (P_k \phi(P_{<k} \phi, \cdot) - P_{<k} \phi(P_k \phi, \cdot)) U_{<k},$$

$$U_{<k} := I + \sum_{k' < k} U_{k'}$$

⁽¹²⁾Equation (20) would suggest that A is a trivial $O(n)$ connection, which of course requires that its curvature vanishes. It turns out indeed that A is a low frequency portion of the pull-back connection $\phi_*(\nabla)$ on the bundle $\phi_*(T\mathbb{S}^{m-1})$ over \mathbb{R}^{n+1} .

⁽¹³⁾For technical reasons the sum in the second definition should extend only to a finite large negative integer $-M$.

This construction can be motivated by the following argument. In the original equation for ϕ_0 the term $A_\alpha \partial_0^\alpha \phi_0$ can be replaced by $\partial_\alpha \Lambda \partial^\alpha \phi_0$ with

$$\Lambda = \sum_{k < 0} (P_k \phi(P_{<k} \phi, \cdot) - P_{<k} \phi(P_k \phi, \cdot))$$

at the expense of generating an error term with an acceptable $L_t^1 L_x^2$ bound: the difference between A_α and $\tilde{A}_\alpha = \partial_\alpha \Lambda$ involves terms where derivatives transferred to terms with lower frequencies. The equation (20) then reads

$$\partial_\alpha U = \partial_\alpha \Lambda U$$

with the solution given by

$$U = e^\Lambda = e^{\sum_{k < 0} (P_k \phi(P_{<k} \phi, \cdot) - P_{<k} \phi(P_k \phi, \cdot))} \approx$$

and thus if we set

$$U_{<k} = e^{\sum_{k' < k} (P_{k'} \phi(P_{<k'} \phi, \cdot) - P_{<k'} \phi(P_{k'} \phi, \cdot))}$$

we have

$$\begin{aligned} U_k &= \left(e^{(P_k \phi(P_{<k} \phi, \cdot) - P_{<k} \phi(P_k \phi, \cdot))} - I \right) e^{\sum_{k' < k} (P_{k'} \phi(P_{<k'} \phi, \cdot) - P_{<k'} \phi(P_{k'} \phi, \cdot))} \\ &\approx (P_k \phi(P_{<k} \phi, \cdot) - P_{<k} \phi(P_k \phi, \cdot)) U_{<k}. \end{aligned}$$

It remains to show that the remaining terms on the right hand side of the equation for w are acceptable error terms so that

$$(22) \quad \square w = F$$

and that the transformation U preserves the space S_0 . The latter follows from the estimates

$$\|U\|_{L_t^\infty L_x^\infty} \lesssim 1, \quad \|U\|_{L_t^\infty L_x^\infty}, \quad \|\partial_\alpha U\|_{L_t^\infty L_x^\infty} \lesssim \epsilon.$$

The standard⁽¹⁴⁾ Strichartz estimates applied to the equation (22) imply that

$$\|w\|_{S_0} \lesssim \|w[0]\|_{L^2} + C_0^3 \epsilon c_0 \lesssim \|\phi[0]_0\|_{L^2} + C_0^3 \epsilon c_0 \lesssim 2c_0$$

and the desired estimate for ϕ_0 follows.

6. TAO'S RESULT IN LOWER DIMENSIONS

The (critical) small data global existence result for the wave map problem with the \mathbb{S}^{m-1} target for the remaining dimensions $2 \leq n \leq 4$ is contained in [31]. Given the space constraints it is very difficult to do the justice to that paper here. The proof is a true tour de force. We will only give a (superficial) description of the spaces used in the proof.

⁽¹⁴⁾not quite, as the Fourier support of w is not supported on unit frequencies, since ϕ_0 has been distorted by U . This can be corrected however without too much difficulty.

As the problem is shifted to the lower dimensions the dispersion properties of the wave equation become weaker and one starts “losing” various key Strichartz estimates (e.g. $L_t^2 L_x^4$ in dimensions $n \leq 4$ and even the $L_t^2 L_x^\infty$ in dimensions $n \leq 3$). This eventually means that the renormalization procedure has to be combined with a refinement of ideas that led to the solution of the division problem. In particular the null structure of the nonlinearity, which did not have much effect on the higher dimensional problem, becomes crucial. In the iteration procedure the Strichartz based spaces are replaced by a combination of the $H^{\frac{n}{2}, \frac{1}{2}}$ and null frame spaces. The iteration space $X = S(c)$, associated with an envelope $c = \{c_k\}$, is built of the following parts:

$$\|\phi\|_{S(c)} = \|\phi\|_{L_t^\infty L_x^\infty} + \sup_k c_k^{-1} \|\phi\|_{S[k]}.$$

The first L^∞ component is very important for the algebra property and reflects the fact that the wave map $\phi \in \mathbb{S}^m$ and thus $|\phi| = 1$. The dyadic spaces $S[k]$ in turn are defined by the norm

$$\|\phi\|_{S[k]} := \|\nabla_{x,t}\phi\|_{L_t^\infty \dot{H}^{\frac{n}{2}-1}} + \|\nabla_{x,t}\phi\|_{\dot{H}_k^{\frac{n}{2}-1, \frac{1}{2}, \infty}} + \sup_{\pm} \sup_{\ell > 0} \left(\sum_{\kappa \in K_\ell} \|P_{k, \pm \kappa} Q_{<k-2\ell}^\pm \phi\|_{S[k, \kappa]}^2 \right)^{\frac{1}{2}}.$$

The first term on the right represents the usual energy space. The second is an $H^{s, \delta}$ type space where the index ∞ reflects an ℓ^∞ norm with respect to a dyadic distance to the cone $|\tau|^2 = |\xi|^2$. The last ingredient is a null frame space built with the help of the Fourier projections $P_{k, \pm \kappa}$ and Q_j^\pm restricting the Fourier transform of a function to the region of (τ, ξ) with $|\tau \pm |\xi|| \sim 2^j$ (coming from the Q_j action) $|\xi| \sim 2^k$ (coming from P_k part) and a spherical cap $\xi/|\xi| \in \kappa$ of size $2^{-\ell}$ for $\kappa \in K_\ell$. Finally the space $S[k, \kappa]$ is defined by the norm

$$\|\phi\|_{S[k, \kappa]} = 2^{\frac{nk}{2}} \|\phi\|_{NFA^*[\kappa]} + |\kappa|^{-\frac{1}{2}} 2^{\frac{k}{2}} \|\phi\|_{PW[\kappa]} + 2^{\frac{nk}{2}} \|\phi\|_{L_t^\infty L_x^2},$$

where

$$\|\phi\|_{NFA^*[\kappa]} = \sup_{\omega \notin 2\kappa} \text{dist}(\omega, \kappa) \|\phi\|_{L_{u\omega}^\infty L_{x\omega}^2},$$

is motivated by the property (15) of free waves, and the space $PW[\kappa]$ is an atomic Banach space whose atoms are functions ϕ with

$$\|\phi\|_{L_{u\omega}^2 L_{x\omega}^\infty} \leq 1$$

for some $\omega \in \kappa$ and its definition is motivated by the property (14).

One of the important new ingredients is an appearance of a true trilinear estimate which deals with the nonlinearity $\phi(\partial_\alpha \phi \cdot \partial^\alpha \phi)$ and provides an exponential gain in the ratio of frequencies in the case where the frequency of the first term is larger than one of the other two frequencies.

7. EXTENSIONS TO OTHER TARGET MANIFOLDS

The work in [30], [31] has already had a serious impact on the field. In particular a lot of effort has been concentrated on the extension of the (critical) small data global existence result to other target manifolds.

In a more intrinsic interpretation of the wave map problem it is cast as a system of equations for the derivatives of the wave map ϕ . For a given orthonormal e_a on (\mathcal{N}, h) denote

$$\phi_\alpha^a = h(\partial_\alpha \phi, e_a).$$

We set $A_{b\alpha}^a = h(\nabla_{\phi^*(\partial_\alpha)} e_a, e_b)$ to be the pull-back of the Levi-Civita connection ∇ on (\mathcal{N}, h) to the bundle $\phi^*(T\mathcal{N})$, represented by anti-symmetric matrices A_α . Then if ϕ is a wave map the components $\Phi = (\phi_\alpha^a)$ satisfy the equation

$$(23) \quad \square \Phi = -2A_\alpha \cdot \partial^\alpha \Phi + E,$$

where E is a term cubic in Φ . The problem (23) can be micro-linearized similar to (19). The low frequency of the connection A_α is split with the help of a Hodge decomposition and a Coulomb gauge. Its gradient part then renormalized following Tao's approach. This led to an extension of Tao's result to a large class of target manifolds in dimensions $n \geq 5$ in [14]. Intuitively, in this picture, the procedure of a Hodge decomposition and renormalization corresponds to a choice of an orthonormal frame e_a with the property that the connection A_α , $\alpha = 0, \dots, n$ satisfies the Coulomb gauge condition $\sum_{i=1}^n \partial^i A_i = 0$. This was made even more explicit in [25], where Tao's work was extended to more general targets in dimensions $n \geq 4$. The geometric choice of a global Coulomb gauge avoided a micro-linearization of the equation. Similar extension for $n \geq 4$ was obtained in [21]. In [16], [17] the results were extended to general targets in dimensions $n \geq 3$ and a hyperbolic space \mathbb{H}^2 in dimension $n \geq 2$. In [35] the result was extended to targets isometrically embedded (with bounded geometry) into \mathbb{R}^m in dimensions $n \geq 2$.

REFERENCES

- [1] A.A. Belavin, A.M. Polyakov, *Metastable states of two-dimensional isotropic ferromagnets*, JETP Lett. **22** (1975), 245–247 (Russian).
- [2] J. Bourgain, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. I. Schrödinger equations*, Geom. Funct. Anal. **3** (1993), no. 2, 107–156.
- [3] J. Bourgain, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations. II. The KdV-equation*, Geom. Funct. Anal. **3** (1993), no. 3, 209–262.

- [4] T. Cazenave, J. Shatah, S. Tahvildar-Zadeh, *Harmonic maps of the hyperbolic space and development of singularities in wave maps and Yang-Mills fields*, Ann. I.H.P., section A **68** (1998), no. 3, 315–349.
- [5] S.Y.A Chang, L. Wang, P. Yang, *Regularity of harmonic maps*, Comm. Pure Appl. Math. **52** (1999), 1099–1111.
- [6] Y. Choquet-Bruhat, *Future complete Einsteinian space times with $U(1)$ symmetry, the unpolarized case*, The Einstein equations and the large scale behavior of gravitational fields, 251–298, Birkhäuser, Basel, 2004.
- [7] D. Christodoulou, S.A. Tahvildar-Zadeh, *On the regularity of spherically symmetric wave maps*, Comm. Pure Appl. Math. **46** (1993), no. 7, 1041–1091.
- [8] E. Fradkin, *Field theories of condensed matter systems*, Addison-Wesley, 1991.
- [9] C. Gu, *On the Cauchy problem for harmonic maps defined on two-dimensional Minkowski space*, Comm. Pure Appl. Math. **33** (1980), 727–737.
- [10] F. Hélein, *Régularité des applications faiblement harmoniques entre une surface et une variété riemannienne*, C.R. Acad. Sci. Paris Sér. I Math. **312** (1991), 591–596
- [11] M. Keel, T. Tao, *Local and global well-posedness of wave maps in R^{1+1} for rough data*, IMRN **21** (1998), 1117–1156.
- [12] S. Klainerman, M. Machedon, *Space-time estimates for null forms and the local existence theorem*, Comm. Pure Appl. Math. **46** (1993), no. 9, 1221–1268.
- [13] S. Klainerman, M. Machedon, *Smoothing estimates for null forms and applications*, Duke Math J. **81** (1995), 99–103.
- [14] S. Klainerman, I. Rodnianski, *On the global regularity of wave maps in the critical Sobolev norm*, Internat. Math. Res. Notices (2001), no. 13, 655–677.
- [15] S. Klainerman, S. Selberg, *Remark on the optimal regularity for equations of wave maps type*, C.P.D.E. **22** (1997), 901–918.
- [16] J. Krieger, *Global regularity of wave maps from R^{3+1} to surfaces*, CMP **238** (2003), 333–366.
- [17] J. Krieger, *Global regularity of wave maps from R^{2+1} to H^2 . Small energy*, Comm. Math. Phys. **250** (2004), no. 3, 507–580.
- [18] J. Krieger, *Stability of Spherically Symmetric Wave Maps*, Mem. Amer. Math. Soc. 181 (2006), no. 853.
- [19] O.A. Ladyzhenskaya, V.I. Shubov, *Unique solvability of the Cauchy problem for the equations of the two dimensional chiral fields, taking values in complete Riemann manifolds*, J. Soviet Math. **25** (1984), 855–864
- [20] N. Manton, P. Sutcliffe, *Topological solitons*, Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, 2004
- [21] A. Nahmod, A. Stefanov, K. Uhlenbeck, *On the well-posedness of the wave map problem in high dimensions*, Comm. Anal. Geom. **11** (2003), 49–83.
- [22] G. Ponce, T. Sideris, *Local regularity of nonlinear wave equations in three space dimensions*, Comm. PDE. **18** (1993), 169-177.

- [23] I. Rodnianski, J. Sterbenz, *On the Formation of Singularities in the Critical $O(3)$ σ -Model*, preprint.
- [24] J. Shatah, *Weak solutions and development of singularities in the $SU(2)$ σ -model*, Comm. Pure Appl. Math. **41** (1988), 459–469.
- [25] J. Shatah, M. Struwe, *The Cauchy problem for wave maps*, IMRN **11** (2002) 555–571.
- [26] J. Shatah, S.A. Tahvildar-Zadeh *On the Cauchy problem for equivariant wave maps*. Comm. Pure Appl. Math. **47** (1994), no. 5, 719–754.
- [27] T. Sideris, *Global existence of harmonic maps in Minkowski space*, Comm. Pure Appl. Math. **42** 1 (1989), 1–13.
- [28] M. Struwe, *Equivariant wave maps in two space dimensions*, Comm. Pure Appl. Math. **56** (2003), 815–823.
- [29] M. Struwe, *Radially symmetric wave maps from $(1 + 2)$ -dimensional Minkowski space to general targets*, Calc. Var. **16** (2003), 431–437.
- [30] T. Tao, *Global regularity of wave maps I. Small critical Sobolev norm in high dimension*, IMRN **7** (2001), 299–328.
- [31] T. Tao, *Global regularity of wave maps II. Small energy in two dimensions*, Comm. Math. Phys. **224** (2001), 443–544.
- [32] D. Tataru, *Local and global results for wave maps I*, CPDE **23** (1998), 1781–1793.
- [33] D. Tataru, *On global existence and scattering for the wave maps equation*, AJM **123** (2001), 37–77.
- [34] D. Tataru, *The wave maps equation*, Bull. Amer. Math. Soc. (N.S.) **41** (2004), no. 2, 185–204.
- [35] D. Tataru, *Rough solutions for the wave maps equation*, Amer. J. Math. **127** (2005), no. 2, 293–377.

Igor RODNIANSKI

Department of Mathematics

Princeton University

PRINCETON, NJ 08544 – U.S.A.

E-mail : irod@math.princeton.edu

**ENSEMBLES DE JULIA DE MESURE POSITIVE ET
DISQUES DE SIEGEL DES POLYNÔMES QUADRATIQUES**

[d'après X. Buff et A. Chéritat]

par Jean-Christophe YOCCOZ

1. ENSEMBLES DE JULIA DE MESURE POSITIVE

1.1. Soit R une fraction rationnelle de degré $d \geq 2$ opérant sur la sphère de Riemann $\overline{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$.

Le multiplicateur d'un point périodique z_0 de R , de période minimale n , est $\lambda = (R^n)'(z_0)$. On dit que z_0 est attractif si $|\lambda| < 1$, répulsif si $|\lambda| > 1$, indifférent si $|\lambda| = 1$. Un point périodique indifférent est parabolique si $\lambda = 1$, rationnel si λ est racine de l'unité, irrationnel dans le cas contraire.

L'ensemble de Fatou $F(R)$ est l'ensemble des points $z \in \overline{\mathbf{C}}$ au voisinage desquels les itérés $(R^n)_{n \geq 1}$ forment une famille normale. Son complémentaire dans $\overline{\mathbf{C}}$ est l'ensemble de Julia $J(R)$.

L'ensemble de Fatou est ouvert et totalement invariant : $R^{-1}(F(R)) = F(R)$. L'image par R d'une composante connexe de $F(R)$ est encore une composante connexe de $F(R)$. Un célèbre théorème de D. Sullivan affirme que chaque composante est prépériodique sous l'action de R . Il y a au plus $2d - 2$ cycles de composantes périodiques. La dynamique dans une composante fixe est d'un des 3 types suivants :

- convergence vers un point fixe attractif ;
- convergence vers un point fixe parabolique situé sur le bord de la composante ;
- dynamique quasi périodique conjuguée à une rotation irrationnelle sur un disque (dit de Siegel) ou un anneau (dit de Herman).

L'ensemble de Julia est une partie compacte, totalement invariante, qui n'est jamais vide. C'est l'adhérence de l'ensemble des points périodiques répulsifs de R . La dynamique y est chaotique, au moins topologiquement, dans le sens suivant : pour tout ouvert U rencontrant $J(R)$, il existe $N \geq 1$ tel que $R^N(U)$ contienne $J(R)$.

L'ensemble de Fatou peut être vide. C'est le cas, en particulier, pour les exemples de Lattès déduits via la fonction de Weierstrass des endomorphismes non inversibles des courbes elliptiques. M. Rees a montré que dans l'espace (de dimension complexe $2d - 2$) des fractions rationnelles de degré d , celles pour lesquelles l'ensemble de Fatou est vide

forment un ensemble de mesure de Lebesgue positive. Cependant, lorsque l'ensemble de Julia n'est pas égal à la sphère de Riemann toute entière, c'est un ensemble d'intérieur vide.

Pour un polynôme P de degré $d \geq 2$, la présence d'un point fixe (super)attractif à l'infini rend les choses un peu plus simples : l'ensemble de Fatou n'est jamais vide. On appelle ensemble de Julia rempli l'ensemble $K(P)$ des points $z \in \mathbf{C}$ d'orbite bornée. C'est une partie compacte totalement invariante de \mathbf{C} dont le bord est égal à $J(P)$.

1.2. Dès le début du siècle dernier, P. Fatou suggère d'étudier les ensembles de Julia par les méthodes de la théorie de la mesure de Borel-Lebesgue. Vers 1980, D. Sullivan développe et met à profit des analogies profondes entre l'itération des fractions rationnelles et l'action des groupes kleinien (i.e. des sous-groupes discrets de type fini de $PSL(2, \mathbf{C})$) sur la sphère de Riemann. Pour un groupe kleinien, l'ensemble limite joue le rôle de l'ensemble de Julia. Une conjecture d'Ahlfors, maintenant démontrée, affirme que l'ensemble limite est de mesure nulle s'il n'est pas égal à la sphère de Riemann. La question analogue pour les ensembles de Julia s'impose alors rapidement comme l'un des problèmes majeurs de la théorie, d'autant plus qu'une réponse positive aurait de nombreuses conséquences intéressantes. Dans les années suivantes on montre en particulier pour de nombreuses classes de polynômes (cf. ci-dessous) que l'ensemble de Julia est de mesure nulle.

Cependant, vers 1990, T. Nowicki et S. Van Strien d'une part, A. Douady d'autre part, commencent à suspecter qu'il pourrait exister des polynômes dont l'ensemble de Julia est de mesure de Lebesgue positive. T. Nowicki et S. Van Strien considèrent des polynômes $P(z) = z^d + c$, de grand degré d , pour lesquels l'orbite de l'unique point critique 0 s'organise suivant une combinatoire quasipériodique dite de Fibonacci ; mais leur programme n'aboutira pas.

A. Douady formule de son côté un programme visant à construire des polynômes quadratiques possédant un point fixe indifférent irrationnel non linéarisable dont l'ensemble de Julia (égal dans ce cas à l'ensemble de Julia rempli) est de mesure positive. Dans sa thèse [9], A. Chéritat réalise des progrès majeurs qui ne laissent plus de doute sur l'existence de tels polynômes. S'appuyant sur des travaux récents de Inou-Shishikura [13], X. Buff et A. Chéritat ont finalement obtenu ([6], [8]) le

THÉORÈME — *Il existe des polynômes quadratiques dont l'ensemble de Julia est de mesure de Lebesgue strictement positive.*

1.3. Dans la prochaine section, nous présentons quelques préliminaires sur la dynamique des polynômes quadratiques, en particulier ceux possédant un point fixe indifférent irrationnel. Cela nous permettra d'énoncer des versions un peu plus précises du théorème précédent. Dans la section suivante, on présente le plan de la construction, une version simplifiée de la proposition initiale de Douady, qui comporte 3 étapes. Ces étapes sont passées en revue au cours des 3 sections suivantes. Dans la dernière section, nous

présenterons plusieurs résultats spectaculaires sur les disques de Siegel des polynômes quadratiques obtenus par X. Buff et A. Chéritat, résultats qui utilisent certains des ingrédients essentiels de la construction précédente.

Je remercie Xavier Buff, Arnaud Chéritat et Adrien Douady pour de nombreuses et précieuses conversations, et Dominique Bidois sans laquelle ce texte n'aurait pas vu le jour à temps.

2. RAPPELS SUR LES POLYNÔMES QUADRATIQUES

2.1. À conjugaison affine près, un polynôme quadratique s'écrit de façon unique sous la forme

$$Q_c(x) = x^2 + c .$$

L'ensemble des paramètres c pour lesquels le point critique 0 a une orbite bornée est l'ensemble de Mandelbrot M . C'est aussi l'ensemble des paramètres pour lesquels l'ensemble de Julia rempli $K(Q_c)$ est connexe.

Quand on s'intéresse à un point fixe et à son multiplicateur, il est plus pratique d'effectuer un revêtement ramifié $c = \lambda/2 - \lambda^2/4$ dans le plan des paramètres et une translation $x = z + \lambda/2$ dans le plan dynamique : le point 0 est alors fixe, de multiplicateur λ . Seul le cas où $|\lambda|$ est égal ou voisin de 1 nous intéresse, on écrira $\lambda = \exp 2\pi i\alpha$ et on considérera donc la famille

$$P_\alpha(z) = \lambda z + z^2 .$$

L'image du disque $\{|\lambda| < 1\}$ par le revêtement $\lambda \rightarrow c = \lambda/2 - \lambda^2/4$ (ramifié en $\lambda = 1$ au-dessus de $c = 1/4$) est la cardioïde principale de M .

2.2. Si $c \notin M$, le point critique 0 s'échappe vers le point fixe attractif à l'infini ; l'ensemble de Julia $J(Q_c) = K(Q_c)$ est un ensemble de Cantor sur lequel le polynôme est uniformément dilatant.

Si Q_c possède une orbite périodique \mathcal{O} attractive ou indifférente, alors $c \in M$ et les autres orbites périodiques sont répulsives ; si \mathcal{O} est attractive ou indifférente rationnelle, l'orbite du point critique 0 converge vers \mathcal{O} ; l'intérieur de $K(Q_c)$ est alors exactement égal au bassin de \mathcal{O} . Si \mathcal{O} est attractive, le polynôme Q_c est uniformément dilatant sur $J(Q_c)$. On dit que Q_c est hyperbolique si $c \notin M$ ou si Q_c possède une orbite périodique attractive (à distance finie).

2.3. Soit n un entier au moins égal à 2. On dit que Q_c est n -renormalisable s'il existe des disques topologiques U, V avec $0 \in U \subset\subset V$ tels que la restriction de Q_c^n à U soit un revêtement ramifié sur V de degré 2 et $Q_c^{nk}(0)$ appartienne à U pour tout $k \geq 0$.

On dit que Q_c est infiniment renormalisable s'il existe une infinité d'entiers n tels que Q_c soit n -renormalisable.

2.4. Soit Q_c un polynôme quadratique possédant un point périodique indifférent irrationnel x_0 , de période minimale N et multiplicateur $\lambda = \exp 2\pi i\alpha$. On dit que x_0 est linéarisable si x_0 appartient à l'intérieur de $K(Q_c)$; on appelle alors disque de Siegel la composante connexe Δ de $\text{int}K(Q_c)$ qui contient x_0 ; c'est un disque topologique et toute représentation conforme $\alpha : (\mathbf{D}, 0) \rightarrow (\Delta, x_0)$ conjugue la rotation $R_\alpha(z) = \lambda z$ à la restriction de Q_c^N à Δ .

Lorsque x_0 n'est pas linéarisable, on a $K(Q_c) = J(Q_c)$; on dit que x_0 est un point de Cremer.

Notons

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots}}}} = [a_0, a_1, a_2 \dots]$$

le développement en fraction continue de α et $(p_n/q_n)_{n \geq 0}$ les réduites associées. Alors, pour que x_0 soit linéarisable, il faut et il suffit que α vérifie la condition de Brjuno (cf. [22], [2], [23])

$$\sum_{n \geq 0} q_n^{-1} \log q_{n+1} < +\infty.$$

On dit que α est de type constant si la suite $(a_n)_{n \geq 0}$ est bornée. La condition de Brjuno est alors satisfaite. Un théorème de Herman-Swiatek affirme que Δ est un quasidisque et que le point critique appartient au bord de l'orbite de Δ .

2.5. On sait que l'ensemble de Julia $J(Q_c)$ est de mesure nulle dans chacun des cas suivants :

- Q_c est hyperbolique ;
- Q_c possède un point périodique indifférent rationnel ([11]) ;
- Q_c n'est pas infiniment renormalisable et toutes ses orbites périodiques sont répulsives ([15], [21]) ;
- Q_c possède un point périodique indifférent irrationnel dont le multiplicateur $\lambda = \exp 2\pi i\alpha$ vérifie $\log a_n = \mathcal{O}(\sqrt{n})$ ([20]).

Dans tous les cas restants, Buff et Chéritat ont construit des exemples pour lesquels l'ensemble de Julia est de mesure positive.

THÉORÈME. — 1) *Il existe des paramètres c tels que Q_c a un point fixe de Cremer et $J(Q_c)$ est de mesure positive.*

2) *Il existe des paramètres c tels que Q_c a un disque de Siegel fixe et $J(Q_c)$ est de mesure positive.*

3) Il existe des paramètres c tels que Q_c est infiniment renormalisable et $J(Q_c)$ est de mesure positive.

Dans la suite, nous décrivons la construction dans le cas des points de Cremer. Les autres cas sont basés sur les mêmes méthodes, avec quelques subtilités supplémentaires.

3. PRINCIPE DE LA CONSTRUCTION

3.1. Comme on s'intéresse à des points fixes indifférents, on va écrire (cf. 2.1)

$$P_\alpha(z) = \lambda z + z^2, \quad \lambda = \exp 2\pi i \alpha.$$

On notera pour simplifier $K_\alpha = K(P_\alpha)$, $J_\alpha = J(P_\alpha)$.

Soit N un entier assez grand, qui sera déterminé ultérieurement (voir 5). On note $\mathcal{C}(N)$ l'ensemble des nombres de type constant tels que $a_n \geq N$ pour tout $n \geq 1$. Pour $\alpha = [a_0, a_1, \dots, a_n, \dots]$, $n \geq 0$ et $A \geq 1$, on pose

$$\alpha(n, A) = [a_0, a_1, \dots, a_n, A, N, N, N, \dots]$$

On a donc $\alpha(n, A) \in \mathcal{C}(N)$ si $\alpha \in \mathcal{C}(N)$ et $A \geq N$.

PROPOSITION. — Soient $\alpha \in \mathcal{C}(N)$, A_n une suite telle que $\lim q_n^{-1} \log A_n = +\infty$, et soit $\varepsilon > 0$. Si n est assez grand, on a

$$\text{Leb}(K_{\alpha(n, A_n)}) \geq (1 - \varepsilon) \text{Leb}(K_\alpha)$$

et le disque $\{|z| < \varepsilon\}$ contient un cycle périodique de $P_{\alpha(n, A_n)}$ distinct de $\{0\}$.

REMARQUE. — Cet énoncé est très probablement vrai pour tout entier $N \geq 1$, mais n'est démontré que si l'entier N est assez grand.

3.2. La proposition permet de réaliser la construction recherchée. On définit en effet une suite $(\alpha_\ell)_{\ell \geq 0}$ dans $\mathcal{C}(N)$ comme suit. On choisit arbitrairement $\alpha_0 \in \mathcal{C}(N)$, puis on construit $\alpha_{\ell+1}$ à partir de α_ℓ en posant

$$\alpha_{\ell+1} = \alpha_\ell(n, A_n).$$

Ici, la suite (A_n) est choisie de façon à vérifier l'hypothèse de la proposition et l'entier $n = n_\ell$ est choisi assez grand pour que les conclusions de la proposition soient vérifiées avec $\varepsilon_\ell = 2^{-\ell-1}$. Quitte à augmenter n_ℓ , on peut de plus garantir que la suite α_ℓ converge vers une limite irrationnelle α_∞ , et que le cycle périodique dans $\{|z| < 2^{-j}\} - \{0\}$ garanti par la proposition pour P_{α_j} (pour $j \leq \ell$) soit encore contenu dans le même disque pour $P_{\alpha_{\ell+1}}$.

Le polynôme P_{α_∞} possède les propriétés requises : tout voisinage de 0 contient un cycle périodique distinct de 0, donc 0 est un point de Cremer. On a donc $K_{\alpha_\infty} = J_{\alpha_\infty}$. Toute valeur d'adhérence d'une suite $(z_\ell)_{\ell \geq 0}$ vérifiant $z_\ell \in K_{\alpha_\ell}$ appartient à K_{α_∞} .

On a donc

$$\text{Leb}(K_{\alpha_\infty}) \geq \limsup \text{Leb}(K_{\alpha_\ell}).$$

Comme le produit $\prod(1 - \varepsilon_\ell)$ est convergent et K_{α_0} est d'intérieur non vide, on obtient bien que J_{α_∞} est de mesure de Lebesgue positive.

3.3. La démonstration de la proposition comporte trois étapes.

Dans une première étape, on contrôle le cycle périodique voisin de 0 de $P_{\alpha(n, A_n)}$ au moyen d'outils introduits par A. Chéritat. On couple ensuite ces outils aux techniques de renormalisation introduites il y a une vingtaine d'années par Douady, Ghys et moi-même pour faire un premier pas vers l'estimation de mesure : on montre que, pour tout ouvert V contenu dans le disque de Siegel Δ_α de P_α , et tout $\varepsilon > 0$, on a

$$\text{Leb}(V \cap \Delta_{\alpha(n, A_n)}) \geq \left(\frac{1}{2} - \varepsilon\right) \text{Leb}(V)$$

si n est assez grand.

Dans une deuxième étape, on montre que pour tout ouvert W contenant Δ_α , et tout nombre $\alpha' \in \mathcal{C}(N)$ suffisamment proche de α , l'orbite du point critique pour $P_{\alpha'}$ est contenue dans W . Cette affirmation est une conséquence des résultats de renormalisation d'Inou-Shishikura [13] ; c'est ici qu'il est (pour l'instant) nécessaire de supposer que l'entier N est assez grand.

Dans la troisième étape, on met à profit ce contrôle de l'orbite postcritique pour $P_{\alpha(n, A_n)}$ pour obtenir l'estimation de mesure de $K_{\alpha(n, A_n)}$. Pour promouvoir la constante $\frac{1}{2} - \varepsilon$ en $1 - \varepsilon$, on reprend en le modifiant légèrement un argument de C. Mc Mullen.

Chacune de ces étapes est détaillée dans les trois sections suivantes.

4. EXPLOSION PARABOLIQUE ET TAILLE ASYMPTOTIQUE

4.1. Soit p/q un nombre rationnel. Écrivons au voisinage de l'origine

$$P_{p/q}^q(z) = z + Az^{q+1} + \mathcal{O}(z^{q+2}).$$

Le coefficient A n'est pas nul : sinon, la dynamique locale en 0 comporterait au moins deux cycles de pétales, et chacun devrait contenir une orbite postcritique. Suivant Chéritat, définissons la taille asymptotique par

$$L(p/q) = |qA|^{-1/q}.$$

Pour α voisin de p/q , P_α possède au voisinage de 0 un cycle périodique de période q . Plus précisément, en écrivant $\alpha = p/q + \delta^q$, il existe une fonction holomorphe $\chi = \chi_{p/q}$ définie au voisinage de 0 telle que les points de ce cycle périodique soient exactement

$$\chi(\delta), \chi(\zeta\alpha), \dots, \chi(\zeta^{q-1}\delta)$$

avec $\zeta = \exp 2\pi i/q$. On a

$$|\chi'(0)| = (2\pi q^2)^{1/q} L(p/q)$$

donc la taille asymptotique exprime la vitesse de l'explosion parabolique [10].

On peut suivre le cycle périodique de période q tant qu'il n'est pas parabolique. Par conséquent, χ est holomorphe dans un disque $\{|\delta| < \rho(p/q)\}$, où $r(p/q) = \rho(p/q)^q$ est la distance de p/q à l'ensemble des paramètres $\alpha \neq p/q$ pour lesquels P_α a un q -cycle parabolique. Comme le cycle est contenu dans l'ensemble de Julia rempli, qui est uniformément borné dans la région considérée, la fonction χ est bornée dans $\{|\delta| < \rho(p/q)\}$.

On a

$$1 \geq r(p/q) \geq q^{-3},$$

où la majoration est triviale et la minoration résulte d'une inégalité sur la taille des membres de M que j'ai établie il y a 20 ans. En particulier, on a

$$\lim_{q \rightarrow +\infty} \rho(p/q) = 1.$$

4.2. Soit α un nombre vérifiant la condition de Brjuno ; on note $(p_n/q_n)_{n \geq 0}$ les réduites de α et $r(P_\alpha)$ le rayon conforme du disque de Siegel Δ_α (rayon de convergence de l'application linéarisante normalisée). Après Jellouli [14], on observe que $P_{p_n/q_n}^{q_n}$ converge vers l'identité uniformément sur les compacts de Δ_α . En particulier $P_{p_n/q_n}^{q_n}$ est univalente sur de tels compacts et on en déduit facilement

$$\liminf_n L(p_n/q_n) \geq r(P_\alpha).$$

Soit $\rho < 1$. Les fonctions χ_{p_n/q_n} sont définies dans $\{|z| < \rho\}$ pour n assez grand et y forment une famille normale. Toute valeur d'adhérence χ n'est pas constante d'après l'inégalité précédente ; comme on a

$$\chi_{p_n/q_n}(\delta \exp 2\pi i p_n/q_n) = P_{p_n/q_n}(\chi_{p_n/q_n}(\delta)),$$

la fonction χ va linéariser P_α . Comme $\rho < 1$ est arbitraire, on conclut après Chéritat que $(\chi_{p_n/q_n})_{n \geq 0}$ converge sur les compacts de $\{|z| < 1\}$ vers le biholomorphisme $\chi_\alpha : \{|z| < 1\} \rightarrow \Delta_\alpha$ tel que $\chi_\alpha(0) = 0$ et $\chi_\alpha'(0) = r(P_\alpha)$. En particulier, on a

$$\lim L(p_n/q_n) = r(P_\alpha).$$

4.3. Dans le cadre de la proposition du 3.1, on a

$$|\alpha(n, A) - p_n/q_n| \approx q_n^{-2} A^{-1}.$$

Au vu des résultats ci-dessus, l'hypothèse de la proposition ($A_n^{1/q_n} \rightarrow +\infty$) garantit bien l'existence d'un q_n -cycle périodique proche de 0 pour $P_{\alpha(n, A_n)}$ lorsque n est grand.

4.4. Il s'agit maintenant de montrer que, pour tout ouvert V de Δ_α et tout $\varepsilon > 0$, on a

$$Leb(V \cap \Delta_{\alpha(n, A_n)}) \geq \left(\frac{1}{2} - \varepsilon\right) Leb(V)$$

si n est assez grand.

Introduisons

$$f_n = \chi_{p_n/q_n}^{-1} \circ P_{\alpha(n, A_n)} \circ \chi_{p_n/q_n} ,$$

Cette suite tend vers R_α uniformément sur les compacts de $\{|z| < 1\}$. Posons aussi

$$\varepsilon_n = \alpha(n, A_n) - p_n/q_n ,$$

$$\psi_n(z) = \frac{z^{q_n}}{z^{q_n} - \varepsilon_n} = v = \exp 2\pi i q_n^2 |\varepsilon_n| w$$

et, pour tout $\rho < 1$

$$X_n(\rho) = \psi_n^{-1}(\{|v| < \frac{\rho^{q_n}}{\rho^{q_n} + |\varepsilon_n|}\}) .$$

Pour tout ouvert $W \subset \{|z| < \rho\}$ et tout $\varepsilon > 0$, on a clairement, si n est assez grand

$$Leb(W \cap X_n(\rho)) \geq \left(\frac{1}{2} - \varepsilon\right) Leb(W) .$$

Il suffit donc de montrer que, pour tout $\rho < 1$, $X_n(\rho)$ est contenu dans le disque de Siegel de f_n si n est assez grand.

Posons

$$\xi_n(z) = 2\pi i q_n z (\varepsilon_n - z^{q_n})$$

La définition de χ_{p_n/q_n} permet de factoriser :

$$f_n^{q_n}(z) - z = \xi_n(z) k_n(z) ,$$

$$\exp((-1)^n \frac{2\pi i}{q_n}) f_n^{q_n-1}(z) - z = \xi_n(z) g_n(z) ,$$

et il n'est pas difficile de voir que k_n tend vers 1 uniformément sur les compacts de $\{|z| < 1\}$, tandis que $|g_n|^{1/q_n}$ est pour tout $\varepsilon > 0$ et tout $\rho < 1$ majoré par $1 + \varepsilon$ sur $\{|z| < \rho\}$ si n est assez grand.

L'image par le revêtement ramifié χ_n du champ de vecteurs $\xi_n(z) \partial/\partial z$ est le champ $2\pi i q_n^2 \varepsilon_n v \partial/\partial v$, qui se relève en $(-1)^n \frac{\partial}{\partial w}$. Le domaine $X_n(\rho)$ correspond à

$$H_n(\rho) = \{Im w > \frac{1}{2\pi |\varepsilon_n| q_n^2} \log(1 + |\varepsilon_n| \rho^{-q_n})\} .$$

Dans la coordonnée w , les applications $f_n^{q_n}, f_n^{q_n-1}$ s'écrivent respectivement

$$F_n(w) = w + u_n(w) ,$$

$$G_n(w) = w + v_n(w) - \frac{1}{q_n^2 \varepsilon_n} .$$

Le contrôle de u_n, v_n dans $H_n(\rho)$ se déduit de celui de k_n, g_n . On a en particulier que $|v_n|^{1/q_n}$ est majoré par $1 + \varepsilon$ si n est assez grand ; comme $|\varepsilon_n|^{1/q_n}$ tend vers 0, c'est le terme de translation qui domine dans G_n . On obtient surtout dans $H_n(\rho)$

$$|u_n(w) - 1| \leq B_n |\varepsilon_n| \kappa_n(\operatorname{Re} w),$$

avec $\lim_{n \rightarrow +\infty} B_n^{1/q_n} = 1$ et

$$\kappa_n(X) = 1 + \left| 1 + \frac{\rho^{q_n}}{\rho^{q_n} + |\varepsilon_n|} \exp(2\pi i q_n^2 |\varepsilon_n| X) \right|^{-1}.$$

Ce contrôle permet (tout juste) d'itérer suffisamment F_n pour traverser une bande de largeur $(q_n^2 |\varepsilon_n|)^{-1}$. Cela suffit pour pouvoir utiliser les techniques de renormalisation introduites par Douady, Ghys et moi-même et conclure qu'on peut itérer indéfiniment les points de $H_n(\rho)$, ce qui mène à la conclusion recherchée.

5. CONTRÔLE DE L'ORBITE CRITIQUE ET RENORMALISATION SUIVANT INOU-SHISHIKURA

5.1. Les techniques de renormalisation évoquées précédemment s'appliquent à des transformations (possédant un point fixe indifférent irrationnel) dont la seule caractéristique globale est d'être univalentes dans le domaine où on les considère. Cela s'avère suffisant pour contrôler l'intérieur des disques de Siegel. Cependant, pour presque tout nombre de rotation, le bord du disque de Siegel du polynôme quadratique contient le point critique. Une autre difficulté associée à ces techniques est la présence de choix arbitraires dans la définition de la renormalisation : on considère en effet l'application de premier retour dans un secteur délimité par une courbe ℓ issue du point fixe et son image ; pour uniformiser un tel secteur, il faut le clore par une courbe joignant l'extrémité de ℓ à son image ; changer cette dernière courbe conjugue l'application renormalisée.

Pour étudier de façon plus approfondie les bords des disques de Siegel des polynômes quadratiques (et des transformations de même nature), il est donc souhaitable de travailler avec des transformations de nature plus globale qui permettent en particulier de rigidifier l'uniformisation. Le problème est alors de trouver des classes de transformation qui soient stables pour l'application de renormalisation, de façon à ce qu'on puisse itérer celle-ci autant qu'il est nécessaire.

Inou et Shishikura ont résolu partiellement ce problème en découvrant de telles classes lorsque le nombre de rotation est assez petit. Pour un nombre de rotation α de type constant, c'est donc seulement lorsque $\alpha \in \mathcal{C}(N)$, avec N assez grand, qu'on pourra itérer indéfiniment leur application de renormalisation. C'est suffisant pour la construction de Buff et Chéritat.

5.2. Il est pratique de travailler dans une coordonnée qui rejette le point fixe à l'infini.

Considérons la fraction rationnelle de degré 6

$$Q(z) = z\left(1 + \frac{1}{z}\right)^6\left(1 - \frac{1}{z}\right)^{-4}$$

dans $\{|z| > 1\}$. À l'infini, on a

$$Q(z) = z + 10 + 49z^{-1} + \dots$$

donc ∞ est un point fixe parabolique de multiplicité minimale 2. Dans $\{|z| > 1\}$, le seul point critique est $5 + \sqrt{24}$, la valeur critique correspondante est 27. Sur le cercle $\{|z| = 1\}$, on a aussi les points critiques -1 (de degré local 6) et 1 (de degré local 4), les valeurs critiques correspondantes sont 0 et ∞ .

Décrivons l'image du cercle $\{|z| = 1\}$ (en contournant par l'extérieur les points ± 1 : on parcourt l'axe réel positif entre ε ($0 < \varepsilon \ll 1$) et $R \gg 1$ puis on tourne 2 fois dans le sens négatif sur $\{|z| = R\}$ avant de revenir de R à ε et de tourner 3 fois sur $\{|z| = \varepsilon\}$ dans le sens positif.

Suivant Inou-Shishikura, définissons

$$E = \left\{ z = x + iy, \left(\frac{x - x_E}{a_E}\right)^2 + \left(\frac{y}{b_E}\right)^2 \leq 1 \right\}$$

avec $x_E = -0.18$, $a_E = 1.24$, $b_E = 1.04$. L'intérieur de E contient $\{|z| \leq 1\}$. Notons \mathcal{F}_0 la classe des applications $f = Q \circ \varphi^{-1}$, où φ est une application univalente sur $\overline{\mathbf{C}} - E$ vérifiant $\varphi(\infty) = \infty$, $\varphi'(\infty) = 1$, et telle que $\varphi(\overline{\mathbf{C}} - E)$ est un quasidisque ne contenant pas 0. Le domaine de f est donc ce quasidisque et l'image est $Q(\overline{\mathbf{C}} - E)$.

Les applications de \mathcal{F}_0 ont donc toutes un point fixe parabolique à l'infini, et il est facile de voir qu'il est toujours de multiplicité minimale 2.

Pour $\alpha \in (0, 1)$, notons \mathcal{F}_α la classe des transformations $R_{-\alpha} \circ f$, $f \in \mathcal{F}_0$. Le multiplicateur du point fixe à l'infini est alors $\exp 2\pi i \alpha$.

Une transformation dans \mathcal{F}_α s'écrit

$$f(z) = \exp(-2\pi i \alpha)z + c_0 + o(z^{-1}),$$

avec $|c_0 - 10| \leq 3$. Pour α petit, f a un autre point fixe z^* voisin de $\frac{c_0}{2\pi i \alpha}$. Supposons par exemple $\alpha > 0$. Notons ℓ la demi-droite verticale joignant z^* à $-i\infty$, U le secteur infini délimité par ℓ et $f(\ell)$; la surface de Riemann \mathcal{U} obtenue en recollant les bords de U par f est isomorphe à \mathbf{C}^* ; on l'uniformise en envoyant les bouts correspondant à z^* et ∞ en $0, \infty$ respectivement. La coordonnée uniformisante est alors définie à multiplication par un nombre complexe non nul près, et sera complètement déterminée en spécifiant l'unique valeur critique de l'application renormalisée (on la prendra égale à $27 \exp(-2\pi i \alpha^{-1})$).

La renormalisation Rf de $f \in \mathcal{F}_\alpha$ est définie au voisinage de l'infini comme l'application de premier retour dans U , lue dans la coordonnée uniformisante w de \mathcal{U} . L'application Rf a un point fixe indifférent à l'infini de multiplicateur $\exp(-2\pi i \alpha^{-1})$.

Après conjugaison par $w \rightarrow \bar{w}$, ce multiplicateur devient $\exp(2\pi i\alpha^{-1}) = \exp(2\pi i\alpha_1)$ avec $\alpha^{-1} = a_1 + \alpha_1$. Inou et Shishikura ont montré que Rf appartient à \mathcal{F}_{α_1} !

Plus précisément, notons V_1 (resp. V_{-1}) la région voisine de 1 (resp. de -1) qui s'envoie par Q sur $\{|z| \geq 27e^{4\pi}\}$ (resp. sur $\{|z| \leq 27e^{-4\pi}\}$).

L'ensemble $E' = \{|z| \leq 1\} \cup V_1 \cup V_{-1}$ est contenu dans l'intérieur de E . L'image par Q de $\bar{\mathcal{C}} - E'$ est égale à $\{|z| > 27e^{-4\pi}\}$; un point z dans ce domaine qui n'est pas sur le segment $[27e^{-4\pi}, 27e^{4\pi}]$ a 1 image inverse par Q dans $\bar{\mathcal{C}} - E'$ si $|z| > 27e^{4\pi}$, 3 si $|z| < 27e^{-4\pi}$.

Sur la surface de Riemann \mathcal{U} , munie de sa coordonnée uniformisante normalisée w , Inou et Shishikura construisent un quasidisque E'_1 contenant 0 dans son intérieur tel que

- Rf est défini sur $\bar{\mathcal{C}} - E'_1$ et envoie ce domaine sur $\{|w| > 27e^{-4\pi}\}$
- il existe un isomorphisme Φ de $\bar{\mathcal{C}} - E'$ sur $\mathcal{U} - E'_1$ tel que

$$Rf \circ \Phi = Q.$$

On voit que Rf est de la forme cherchée sur un domaine plus grand (car $\bar{\mathcal{C}} - E$ est d'adhérence compacte dans $\bar{\mathcal{C}} - E'$). Cela leur permet de conclure, à l'aide d'un argument de type lemme de Schwartz dans un espace de Teichmüller approprié, que l'opérateur de renormalisation est uniformément contractant (à nombre de rotation fixé). Résumons :

THÉORÈME (INOUSHISHIKURA), [13]. — *Si α est assez petit, l'opérateur de renormalisation envoie \mathcal{F}_α dans \mathcal{F}_{α_1} , et est une contraction uniforme pour la distance de Teichmüller.*

Inou et Shishikura déduisent ce résultat par perturbation d'un résultat analogue pour \mathcal{F}_0 . Leur démonstration entremêle des calculs numériques délicats et des estimations raffinées sur les fonctions univalentes.

5.3. On prend N assez grand de façon à ce que le théorème d'Inou-Shishikura s'applique lorsque $|\alpha| < 1/N$.

Soit alors $\alpha \in \mathcal{C}(N)$. Bien que l'application $z \rightarrow z^2(1+\lambda z)^{-1}$ (conjuguée au polynôme quadratique par l'inversion $z \rightarrow \frac{1}{z}$) n'appartienne pas à \mathcal{F}_α on peut définir sa renormalisée et montrer après Inou-Shishikura qu'elle appartient à \mathcal{F}_{α_1} , ce qui permet donc de définir les renormalisées de tous ordres.

Soit W_n le domaine formé des points dont l'orbite jusqu'à l'ordre q_n est représentée par la renormalisée d'ordre n (une définition précise de W_n dépend de choix arbitraires, mais les assertions qui suivent sont indépendantes de ces choix). La suite des domaines $(W_n)_{n \geq 0}$ est décroissante et on peut montrer que l'intersection des W_n est exactement égale à l'adhérence du disque de Siegel Δ_α . (On utilise ici que α est de type constant : $\partial\Delta_\alpha$ est un quasicercle sur lequel l'orbite postcritique est dense.)

On peut maintenant conclure que, pour tout ouvert W contenant $\overline{\Delta}_\alpha$, on a $\overline{\Delta}_{\alpha'} \subset W$ si $\alpha' \in \mathcal{C}(N)$ est assez proche de α : on a en effet $\overline{W}_{n_0} \subset W$ pour un entier n_0 choisi assez grand ; la continuité évidente de l'opérateur de renormalisation implique qu'on a encore $\overline{W}_{n_0}(\alpha') \subset W$ si α' est assez proche de α , et donc $\overline{\Delta}_{\alpha'} \subset W$.

6. FIN DE LA DÉMONSTRATION DE LA PROPOSITION

6.1. Il reste à voir que, pour tout $\varepsilon > 0$, on a

$$\text{Leb}(K_{\alpha(n, A_n)}) \geq (1 - \varepsilon)\text{Leb}(K_\alpha)$$

si n est assez grand.

Comme J_α est de mesure nulle, la mesure de Lebesgue de K_α est égale à celle de son intérieur. L'intérieur de K_α est l'union des préimages de Δ_α . Il suffit donc de montrer que pour tout $\varepsilon > 0$, on a

$$\text{Leb}(K_{\alpha(n, A_n)} \cap \Delta_\alpha) \geq (1 - \varepsilon)\text{Leb}(\Delta_\alpha)$$

si n est assez grand. On sait déjà d'après la section 4 qu'on a, pour tout ouvert $V \subset \Delta_\alpha$,

$$\text{Leb}(\Delta_{\alpha(n, A_n)} \cap V) \geq \left(\frac{1}{2} - \varepsilon\right)\text{Leb}(V)$$

si n est assez grand. Il est facile de se convaincre que la constante $1/2$ est optimale ; donc, pour récupérer l'autre moitié de la mesure, il faut utiliser les préimages de $\Delta_{\alpha(n, A_n)}$.

On va esquisser l'adaptation par Buff et Chéritat d'un argument de Mc Mullen [18]. Celui-ci montrait (entre autres) que, si α est de type constant, J_α est de dimension de Hausdorff strictement inférieure à 2 et tout point du bord de Δ_α est un point de densité de K_α .

6.2. On note Δ'_α la composante de $P^{-1}(\Delta_\alpha)$ distincte de Δ_α , c'est-à-dire le domaine symétrique de Δ_α par rapport au point critique.

Un voisinage W d'un point z est dit équilibré s'il existe $r > 0$ tel qu'on ait

$$B(z, C^{-1}r) \subset W \subset B(z, Cr).$$

Ici et dans la suite, C désigne des constantes ne dépendant que de α .

Pour $\ell \geq 0$, on pose $\delta_\ell(z) = |P_\alpha^{q_\ell}(z) - z|$.

Comme Δ_α est un quasidisque, et α est de type constant, on peut construire une suite décroissante de domaines de Jordan V_ℓ qui possède les propriétés suivantes :

(i) pour tout $z \in \partial\Delta_\alpha$, on a

$$B(z, C^{-1}\delta_\ell(z)) \subset V_\ell ;$$

(ii) pour tout $z \in V_\ell - V_{\ell+1}$, on a

$$B(z, C^{-1}\delta_\ell(z)) \subset V_{\ell-1} - V_{\ell+2} ;$$

$$(iii) \quad V_\ell \subset \bigcup_{\partial \Delta_\alpha} B(z, C\delta_\ell(z)) \bigcup \Delta_\alpha ;$$

$$(iv) \quad P_\alpha(V_\ell) \subset \text{int } V_{\ell-1} ;$$

(v) pour tout $z \in V_\ell - V_{\ell+1}$, il existe un voisinage équilibré W_z de z et un entier $m \leq q_\ell$ tel que $P^j(W_z)$ soit contenu dans $V_{\ell-1} - V_{\ell+2}$ pour $0 \leq j \leq m$, P^m soit univalente de distortion bornée sur W_z , et on ait

$$\text{Leb}(P_\alpha^m(W_z) \cap \Delta'_\alpha) \geq c_0 \text{Leb}(P_\alpha^m(W_z)) .$$

La densité relative uniforme de $\Delta_{\alpha(n, A_n)}$ dans Δ_α et un argument de compacité impliquent, si n est assez grand (ℓ étant fixé)

$$\text{Leb}(P_{\alpha(n, A_n)}^m(W_z) \cap \Delta'_{\alpha(n, A_n)}) \geq \frac{c_0}{3} \text{Leb}(P_{\alpha(n, A_n)}^m(W_z))$$

d'où l'on tire

$$\text{Leb}(W_z \cap P_{\alpha(n, A_n)}^{-(m+1)}(\Delta_{\alpha(n, A_n)})) \geq c_1 \text{Leb}(W_z) .$$

uniformément en $z \in V_\ell - V_{\ell+1}$ si n est assez grand.

6.3. Notons $Z_{n, \ell}$ l'ensemble des points $z \in \Delta_\alpha$ dont l'orbite sous $P_{\alpha(n, A_n)}$ n'est pas contenue dans V_ℓ .

Soit $u \in Z_{n, \ell+1}$. D'après (iv) ci-dessus, si n est assez grand, le premier point $z = P_{\alpha(n, A_n)}^k(u)$ de l'orbite de u pour $P_{\alpha(n, A_n)}$ qui n'appartient pas à $V_{\ell+1}$ appartient à V_ℓ .

D'après la section 5, l'orbite postcritique de $P_{\alpha(n, A_n)}$ est contenue dans $V_{\ell+3}$ si n est assez grand. On a donc une branche inverse de $P_{\alpha(n, A_n)}^k$ définie et à distortion bornée sur W_z envoyant z sur u . L'image est un voisinage équilibré W_u de u . On a

$$\text{Leb}(W_u - Z_{n, \ell-1}) \geq c_2 \text{Leb}(W_u) .$$

Par ailleurs, on a $W_u \cap \Delta_\alpha \subset Z_{n, \ell+2}$ puisque $W_z \cap V_{\ell+2} = \emptyset$.

Un argument à la Vitali permet de conclure qu'il existe $\lambda < 1$ tel qu'on ait

$$\text{Leb}(Z_{n, \ell-1}) \leq \lambda \text{Leb}(Z_{n, \ell+2})$$

pour $n \geq n(\ell)$. Cela suffit pour conclure qu'on a, pour tout $\ell \geq 0$

$$\lim_{n \rightarrow +\infty} \text{Leb } Z_{n, \ell} = 0 .$$

Pour $\ell = 0$, c'est le résultat recherché. Cela conclut l'esquisse de preuve de la proposition de la section 3.

7. DISQUES DE SIEGEL DES POLYNÔMES QUADRATIQUES

7.1. Taille des disques

Soit α un nombre irrationnel, $(p_n/q_n)_{n \geq 0}$ la suite de ses réduites. Définissons

$$\alpha_0 = \{\alpha\}, \quad \alpha_n = \{\alpha_{n-1}^{-1}\} \quad \text{pour } n \geq 1,$$

et posons, pour $n \geq -1$

$$\beta_n = (-1)^n (q_n \alpha - p_n) = \prod_{\ell=0}^n \alpha_\ell.$$

La série

$$B(\alpha) = \sum_{n \geq 0} \beta_{n-1} \log \alpha_n^{-1}$$

converge si et seulement si α vérifie la condition de Brjuno.

Rappelons que pour un nombre α vérifiant la condition de Brjuno, on note $r(P_\alpha)$ le rayon conforme du disque de Siegel du polynôme quadratique P_α . Posons alors

$$\Upsilon(\alpha) = B(\alpha) + \log r(P_\alpha).$$

THÉORÈME (BUFF-CHÉRITAT), [4][5]. — *La fonction $\alpha \rightarrow \Upsilon(\alpha)$ s'étend en une fonction continue sur $\mathbf{T} = \mathbf{R}/\mathbf{Z}$.*

On notera que chacun des deux termes composant Υ est extrêmement singulier, le premier de nature arithmétique et le second de nature géométrique.

J'avais auparavant montré [23] que Υ est minorée et qu'on a une majoration

$$\log r(P_\alpha) \leq -(1 - \varepsilon)B(\alpha) + C_\varepsilon$$

pour tout $\varepsilon > 0$. X. Buff a amélioré récemment mon argument [3] et obtient ainsi de façon élémentaire que Υ est majorée. S. Marmi avait conjecturé, suite à des expériences numériques, que Υ est continue [16]. Une conjecture plus forte [17] est que Υ est höldérienne d'exposant $1/2$.

Buff et Chéritat donnent une définition directe de la valeur de Υ aux nombres qui ne vérifient pas la condition de Brjuno.

Considérons d'abord un nombre irrationnel α . Posons

$$\Phi_n(\alpha) = \sum_0^n \beta_{\beta_{\ell-1}} \log \alpha_\ell^{-1}$$

et notons $r_n(\alpha)$ la distance de 0 à l'ensemble des points périodiques (distincts de 0) de période $\leq q_n$. Ils ont montré qu'on a

$$\Upsilon(\alpha) = \lim_{n \rightarrow +\infty} \Phi_n(\alpha) + \log r_n(\alpha).$$

Pour un nombre rationnel $\alpha = p/q$, soit m l'entier tel que $\alpha_m = 0$; posons

$$\Phi_{\text{trunc}}(\alpha) = \sum_{0 \leq \ell < m} \beta_{\ell-1} \log \alpha_{\ell}^{-1}.$$

On a alors

$$\Upsilon(p/q) = \Phi_{\text{trunc}}(p/q) + \frac{\log 2\pi}{q} + \log L(p/q)$$

où $L(p/q)$ est la taille asymptotique de la section 4. La preuve de Buff et Chéritat s'appuie principalement sur les méthodes évoquées dans la section 4.

7.2. Géométrie des disques

Perez-Marco a construit [19] des applications univalentes dans le disque unité possédant un disque de Siegel relativement compact dont le bord est une courbe de Jordan de classe C^{∞} .

Par une méthode totalement différente, Buff et Chéritat ont montré que ceci se produisait dans de très nombreuses familles, y compris dans la famille des polynômes quadratiques. Leur démonstration a été simplifiée en collaboration avec A. Avila. Une simplification ultérieure est due à L. Geyer [12].

THÉORÈME (AVILA-BUFF-CHÉRITAT), [1] — *Soient α_0 un nombre de Brjuno, $0 < r < r(P_{\alpha_0})$, $\varepsilon > 0$. Il existe un ensemble de Cantor $K \subset [\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]$ formé de nombre de Brjuno α tels que le disque de Siegel Δ_{α} soit de rayon conforme égal à r et ait pour bord une courbe de Jordan de classe C^{∞} .*

La méthode de démonstration est suffisamment souple pour obtenir d'autres classes de régularité, par exemple des courbes de classes C^n mais pas C^{n+1} [7].

On notera que, pour les nombres α du théorème ci-dessus, le point critique de P_{α} ne peut se trouver sur le bord de Δ_{α} ; d'après un théorème de Herman, cela veut dire qu'aucun de ces nombres α ne peut vérifier la condition arithmétique \mathcal{H} associée à la linéarisation analytique des difféomorphismes analytiques du cercle [24].

RÉFÉRENCES

- [1] A. AVILA, X. BUFF, A. CHÉRITAT – *Siegel disks with smooth boundaries*, Acta Mathematica (2004) 193, 1-30.
- [2] A.D. BRJUNO – *Analytic forms of differential equations*, Trans. Mosc. Math. Soc. **25** (1971).
- [3] X. BUFF – *Disques de Siegel et ensembles de Julia d'aire strictement positive*, manuscrit.

- [4] X. BUFF, A. CHÉRITAT – *Upper bound for the Size of Quadratic Siegel Disks*, *Inventiones Mathematicae* (2004) 156/1, 1-24.
- [5] X. BUFF, A. CHÉRITAT – *The Brjuno function continuously approximates the size of quadratic Siegel disks*, à paraître aux *Annals of Math*.
- [6] X. BUFF, A. CHÉRITAT – *Ensembles de Julia quadratiques de mesure de Lebesgue strictement positive*, *Comptes Rendus Mathématiques* (2005) 341/11, 669-674.
- [7] X. BUFF, A. CHÉRITAT – *How regular can the boundary of a quadratic Siegel disk be ?*, à paraître aux *Proc. of the A.M.S.*
- [8] X. BUFF, A. CHÉRITAT – *Quadratic Julia sets with positive area*, Preprint.
- [9] A. CHÉRITAT – *Recherche d'ensembles de Julia de mesure de Lebesgue positive*, Thèse, Orsay (2001).
- [10] A. CHÉRITAT – *Sur la vitesse d'explosion des points paraboliques dans la famille quadratique*, *Comptes Rendus Mathématiques* (2002) 334/12, 1107-1112.
- [11] A. DOUADY, J. HUBBARD – *Étude dynamique des polynômes complexes I et II*, *Publ. Math. d'Orsay* (1984-85).
- [12] L. GEYER – *Smooth Siegel discs without number theory : A remark on a proof by Buff and Chéritat*, preprint (2003).
- [13] H. INOU, M. SHISHIKURA – *The renormalization for parabolic fixed points and their perturbation*, Preprint.
- [14] H. JELLOULI – *Perturbation d'une fonction linéarisable*, *London Math. Soc. Lect. Note* **274**, Ed. Tan Lei, Cambridge Univ. Press (2000), 227-252.
- [15] M. LYUBICH – *On the Lebesgue measure of the Julia set of a quadratic polynomial*, Stonybrook IMS, preprint (1991/10).
- [16] S. MARMI – *Critical Functions for Complex Analytic Maps*, *J. Phys. A : Math. Gen.* **23** (1990), 3447-3474.
- [17] S. MARMI, P. MOUSSA, J-C. YOCCOZ – *The Brjuno functions and their regularity properties*, *Comm. Math. Phys.* **186** (1997), 265-293.
- [18] C. McMULLEN – *Self-similarity of Siegel disks and Hausdorff dimension of Julia sets*, *Acta Math.*, **180** (1998), 247-292.
- [19] R. PÉREZ-MARCO – *Siegel disks with smooth boundary*, Preprint.
- [20] C.L. PETERSEN, S. ZAKERI – *On the Julia set of a typical quadratic polynomial with a Siegel disk*, *Ann. of Math.* **159** (2004) 1-52.
- [21] M. SHISHIKURA – Unpublished.
- [22] C.L. SIEGEL – *Iteration of analytic functions*, *Ann. of Math.* **vol 43** (1942).
- [23] J-C. YOCCOZ – *Petits diviseurs en dimension 1*, *S.M.F., Astérisque* **231** (1995).

- [24] J-C. YOCCOZ – *Analytic linearization of circle diffeomorphisms, in Dynamical systems and small divisors*, LNM **1784** (2002).

Jean-Christophe YOCCOZ

Collège de France

Département de Mathématiques

3, rue d'Ulm

F-75231 Paris Cedex 05

E-mail : `jean-c.yoccoz@college-de-france.fr`